Robust Out-of-Sample Data Recovery

Bo Jiang[1], Chris Ding[2,1] and **Bin Luo[1]**
[1]School of Computer Science and Technology, Anhui University, Hefei, 230601, China
[2]CSE Department, University of Texas at Arlington, Arlington, TX 76019, USA
jiangbo@ahu.edu.cn, chqding@uta.edu, luobin@ahu.edu.cn

Abstract

Trace norm based rank regularization techniques have been successfully applied to learn a low-rank recovery for high-dimensional noise data. In many applications, it is desirable to add new samples to previously recovered data which is known as out-of-sample data recovery problem. However, traditional trace norm based regularization methods can not naturally cope with new samples and thus fail to deal with out-of-sample data recovery. In this paper, we propose a new robust out-of-sample data recovery (ROSR) model for trace norm based regularization methods. An effective iterative algorithm, with the proof of convergence, is presented to find the optimal solution of ROSR problem. As an application, we apply our ROSR to image classification task. Experimental results on six image datasets demonstrate the effectiveness and benefits of the proposed ROSR method.

1 Introduction

Image recovery and reconstruction is an important research topic in computer vision and machine learning area. There exist many studies on image data recovery. One kind of popular methods is to use matrix factorization techniques [Lee and Seung, 1999; Duda *et al.*, 2001; Aharon *et al.*, 2006] which aim to learn an explicit low-rank recovery/representation for original high-dimensional data. In low-rank space, the noise can be suppressed and the class distribution becomes more apparent which significantly improves the machine learning results. In order to deal with gross errors or outliers, recent works use more robust matrix norms such as ℓ_1-norm [Ke and Kanade, 2005; Kasiviswanathan *et al.*, 2012; Peng *et al.*, 2010; Zhao and Cham, 2011; Zhang *et al.*, 2011; Feng *et al.*, 2013; Yu *et al.*, 2012], ℓ_{21}-norm [Ding *et al.*, 2006; Kwak, 2008; Kong *et al.*, 2011] to develop robust matrix factorization formulations which have been shown to be able to deal with gross errors or outliers effectively.

In additional to matrix factorization methods, trace norm based rank regularization approaches [Cai *et al.*, 2010; Ma *et al.*, 2009; Liu *et al.*, 2010; Liu and Yan, 2011] have also been applied to reduce the rank of the data and thus can perform data recovery robustly. Among them, one of popular methods is Robust Principal Component Analysis (RPCA) [Wright *et al.*, 2009a; Peng *et al.*, 2010; Xu *et al.*, 2012; Ma *et al.*, 2009] which has been successfully applied in many problems such as data reconstruction, image denoising and background modeling, etc. Liu et al., [Liu *et al.*, 2010] proposed a robust subspace segmentation and data recovery method based on Low Rank Representation (LRR). Liu and Yan [Liu and Yan, 2011] extend LRR to Latent Low Rank Representation (LatLRR) which uses both observed and unobserved hidden data and thus provides a more robust representation. Zhang et al., [Zhang *et al.*, 2012] provided a matrix completion method via truncated nuclear norm regularization which can better approximate the matrix rank function. Recently, some other methods have also been proposed [Nie *et al.*, 2014; Kim *et al.*, 2015]. The above rank regularization methods can correctly recover underlying low-rank structure in the data, even in the presence of noise. One main advantage of these trace-norm based approaches is that the trace-norm is the convex envelop of matrix rank and thus the optimization is usually convex. A unique optimal solution exists.

However, one important issue for these trace norm recovery methods is how to do robust recovery for a new incoming 'query/test' sample, which we call *out-of-sample data recovery*. Since trace norm based methods are generally computed in a batch manner, i.e., they process the whole data set X^0 simultaneously and thus can not naturally cope with the new incoming data. When a new query data x is added, one possible way is to re-compute the whole available data $\{x, X^0\}$ simultaneously using traditional trace norm methods and then obtain the optimal recovery for the new data x. Obviously, this is computationally cost and also does not provide a prediction for the new data.

