Neurocomputing 145 (2014) 369–373

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust latent low rank representation for subspace clustering

Hongyang Zhang [a], Zhouchen Lin [a], Chao Zhang [a,*], Junbin Gao [b]

[a] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, PR China
[b] School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW 2795, Australia

ARTICLE INFO

Article history:
Received 7 September 2013
Received in revised form
30 April 2014
Accepted 15 May 2014
Communicated by Zhi Yong Liu
Available online 27 May 2014

Keywords:
Subspace clustering
Latent low rank representation

ABSTRACT

Subspace clustering has found wide applications in machine learning, data mining, and computer vision. Latent Low Rank Representation (LatLRR) is one of the state-of-the-art methods for subspace clustering. However, its effectiveness is undermined by a recent discovery that the solution to the noiseless LatLRR model is non-unique. To remedy this issue, we propose choosing the sparest solution in the solution set. When there is noise, we further propose preprocessing the data with robust PCA. Experiments on both synthetic and real data demonstrate the advantage of our robust LatLRR over state-of-the-art methods.

1. Introduction

High-dimensional data in the real world are often structured. As the basic assumption of manifold learning, the data usually distribute near low-dimensional manifolds. Subspace clustering aims at clustering data into a union of subspaces [1] while recognizing possible noises and outliers simultaneously [2]. It has found wide applications in machine learning, data mining, and computer vision, such as image segmentation [3], motion segmentation [2,4–7], saliency detection [8], and face clustering [2,4–6].

1.1. Related work

According to Vidal [1], subspace clustering algorithms can be divided into four approaches: statistical, algebraic, iterative, and spectral clustering based. Spectral clustering based methods first learn a similarity matrix for measuring the similarity between the data points and then perform spectral clustering to segment the samples. As a successful example of spectral clustering based methods, Sparse Subspace Clustering (SSC) [7] expresses each data sample as a sparse linear or affine combination of other samples and uses the linear combination coefficients as similarity between samples. The mathematical model of SSC is as follows:

$$\min_Z \|Z\|_1, \quad \text{s.t. } X = XZ,\ \operatorname{diag}(Z) = 0, \tag{1}$$

where X is the data matrix with samples as its columns. More notations can be found in Table 1. The last constraint in problem (1) is to avoid the trivial solution I. It is shown in [7] and [9] that, when the subspaces are either disjoint or independent, the solution to problem (1) is block diagonal, i.e., $Z_{jk} = 0$ if the corresponding samples X_j and X_k lie in different subspaces. The block diagonality reveals the structure of subspaces.

Low Rank Representation (LRR) [2,4] is another recently proposed spectral clustering based method for subspace clustering. It seeks the lowest-rank representation of the data samples. The model of LRR in the noiseless case is

$$\min_Z \|Z\|_*, \quad \text{s.t. } X = XZ, \tag{2}$$

where the nuclear norm $\|\cdot\|_*$ serves as a convex surrogate of the rank function. It is shown by Liu et al. [4] that when the data are noise free and drawn from independent subspaces, the optimal solution to problem (2) is also block diagonal.

There are several successful trials on improving LRR. By combining SSC with LRR, Zhuang et al. [10] computed non-negative low rank and sparse graphs for semi-supervised learning. Wei and Lin [6] proposed Robust Shape Interaction (RSI) by first denoising the data using robust PCA [11] and then applying LRR to the denoised data. Favaro et al. [12] proposed the same idea but integrated denoising with LRR.

LRR works well only when the samples are sufficient. To overcome such a drawback, Liu et al. [5] proposed Latent Low Rank Representation (LatLRR). Another model for resolving such an issue is Fixed Rank Representation [13] which bounds the rank of representation matrix Z. LatLRR assumes that the observed samples are linear combinations of both observed and un-observed data. After some deduction, the noiseless LatLRR is modeled as follows:

$$\min_{Z,L} \|Z\|_* + \|L\|_*, \quad \text{s.t. } X = XZ + LX, \tag{3}$$

* Corresponding author.
E-mail addresses: hy_zh@pku.edu.cn (H. Zhang), zlin@pku.edu.cn (Z. Lin), chzhang@cis.pku.edu.cn (C. Zhang), jbgao@csu.edu.au (J. Gao).

http://dx.doi.org/10.1016/j.neucom.2014.05.022
0925-2312/

Table 1
Summary of main notations used in this paper.