In this paper, we propose a new robust out-of-sample data recovery (ROSR) for trace norm based recovery methods. An effective iterative algorithm with the proof of convergence is presented to find the global optimal solution of the proposed ROSR problem. As an application, we apply our ROSR to face recognition and handwritten character recognition tasks. Experimental results on several datasets demonstrate the effectiveness and benefits of the proposed ROSR method.

2 Brief Review of Low Rank Recovery

Let the input data matrix $X = (x_1, ..., x_n) \in \mathbb{R}^{p\times n}$ contains the collection of n data vectors in p dimension space.

In image processing, each column x_i is the linearized array of pixels gray levels. In many applications, the input data X usually contains noises, i.e., $X = Z + E$, where Z is the true signal data, and E is the sparse distortion. The aim of low-rank recovery is to find the optimal recovery Z by minimizing

$$\min_{Z} \ \|X - Z\|_\ell + \lambda \mathrm{rank}(Z), \tag{1}$$

where $\lambda > 0$ is a parameter and $\|\cdot\|_\ell$ denotes the certain norms, such as Frobenius norm, $\ell_{2,1}$ and ℓ_1-norm. The explicit rank constraint $\mathrm{rank}(Z)$ is difficult to enforce in numerical computation. Many works suggest to use trace norm $\|Z\|_{tr}$ to approximate the rank constraint $\mathrm{rank}(Z)$. Thus, the above low-rank recovery becomes,

$$\min_{Z} \ \|X - Z\|_\ell + \lambda \|Z\|_{tr}, \tag{2}$$

where $\|Z\|_{tr} = \mathrm{Tr}(ZZ^T)^{1/2}$. When $\|\cdot\|_\ell$ denotes $\|\cdot\|_1$ norm, this model refers to robust PCA [Wright *et al.*, 2009a; Candès *et al.*, 2011] which has been shown robustly in dealing with corruption/sparse noise. When $\|\cdot\|_\ell$ denotes $\|\cdot\|_{2,1}$ norm, this model becomes to outlier pursuit model [Xu *et al.*, 2012] which is better to pursuit the outliers in data recovery.

The above rank regularization methods can correctly recover underlying low-rank structure in the data, even in the presence of noise or outliers. Since trace norm is convex, the above optimization (Eq.(2)) is convex and a unique optimal solution exists.

3 Robust Out-of-Sample Recovery

Suppose we have obtained the optimal low-rank recovery $Z^0 \in \mathbb{R}^{p\times n}$ from original observed data $X^0 \in \mathbb{R}^{p\times n}$ using the above RPCA model. Now, we consider a new query/test image data $x \in \mathbb{R}^{p\times 1}$ in original observed space, and wish to obtain its recovery/representation z in the low-rank space. Our aim in this paper is to seek a way to obtain z for x while fixing the already learned variable Z^0. This can be formulated by solving the following model [1],

$$\min_{z} \ \|[X^0, x] - [Z^0, z]\|_1 + \beta \mathrm{rank}([Z^0, z]). \tag{3}$$

Since both X^0 and Z^0 are fixed, this problem can be reduced as,

$$\min_{z} \ \|x - z\|_1 + \beta \mathrm{rank}([Z^0, z]). \tag{4}$$

Using trace norm $\|\cdot\|_{tr}$ replacement, this problem becomes

$$\min_{z} \ \|x - z\|_1 + \beta \|[Z^0, z]\|_{tr}. \tag{5}$$

Let z^* be the optimal solution of model Eq.(5), then z^* can be regarded as a kind of recovery/reconstruction for the new query data x. In this paper, we call it Robust Out-of-Sample Recovery (ROSR). There is no new parameter in ROSR: β is set to the value when Z^0 are obtained.

Illustration. To illustrate the data recovery/reconstruction ability of ROSR model, we run ROSR on the occluded images selected from AT&T face dataset (more details are given in the Experiments section). Here, we select 8 images of each person for training and the rest 2 images for testing. Figure 1 (left) shows some examples of the occluded images X^0 and corresponding low-rank reconstruction Z^0 of RPCA. Figure 1 (right) shows the occluded test image x and corresponding recovery z^* of ROSR. Each panel shows a test image with 3 different corruptions. Here we can note that large errors (occlusions) are suppressed in ROSR recovery. This clearly demonstrates the robust of ROSR method for out-of-sample data recovery.

[1] Here we focus on ℓ_1. $\ell_{2,1}$ is similarly obtained.

4 Optimization

In this section, we derive an effective update algorithm to solve the proposed ROSR problem and provide the theoretical analysis on the convergence of the algorithm.

4.1 Computational algorithm

Given any initial $z^{(0)}$, the algorithm updates the solution $z^{(t)}, t = 1, 2...$ until convergence as summarized in Algorithm 1. Since the objective function of ROSR problem (Eq.(5)) is convex, thus staring from any initialization, the algorithm converges to the global optimal solution.

Algorithm 1 Robust Out-of-Sample Recovery

1: **Input:** Training data $X^0, Z^0 \in \mathbb{R}^{p\times n}$, query/test data $x \in \mathbb{R}^p$, parameters β, maximum number of iteration T_{max}, and convergence tolerance $\epsilon > 0$;
2: Initialize $z^{(0)} = \mathbf{0}, t = 0$,
3: **while** $t < T_{max}$ or $\frac{\|z^{(t+1)} - z^{(t)}\|_2}{\|z^{(t)}\|_2} > \epsilon$ **do**
4: Compute $w^{(t)}$ and matrix $B^{(t)}$ as

$$w^{(t)} = |x - z^{(t)}| \tag{6}$$

$$B^{(t)} = \left([Z^0, z^{(t)}][Z^0, z^{(t)}]^T\right)^{1/2} \tag{7}$$

5: Update $z^{(t+1)}$ as

$$z^{(t+1)} = B^{(t)}(B^{(t)} + \beta D^{(t)} B^{(t)})^{-1} x \tag{8}$$

where $D^{(t)} = \mathrm{diag}(w_1^{(t)}, w_2^{(t)}, \cdots w_p^{(t)})$.
6: **end while**
7: **Output:** The converged recovery $z^* = z^{(t+1)}$.

4.2 Convergence analysis

Let $L(z)$ denote the objective function of ROSR problem Eq.(5). Since $L(z)$ is a convex function of z, thus we only need to prove that the objective function value $L(z)$ is non-increasing in each iteration of Algorithm 1. This is summarized in Theorem 1.

Theorem 1 *The objective function value $L(z)$ of problem Eq.(5) is non-increasing,*

$$L(z^{(t+1)}) \le L(z^{(t)}), \tag{9}$$

along with each iteration of the update rule Eq.(8) in Algorithm 1.

To prove the Theorem 1, we need to prove the following three Lemmas firstly.