Notations	Meanings
Capital letter	A matrix
I, $\mathbf{0}$, $\mathbf{1}$	The identity matrix, all-zero matrix, and all-one vector
V_i	The ith column of the matrix V
V_{ij}	The entry at the ith row and jth column of matrix V
$\|\cdot\|_*$	Nuclear norm, the sum of all the singular values
$\|\cdot\|_1$, $\|\cdot\|_F$	$\|V\|_1=\sum_{i,j}\|V_{ij}\|$, $\|V\|_F=\sqrt{\sum_{i,j}(V_{ij})^2}$
tr $(\cdot)$, $\|\cdot\|$	tr $(V)=\sum_i V_{ii}$, $[\|V\|]_{ij}=\|V_{ij}\|$
diag $(\cdot)$	The diagonal entries of a matrix

where Z can be used for subspace clustering, while L is for feature extraction, thus integrating these two tasks into a unified framework.

1.2. Our contributions

Although LatLRR has been successfully applied, e.g., to motion segmentation [5] and image classification [14], we have found that the solution to noiseless LatLRR is not unique [15]. Actually, there are infinitely many solutions to problem (3) as stated in the following theorem.

Theorem 1.1 (*Zhang et al. [15]*). *The complete solutions to the noiseless LatLRR (3) are as follows:*

$$Z^*=V_X W V_X^T \quad \text{and} \quad L^*=U_X(I-W)U_X^T, \tag{4}$$

where $U_X\Sigma_X V_X^T$ is the skinny singular value decomposition (SVD) of the data matrix X and W is any block diagonal matrix satisfying

$$\begin{cases} \text{its blocks are compatible with } \Sigma_X, \text{ i.e.,} \\ \quad \text{if } [\Sigma_X]_{ii}\neq[\Sigma_X]_{jj} \text{ then } W_{ij}=0; \\ \text{both } W \text{ and } I-W \text{ are positive semi-definite.} \end{cases} \tag{5}$$

As a result, the effectiveness of LatLRR reported in the previous literature may be questionable [15], because which one in the solution set is chosen actually depends on the iteration process for solving LatLRR. This paper aims at addressing this issue of LatLRR. The contributions of this paper include the following:

- For noiseless LatLRR, we propose choosing the sparsest solution Z^* from set (4). We also design an Alternating Direction Method (ADM) based algorithm to efficiently solve for the sparsest Z^*. In this way, we naturally combine the low-rankness, sparsity, and positive semi-definiteness criteria for the solution, which some researchers have elaborated on [10,16,17].
- When there are noises or sparse corruptions, we propose first denoising the data with robust PCA [11] and then finding the low rank and sparse solution by the method sketched above. We call our approach as robust LatLRR. Experiments show that our robust LatLRR outperforms state-of-the-art subspace clustering methods.

2. Robust latent low rank representation

2.1. Choosing the sparsest solution

It is easy to see that there are infinitely many block diagonal matrices W that satisfy properties (5). So we have to choose the optimal solution $Z^*=V_X W V_X^T$ such that it is most suitable for subspace clustering.