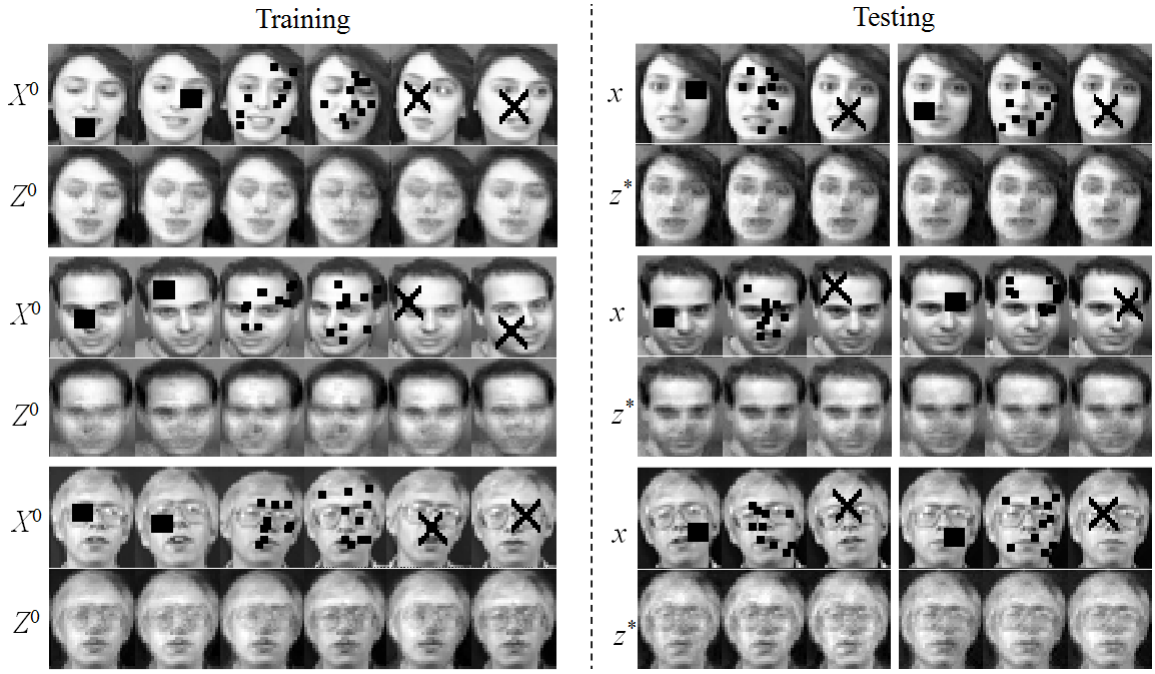

Figure 1: Image recovery of ROSR on occluded AT&T face data. LEFT: examples of the occluded images X^0 and corresponding RPCA recovery Z^0. RIGHT: test image x and corresponding ROSR recovery z^*.

Lemma 2 *For any matrices Y, Z with the same size, let $A = (YY^T)^{1/2}, B = (ZZ^T)^{1/2}$. Then, the following inequality holds,*

$$\mathrm{Tr}(A) \leq \mathrm{Tr}(B) - \frac{1}{2}\mathrm{Tr}(Z^T B^{-1} Z) + \frac{1}{2}\mathrm{Tr}(Y^T B^{-1} Y) \quad (10)$$

The detail proof of this property can refer to the work [Luo *et al.*, 2011].

Lemma 3 *Define an auxiliary function*

$$G(z) = \sum_i \frac{(x-z)_i^2}{2|x - z^{(t)}|_i} + \frac{\beta}{2}\mathrm{Tr}\big([Z^0, z]^T {B^{(t)}}^{-1} [Z^0, z]\big) \quad (11)$$

where $B^{(t)} = \big([Z^0, z^{(t)}][Z^0, z^{(t)}]^T\big)^{\frac{1}{2}}$. Along with the $\{z^{(t)}, t = 0, 1, \cdots\}$ sequence of the update rule Eq.(8) in Algorithm 1, the following inequality holds,

$$G(z^{(t+1)}) \leq G(z^{(t)}).$$

Proof Since the two terms in auxiliary function $G(z)$ are semi-definite positive (SDP) problems, we can obtain the global optimal solution of $G(z)$ by taking the derivatives and let them equal to zero.

Take the derivative of $G(z)$ with respect to z, and we get

$$\frac{\partial G(z)}{\partial z_i} = -\frac{(x-z)_i}{w_i^{(t)}} + \beta({B^{(t)}}^{-1} z)_i \quad (12)$$

where $w_i^{(t)} = |x - z^{(t)}|_i$.

By setting $\dfrac{\partial G(z)}{\partial z_i} = 0$, we have,

$$\frac{z_i}{w_i^{(t)}} + \beta({B^{(t)}}^{-1} z)_i = \frac{x_i}{w_i^{(t)}}. \quad (13)$$

Let $D^{(t)} \in \mathbb{R}^{p \times p}$ be the diagonal matrix with $D_{ii}^{(t)} = w_i^{(t)}$, then, Eq.(13) can be reformulated as follows,

$${D^{(t)}}^{-1} z + \beta {B^{(t)}}^{-1} z = {D^{(t)}}^{-1} x. \quad (14)$$

Thus, we have

$$\begin{aligned} z &= \big({D^{(t)}}^{-1} + \beta {B^{(t)}}^{-1}\big)^{-1} {D^{(t)}}^{-1} x \\ &= B^{(t)}\big(B^{(t)} + \beta D^{(t)}\big)^{-1} x \end{aligned} \quad (15)$$

This completes the proof.