According to Wright et al. [18], an informative similarity matrix should have three characteristics: high discriminating power, adaptive neighborhood, and high sparsity [10]. Since the graphs constructed by LatLRR have been reported to have high discriminating power and adaptive neighborhood [5], we consider high sparsity as the criterion to choose the optimal solution Z^*. This leads to the following optimization problem for selecting the best Z^* from solution set (4):

$$\begin{aligned} &\min_{Z,W}\|Z\|_1, \quad \text{s.t. } Z=V_X W V_X^T,\ W \text{ is diagonal,} \\ &0\le \text{diag}\,(W)\le 1, \quad \text{tr}(W)=1, \end{aligned} \tag{6}$$

where we require the trace of W to be 1 so as to avoid the trivial solution $Z=\mathbf{0}$. Note that here we require W to be diagonal because for randomly sampled data matrix X, its singular values are usually distinct. So W has to be diagonal in order to be compatible with Σ_X. In this case, both W and $I-W$ being positive semi-definite becomes $0\le \text{diag}(W)\le 1$. Notice that $V_X V_X^T$ satisfies the constraint in (6) and is block-diagonal when the data lie strictly in several low-dimensional subspaces [2,4,6]. We could expect $V_X V_X^T$ to be our sparsest solution when there is no noise.

2.2. Solving the optimization problem

Since W is diagonal, we may parameterize it by its diagonal entries: $w=\text{diag}(W)$, where $w=(w_1,w_2,\ldots,w_r)^T$ and r is the rank of X. Next, we introduce auxiliary variables $c=(c_1,c_2,\ldots,c_r)^T$ in order to decouple the constraints. So we reformulate (6) as

$$\begin{aligned} &\min_{Z,w,c}\|Z\|_1, \quad \text{s.t. } Z=\sum_{i=1}^{r} w_i V_i V_i^T,\ \mathbf{1}^T w=1, \\ &w=c, \quad c\ge 0, \end{aligned} \tag{7}$$

where we write $[V_X]_i$, the ith column of V_X, as V_i for brevity, and $c\ge 0$ means $c_i\ge 0$, $i=1,\ldots,r$.

Problem (7) could be easily solved by ADM, where the variables are updated alternately by minimizing the augmented Lagrangian function of (7):

$$\begin{aligned} &L(Z,w,c,Y,\alpha,\mu) \\ &\quad=\|Z\|_1+\text{tr}\left[Y^T\left(\sum_{i=1}^{r} w_i V_i V_i^T-Z\right)\right]+\alpha^T(w-c) \\ &\qquad+\frac{\mu}{2}\left(\left\|\sum_{i=1}^{r} w_i V_i V_i^T-Z\right\|_F^2+\|w-c\|_F^2\right), \end{aligned} \tag{8}$$

where Y and α are the Lagrange multipliers and $\mu>0$ is a penalty parameter. More specifically, the update of variables goes as follows:

$$\begin{aligned} Z^{(k+1)}&=\arg\min_{Z} L(Z,w^{(k)},c^{(k)},Y^{(k)},\alpha^{(k)},\mu^{(k)}), \\ w^{(k+1)}&=\arg\min_{w,\mathbf{1}^T w=1} L(Z^{(k+1)},w,c^{(k)},Y^{(k)},\alpha^{(k)},\mu^{(k)}), \\ c^{(k+1)}&=\arg\min_{c,c\ge 0} L(Z^{(k+1)},w^{(k+1)},c,Y^{(k)},\alpha^{(k)},\mu^{(k)}), \end{aligned} \tag{9}$$

where the superscript k is the number of iterations. Then Y, α, and μ are updated. More details can be found in Algorithm 1, where we have worked out the solutions to (9).

Algorithm 1. Solving problem (7) by ADM.

Initialize: $Z^{(0)}=Y^{(0)}=\mathbf{0}$, $w^{(0)}=c^{(0)}=1/r\mathbf{1}$, $\alpha^{(0)}=0$, $\mu^{(0)}=10^{-3}$, $\mu_{\max}=10^{20}$, $\rho=1.75$.
while not converged **do**
1. Update Z by
$Z^{(k+1)}=\arg\min_Z \frac{1}{\mu^{(k)}}\|Z\|_1+\frac{1}{2}\|Z-W^{(k)}\|_F^2=\Theta_{1/\mu^{(k)}}(W^{(k)})$,
where $W^{(k)}=\sum_{i=1}^{r} w_i^{(k)} V_i V_i^T+\frac{Y^{(k)}}{\mu^{(k)}}$ and
$\Theta_\varepsilon(x)=\text{sign}(x)\max(|x|-\varepsilon,0)$ is the soft-thresholding operator [11].