Lemma 4 *The $\{z^{(t)}, t = 0, 1, 2, \cdots\}$ sequence obtained by update rule Eq.(8) in Algorithm 1 has the following property,*

$$L(z^{(t+1)}) - L(z^{(t)}) \leq G(z^{(t+1)}) - G(z^{(t)}). \quad (16)$$

Proof First, we rewrite problem Eq.(5) as,

$$\min_z \; L(z) = \|x - z\|_1 + \beta \mathrm{Tr}\big([Z^0, z][Z^0, z]^T\big)^{1/2} \quad (17)$$

Let $\Delta = \big(L(z^{(t+1)}) - L(z^{(t)})\big) - \big(G(z^{(t+1)}) - G(z^{(t)})\big)$. Then, substitute Eq.(17) and Eq.(11) in it, we obtain

$$\begin{aligned} \Delta = &\big(\|x - z^{(t+1)}\|_1 - \|x - z^{(t)}\|_1\big) \\ &- \Big[\sum_i \frac{(x - z^{(t+1)})_i^2}{2|x - z^{(t)}|_i} - \sum_i \frac{(x - z^{(t)})_i^2}{2|x - z^{(t)}|_i}\Big] \\ &+ \beta \mathrm{Tr}\big([Z^0, z^{(t+1)}][Z^0, z^{(t+1)}]^T\big)^{\frac{1}{2}} \\ &- \beta \mathrm{Tr}\big([Z^0, z^{(t+1)}][Z^0, z^{(t)}]^T\big)^{\frac{1}{2}} \\ &- \frac{\beta}{2}\mathrm{Tr}\Big([Z^0, z^{(t+1)}]^T {B^{(t)}}^{-1} [Z^0, z^{(t+1)}]\Big) \\ &+ \frac{\beta}{2}\mathrm{Tr}\Big([Z^0, z^{(t)}]^T {B^{(t)}}^{-1} [Z^0, z^{(t)}]\Big). \end{aligned} \quad (18)$$

Eq.(18) can be rewritten as

$$\begin{aligned}\Delta = &-\sum_i \frac{1}{2|x-z^{(t)}|_i}\big(|x-z^{(t+1)}|_i - |x-z^{(t)}|_i\big)^2 \\ &+ \beta \mathrm{Tr}\big([Z^0, z^{(t+1)}][Z^0, z^{(t+1)}]^T\big)^{\frac{1}{2}} \\ &- \beta \mathrm{Tr}\big([Z^0, z^{(t+1)}][Z^0, z^{(t)}]^T\big)^{\frac{1}{2}} \\ &- \frac{\beta}{2}\mathrm{Tr}\Big([Z^0, z^{(t+1)}]^T {B^{(t)}}^{-1}[Z^0, z^{(t+1)}]\Big) \\ &+ \frac{\beta}{2}\mathrm{Tr}\Big([Z^0, z^{(t)}]^T {B^{(t)}}^{-1}[Z^0, z^{(t)}]\Big). \end{aligned} \quad (19)$$

According to Lemma 2, we have

$$\begin{aligned}&\mathrm{Tr}\big([Z^0, z^{(t+1)}][Z^0, z^{(t+1)}]^T\big)^{\frac{1}{2}} \\ &\le \mathrm{Tr}\big([Z^0, z^{(t)}][Z^0, z^{(t)}]^T\big)^{\frac{1}{2}} \\ &+ \frac{1}{2}\mathrm{Tr}\big([Z^0, z^{(t+1)}]^T {B^{(t)}}^{-1}[Z^0, z^{(t+1)}]\big)^{\frac{1}{2}} \\ &- \frac{1}{2}\mathrm{Tr}\big([Z^0, z^{(t)}]^T {B^{(t)}}^{-1}[Z^0, z^{(t)}]\big)^{\frac{1}{2}} \end{aligned} \quad (20)$$

where $B^{(t)} = \big([Z^0, z^{(t)}][Z^0, z^{(t)}]^T\big)^{1/2}$. Substitute Eq.(20) into Eq.(19), we can have $\Delta \le 0$. This completes the proof of Lemma 4.