2. Update w by
$w_i^{(k+1)} = \frac{1}{2}(V_i^T Z^{(k+1)} V_i + c_i^{(k)} - \frac{1}{\mu^{(k)}}(\beta^{(k)} + V_i^T Y^{(k)} V_i + \alpha_i^{(k)}))$,
$i = 1, \ldots, r$, where
$\beta^{(k)} = \frac{\mu^{(k)}}{r}\left\{\sum_{i=1}^{r}[V_i^T Z^{(k+1)} V_i + c_i^{(k)} - \frac{1}{\mu^{(k)}}(V_i^T Y^{(k)} V_i + \alpha_i^{(k)})] - 2\right\}$.
3. Update c by $c = \max(w^{(k+1)} + \alpha^{(k)}/\mu^{(k)}, 0)$.
4. Update Y and α by
$Y^{(k+1)} = Y^{(k)} + \mu^{(k)}(\sum_{i=1}^{r} w_i^{(k+1)} V_i V_i^T - Z^{(k+1)})$,
$\alpha^{(k+1)} = \alpha^{(k)} + \mu^{(k)}(w^{(k+1)} - c^{(k+1)})$.
5. Update μ by $\mu^{(k+1)} = \min(\rho\mu^{(k)}, \mu_{\max})$.
end while

2.3. Extending to noisy case

When there are noises or corruptions, it is inappropriate to use the noisy data X itself as the dictionary. Both Wei et al. [6] and Favaro et al. [12] have noticed this issue. Wei et al. [6] proposed to denoise the data first by robust PCA [11] and then apply LRR to the denoised data. Favaro et al. [12] proposed to integrate denoising and LRR in a unified framework, resulting in a non-convex optimization problem. Here we follow the idea by Wei et al. [6]. We first use robust PCA [11]

$$\min_{X,E} \|X\|_* + \lambda\|E\|_1, \quad \text{s.t. } D = X + E, \tag{10}$$

to denoise the data, where D is the noisy data. Then we apply the noiseless LatLRR to the denoised data X and use Algorithm 1 to choose the sparsest solution among the solution set, as described in Section 2.1. Since robust PCA can remove noises and corruptions effectively [11], we expect that our robust LatLRR will be robust to noises and corruptions, as verified by our experiments. Note that the parameter of robust PCA (10) is valid for a large range of values under some mild conditions[1] [11]. Thus the λ in our method is insensitive.

Algorithm 2. Robust LatLRR for noisy subspace clustering.

Input: Observed noisy data D, regularization parameter λ, the subspace number k.
1. Conduct robust PCA on D with parameter λ and obtain the skinny SVD $U_X \Sigma_X V_X^T$ of the X.
2. Perform Algorithm 1 on V_X and get the optimal representation matrix Z^*.
3. Cut the graph with weight matrix $|Z^*|$ into k classes by the NCut algorithm.
Output: The label of each sample.

2.4. Clustering with robust LatLRR

After obtaining the optimal representation matrix Z^*, we follow the conventional way for clustering. We first build a graph with a weight matrix $|Z^*|$, then perform Normalized Cut (NCut) [19] on the graph to achieve clustering. The complete algorithm for our robust LatLRR based clustering is summarized in Algorithm 2.

There are three advantages of robust LatLRR compared with other state-of-the-art subspace clustering algorithms:

1. Robust LatLRR achieves both the low-rankness and the sparsity of the representation matrix. Thus, according to [18], the graph constructed by our method is more informative.
2. Robust LatLRR automatically obtains a symmetric representation matrix. So we do not have to use $(|(Z^*)^T| + |Z^*|)/2$ as the weigh matrix to construct a graph, as most of the spectral clustering based methods did.
3. Robust LatLRR also obtains a semi-definite representation matrix, which was advocated by Ni et al. and was explicitly enforced in their formulation [17].