From Lemma 3 and Lemma 4, we have,

$$L(z^{(t+1)}) - L(z^{(t)}) \le G(z^{(t+1)}) - G(z^{(t)}) \le 0, \quad (21)$$

which is to say

$$L(z^{(t+1)}) \le L(z^{(t)}). \quad (22)$$

This completes the proof of Theorem 1.

Figure 2 shows the variation of objective function value $L(z^{(t)})$ across each iteration starting from random initialization. We can note that, as the iteration increases, the objective function $L(z^{(t)})$ decreases monotonously, demonstrating the convergence of the algorithm. As the iteration increases, the algorithm recoveries the true image more and more clearly. Also, the algorithm converges fast, generally in 20-30 iterations.

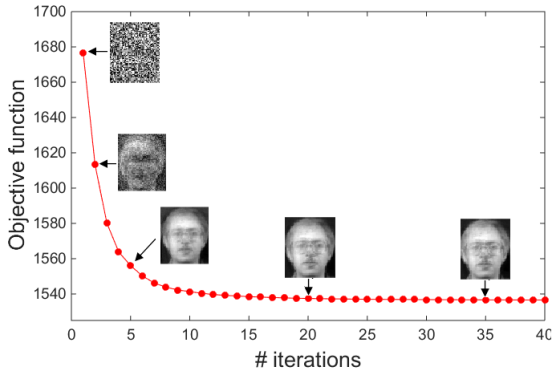

Figure 2: Variation of objective function value $L(z^{(t)})$ across each iteration

5 Experiments

5.1 Image recognition using ROSR

As an application for the proposed ROSR model, we have applied it to achieve face recognition and handwritten character recognition tasks. In short, the recognition process has the following two main steps.

In training phase, we obtain the low-rank recovery Z^0 from input data X^0 using RPCA. In testing phase, for each test image x, we first obtain its ROSR recovery z^* and then identify it by K Nearest Neighbor (KNN) classifier or Sparse Representation Classification (SRC) [Wright *et al.*, 2009b] methods. In KNN classification, the class identity of z^* is determined by its K nearest neighbors in Z^0 based on Euclidean distances. In SRC [Wright *et al.*, 2009b] classification, we first compute the optimal coefficient $\hat{\alpha}$ as

$$\hat{\alpha} = \mathrm{argmin}_\alpha \|\alpha\|_1 \ \ s.t. \ \ \|z^* - Z^0\alpha\|_2 \le \varepsilon.$$

Then, we obtain the class identity of z^* as

$$\mathrm{Identity}(z^*) = \mathrm{argmin}_i \|z^* - Z_i^0 \hat{\alpha}_i\|_2.$$

where Z_i^0 denotes the dataset of the i-th class, and $\hat{\alpha}_i$ is the coding coefficient vector associated with the i-th class.

5.2 Datasets description

Six image datasets are used in the experiments, including three face datasets (AT&T face databases, Extended Yale Database B [Lee *et al.*, 2005] and CMU-PIE [He *et al.*, 2005a]) and three handwritten character datasets (Binary Alphabet Dataset [2], MNIST handwritten digits Database, USPS Handwritten Digits Dataset[3]). The details are introduced below and Table 1 summarizes them.

AT&T Faces Dataset contains 40 distinct persons with each person/class contains 10 different images. In our experiment, each face image was resized into 23× 28 and reshaped into a vector of 644 dimension.

Extended Yale Database B consists of 38 different classes with each class contains about 64 frontal face images. In our experiment, each face image was resized into 28×32 and reshaped into a vector of 896 dimension.

CMU PIE face Database contains 68 different classes with 41368 face images as a whole. In our experiment, we randomly select 30 classes with each class containing 50 face images. Each face image was resized into 32×32 and reshaped into a 1024 dimension vector.

MNIST Hand-written Digit Dataset consists of 8-bit gray-scale images of digits from "0" to "9", about 6000 examples of each class (digit). Each image is centered on a 28×28 grid. Here, we randomly select 100 images from each digit, and convert them to vectors of 784 dimension.

USPS Dataset contains 9298 handwritten digit images from "0" to "9". Each image is centered on a 16×16 grid. In our experiment, we select 1000 images (100 images for every digit) and convert them to vectors of 256 dimension.

Binary Alphabet Dataset contains 26 hand-written alphabets from A∼Z with each alphabet consists 39 samples. Each sample is a 20×16 binary image. We reshape each image into one vector of 320 dimension.

[2] http://olivier.chapelle.cc/ssl-book/benchmarks.html

[3] http://www.cs.nyu.edu/ roweis/data.html

Table 1: Dataset descriptions.

Dataset	# Size	# Dimension	# Class
AT&T	400	644	40
Yale-B	2414	896	32
CMU PIE	1500	1024	30
MNIST	1000	784	10
USPS	1000	256	10
Bin-Alpha	1014	320	26

5.3 Results

To evaluate the robustness of our method, we conduct classification experiments on the corrupted images. Here, we randomly add corruptions on each image of the dataset, as shown in Figure 1. The percentage of corruption is about 10% of image size. All experiments are performed with ten-fold cross validation strategy, i.e., all data sets are randomly splitted into ten equal subsets, iteratively pick one subset for testing and the remaining nine subsets for training, then the performances are averaged over the ten loops.

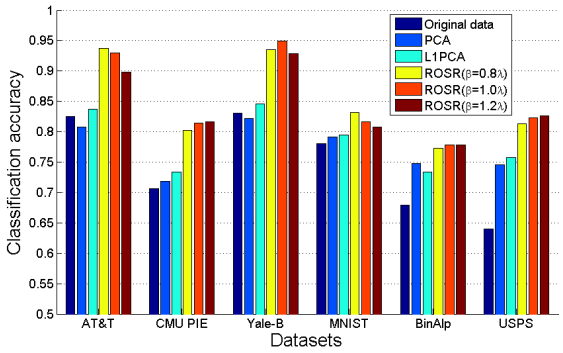

Figure 3: Comparison of classification results using SRC classification

For comparison, we compared our model with original raw data and other data recovery methods including standard Principal Component Analysis (PCA) [Duda *et al.*, 2001], Locality Preserving Projection (LPP) [He *et al.*, 2005a], Neighborhood Preserving Embedding (NPE) [He *et al.*, 2005b] and L1-norm based PCA (L1PCA)[Ke and Kanade, 2005; Yu *et al.*, 2012]. These compared methods can be used for out-of-sample data recovery or representation. For LPP and NPE, we first learn the optimal projection P^0. Then, for a new query data x, we use P^0x as its representation. For L1PCA, we first learn the optimal U^0 from input data X^0 by solving,

$$\{U^0, V^0\} = \text{argmin}_{U,V}\|X^0 - UV\|_1. \quad (23)$$

Then, for a new query data x, we obtain its recovery as,

$$v^* = \text{argmin}_v\|x - U^0v\|_1. \quad (24)$$

At last, we use U^0v^* as the recovery for test image x, i.e.,

$$x \simeq U^0v^*. \quad (25)$$

This is similar to PCA. We set the regularization parameter β in ROSR to $0.8\lambda_0$, $1.0\lambda_0$ and $1.2\lambda_0$, respectively, where $\lambda_0 = \sqrt{p}$ and p is the dimension of data. Note that $\lambda = \lambda_0$ is used in RPCA model in the experiment. For PCA, LPP, NPE and L1PCA, we set the dimension d to 50 and 100, respectively.

Table 2 shows the comparison results of different recovery methods using KNN classification. We can observe that (1) L1PCA generally outperforms standard PCA, indicating the robustness of L1PCA recovery method. (2) ROSR performs better than L1PCA and gives the best classification results, which clearly demonstrates that ROSR can recover the noise images effectively and thus leads to better classification results. Figure 3 shows the comparison results of different recovery methods using SRC classification [Wright *et al.*, 2009b]. Here, we perform SRC on the recovery images of different methods. We compare our ROSR method with original raw data, PCA and L1PCA method, because these methods can return a recovery for out-of-sample image, as shown in Eq.(25). We can note that our ROSR method obviously outperforms other methods, which further demonstrates the robustness of ROSR method on recovering the occluded images and thus leads to better recognition results.