In summary, robust LatLRR elegantly produces a representation matrix with good properties that some researchers have sought after. As a result, robust LatLRR outperforms the state-of-the-art methods for subspace clustering, as shown by our forthcoming experiments.

3. Experiments

We test our robust LatLRR and other methods, including SSC [7], LRR [2,4], RSI [6], LRSC [12], and LatLRR [5], on both synthetic and real data. There are two real data sets: the Hopkins 155[2] motion data set and the Extended Yale B[3] face data set. These data sets contain quite different noise levels, thus are suitable for testing the influence of noise and corruption on the performance. The codes of state-of-the-art methods are all provided by their corresponding authors, including NCut algorithm[4] [19]. *For fair comparison, all the specialized pre- and post-processing steps are removed.*

3.1. Synthetic experiment

We first test the performance of different algorithms on noisy synthetic data with an increasing percentage of corruptions. The data are generated as follows: we construct five independent subspaces $\{S_i\}_{i=1}^5 \subset R^{100}$ whose bases $\{U_i\}_{i=1}^5$ are 100×4 random matrices consisting of orthonormal columns. By computing $X_i = U_i Q_i$, $1 \le i \le 5$, where Q_i is a 4×20 i.i.d. $N(0,1)$ matrix, we sample 20 data vectors from each subspace and obtain a clean data matrix $X = [X_1, X_2, \ldots, X_5]$. For comparing the robustness, we add dense Gaussian noise, with a mean 0 and a variance 0.1^2, to the clean data. We further corrupt the data by randomly choosing different percentages of entries in X and adding them with noises uniformly distributed on $[-1, 1]$.

Since the parameter λ depends on the percentage of corruptions for several methods, we follow the procedure in [4] by tuning λ to be the best at 20% corruptions for every method. We repeat the experiment by 20 times. Fig. 1 reports the mean accuracies of the six methods with respect to the percentage of corruptions. It shows that our robust LatLRR outperforms other state-of-the-art methods with a clear margin.

3.2. Real experiment on Hopkins 155 data set

The Hopkins 155 data set contains 156 sequences, each of which is composed of 39–550 data points drawn from two or three motion objects. Each sequence is independent. So there are 156 subspace clustering tasks in total. Since the data set contains only complete trajectories and no outliers, we regard this test as a slight corruption case.

We compare the performance of the six methods on the 156 tasks, where the parameters have been tuned to be the best so that the mean errors are minimum. Table 2 tabulates the maximums, the means, and the standard deviations of the misclassification

[1] According to [11], the λ in (10) can be simply set as $1/\sqrt{n}$ under some mild conditions, where n refers to the size of the input matrix. In many applications where these conditions do not hold, scholars prefer to tune the parameter by some technology like cross validation.

[2] http://www.vision.jhu.edu/data/hopkins155/.
[3] http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.
[4] http://www.cis.upenn.edu/~jshi/software/.

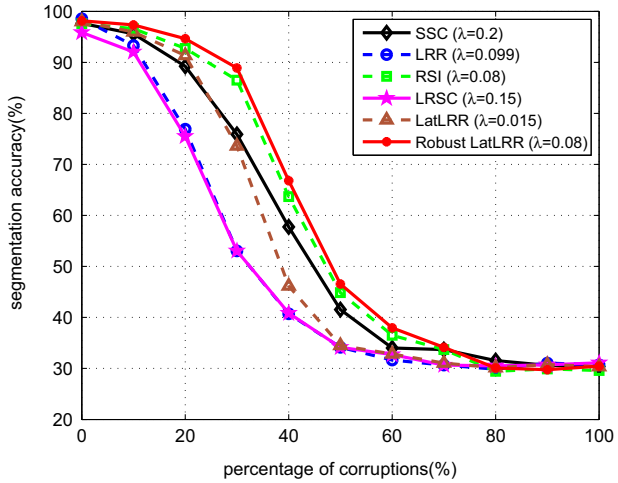

Fig. 1. Comparison on the synthetic data as the percentage of corruptions increases. For every method, the parameter λ is tuned to be the best at 20% percentage of corruptions.

Table 2
Segmentation errors (%) on the Hopkins 155 data set. For robust LatLRR, the parameter λ is set as $0.806/\sqrt{n}$, which is within the suggested range of robust PCA [11]. The parameters of other methods are also tuned to be the best.

Experiment Results on the Hopkins 155 Data Set	SSC	LRR	RSI	LRSC	LatLRR	Robust LatLRR
MAX	46.75	49.88	47.06	40.55	42.03	**35.06**
MEAN	**2.72**	5.64	6.54	4.28	4.17	3.74
STD	8.20	10.35	9.84	8.55	9.14	**7.02**

Table 3
Segmentation accuracy (%) on the Extended Yale B data set, with different numbers of persons. For robust LatLRR, the parameter λ is set as 0.014, 0.013, and 0.0135. The parameters of other methods are also tuned to be the best.

Experiment Results on the Extended Yale B Data Set	SSC	LRR	RSI	LRSC	LatLRR	Robust LatLRR
5 Subjects	90.31	91.25	89.69	75.94	69.06	**95.94**
7 Subjects	87.05	72.77	89.73	65.63	42.86	**93.86**
9 Subjects	71.35	60.42	87.67	51.04	36.11	**91.49**

rates of the six methods on all the 156 tasks. It shows that our robust LatLRR has a clear advantage in two criteria, maximum error and standard deviation, which shows that the performance of robust LatLRR is much more stable. For the mean error, robust LatLRR ranks the second. Actually, the differences among the six methods are limited. This is mainly because the noises and corruptions in the Hopkins 155 data set are small. As a result, all the state-of-the-art methods perform excellently.

3.3. Real experiment on Extended Yale B data set

The Extended Yale B data set consists of face images of 38 persons. For each person, there are 64 frontal face images in different illuminations. Thus each person corresponds to a subspace [20]. More than half of the face images are corrupted by large area "shadows" and are also noisy. So the data set can be regarded as heavily corrupted.

We test the performance of the six methods by choosing the face images of different numbers of persons in the Extended Yale B data set. Since SSC, LRR, and LatLRR all have a high computation complexity, we scale the original images into a size of 42×48 pixels. Table 3 presents the segmentation accuracies of the six methods with different numbers of persons. It shows that in this test our robust LatLRR significantly outperforms other methods, especially when the number of persons is larger. Moreover, the optimal λ for different numbers of persons, 0.014, 0.013, and 0.0135, changes only slightly. This further shows that our robust LatLRR is insensitive to the choice of λ.

4. Conclusion

This paper aims at addressing the non-unique-solution issue of LatLRR. On the basis of the theoretical analysis in [15], we propose robust LatLRR, which first denoises the data by robust PCA and then chooses the sparsest representation matrix among the solution set of LatLRR with denoised data. Tests on both the synthetic and the real data testify to the robustness of our robust LatLRR when compared with other state-of-the-art methods.

Acknowledgment

The authors thank René Vidal for valuable discussions. Hongyang Zhang and Chao Zhang are supported by National Key Basic Research Project of China (973 Program) 2011CB302400 and National Nature Science Foundation of China (NSFC Grant no. 61071156 and 61131003). Zhouchen Lin is supported by NSF China (Grant nos. 61272341, 61231002, and 61121002) and MSRA. Junbin Gao is supported by the Australian Research Council (ARC) through the Grant DP130100364.

References

[1] R. Vidal, Subspace clustering, IEEE Signal Process. Mag. 28 (2) (2011) 52–68.
[2] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[3] B. Cheng, G. Liu, Z. Huang, S. Yan, Multi-task low-rank affinities pursuit for image segmentation, in: IEEE International Conference on Computer Vision, 2011, pp. 2439–2446.
[4] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, in: International Conference on Machine Learning, vol. 3, 2010, pp. 663–670.
[5] G. Liu, S. Yan, Latent low-rank representation for subspace segmentation and feature extraction, in: IEEE International Conference on Computer Vision, 2011, pp. 1615–1622.
[6] S. Wei, Z. Lin, Analysis and Improvement of Low Rank Representation for Subspace Segmentation, 2010, arXiv:1107.1561.
[7] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2790–2797.
[8] C. Lang, G. Liu, J. Yu, S. Yan, Saliency detection by multi-task sparsity pursuit, IEEE Trans. Image Process. 21 (3) (2012) 1327–1338.
[9] E. Elhamifar, R. Vidal, Clustering disjoint subspaces via sparse representation, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2010, pp. 1926–1929.
[10] L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang, N. Yu, Non-negative low rank and sparse graph for semi-supervised learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2328–2335.
[11] E. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis? J. ACM 58 (3) (2011) 1–37.
[12] P. Favaro, R. Vidal, A. Ravichandran, A closed form solution for robust subspace estimation and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1801–1807.
[13] R. Liu, Z. Lin, F. De La Torre, Z. Su, Fixed-rank representation for unsupervised visual learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 598–605.
[14] G. Bull, J. Gao, Transposed low rank representation for image classification, in: International Conference on Digital Image Computing—Techniques and Applications, 2012, pp. 1–7.
[15] H. Zhang, Z. Lin, C. Zhang, A counterexample for the validity of using nuclear norm as a convex surrogate of rank, in: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2013, pp. 226–241.
[16] T. Zhang, B. Ghanem, S. Liu, N. Ahuja, Low-rank sparse learning for robust visual tracking, in: European Conference on Computer Vision, 2012, pp. 470–484.

[17] Y. Ni, J. Sun, X. Yuan, S. Yan, L. Cheong, Robust low-rank subspace segmentation with semidefinite guarantees, in: IEEE International Conference on Data Mining Workshops, 2010, pp. 1179–1188.
[18] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, Sparse representation for computer vision and pattern recognition, Proc. IEEE 98 (6) (2010) 1031–1044.
[19] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
[20] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

Hongyang Zhang received the Bachelor of Mathematics from China University of Geosciences (CUG), Wuhan, China, in 2012. Currently, he is a Master Degree candidate at the Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University. His research interests include machine learning, computer vision and statistics, especially theoretical analysis for optimization models.

Zhouchen Lin received the Ph.D. degree in Applied Mathematics from Peking University, in 2000. He is currently a Professor at Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University. He is also a Chair Professor at Northeast Normal University and a Guest Professor at Beijing Jiaotong University. Before March 2012, he was a Lead Researcher at Visual Computing Group, Microsoft Research Asia. He was a Guest Professor at Shanghai Jiaotong University and Southeast University, and a Guest Researcher at Institute of Computing Technology, Chinese Academy of Sciences. His research interests include computer vision, image processing, computer graphics, machine learning, pattern recognition, and numerical computation and optimization. He is an Associate Editor of International Journal of Computer Vision and Neurocomputing and a Senior Member of the IEEE.

Chao Zhang received the Ph.D. degree in Electrical Engineering from the Beijing Jiaotong University, Beijing, China, in 1995. From 1995 to 1997, he was a Postdoctoral Research Fellow at the National Laboratory on Machine Perception, Peking University. Since 1997 he has been an Associate Professor at Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University. His research interests include image processing, statistical pattern recognition and visual recognition.

Junbin Gao graduated from Huazhong University of Science and Technology (HUST), China, in 1982, with B. Sc. degree in Computational Mathematics and obtained Ph.D. from Dalian University of Technology, China, in 1991. Currently he is a full Professor in Computer Science at Charles Sturt University, Australia. He was a Senior Lecturer, a Lecturer in Computer Science from 2001 to 2005 at University of New England, Australia. From 1982 to 2001 he was an Associate Lecturer, Lecturer, Associate Professor and Professor in Department of Mathematics at HUST. His main research interests include machine learning, data mining, Bayesian learning and inference, and image analysis.