6 Conclusions

This paper proposes a novel Robust Out-of-Sample Recovery (ROSR) method which conducts robust recovery for out-of-sample data naturally based on trace norm regularization. An effective algorithm, with theoretical analysis on convergence, has been developed to solve ROSR problem. As an application, we apply our ROSR to image classification task. Experimental results demonstrate that ROSR is effective in recovering noise image data and thus obviously improves the classification results.

Acknowledgments

This work was supported by the National Key Basic Research Program of China (973 Program) (2015CB351705); National Nature Science Foundation of China (61472002); Co-Innovation Center for Information Supply & Assurance Technology, Anhui University; the Open Projects Program of National Laboratory of Pattern Recognition.

References

[Aharon *et al.*, 2006] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transaction on Signal Processing*, 54(11):4311–4322, 2006.

[Cai *et al.*, 2010] J. F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. on Optimization*, 20(4):1956–1982, 2010.

[Candès *et al.*, 2011] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis ? *Journal of the ACM*, 58(3), 2011.

[Ding *et al.*, 2006] C. Ding, D. Zhou, X. He, and H. Zha. R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. In *ICML*, pages 281–288, 2006.

Table 2: Comparison of classification results using KNN classification

1		Original	PCA		LPP		NPE		L1PCA		ROSR		
			d= 50	d= 100	d = 50	d= 100	d= 50	d = 100	d = 50	d = 100	$\beta = 0.8\lambda_0$	$\beta = \lambda_0$	$\beta = 1.2\lambda_0$
AT&T	1-NN	0.5500	0.5725	0.5725	0.2850	0.3425	0.3300	0.3250	0.6875	0.6225	0.8200	0.8475	0.8425
	5-NN	0.6000	0.6025	0.6200	0.2950	0.3525	0.3250	0.3600	0.6975	0.6550	0.8000	0.8225	0.8225
PIE	1-NN	0.5867	0.4853	0.5273	0.5487	0.6300	0.5733	0.6653	0.5253	0.5587	0.6280	0.6587	0.7047
	5-NN	0.5867	0.5107	0.5473	0.5587	0.6467	0.5827	0.6773	0.5200	0.5913	0.7287	0.7360	0.7387
Yale-B	1-NN	0.6343	0.5216	0.6497	0.5278	0.6636	0.5548	0.6952	0.5841	0.6852	0.7793	0.7840	0.7469
	5-NN	0.6991	0.5602	0.6890	0.5571	0.7014	0.5741	0.7207	0.6334	0.7207	0.8218	0.8194	0.7658
MNIST	1-NN	0.7700	0.7650	0.7560	0.6230	0.6030	0.6120	0.5920	0.7940	0.7890	0.8460	0.8520	0.8380
	3-NN	0.7900	0.7900	0.7890	0.6460	0.6370	0.6360	0.6210	0.8280	0.7990	0.8560	0.8510	0.8500
BinAlp	1-NN	0.7308	0.7731	0.7615	0.6897	0.6385	0.6936	0.7051	0.7641	0.7782	0.7718	0.7731	0.7897
	3-NN	0.7692	0.7923	0.7962	0.7077	0.6603	0.7244	0.7051	0.7846	0.7974	0.8013	0.8013	0.7821
USPS	1-NN	0.6700	0.7510	0.7320	0.6270	0.5400	0.5890	0.5390	0.7930	0.7630	0.7950	0.8170	0.8330
	3-NN	0.6700	0.7700	0.7500	0.6350	0.5610	0.6230	0.5650	0.7990	0.7830	0.8110	0.8300	0.8330

[Duda *et al.*, 2001] R.O. Duda, P.E. Hart, and G.D. Stork, editors. *Pattern Classification (2nd ed)*. Wiley Interscience, New York, 2001.

[Feng *et al.*, 2013] J. Feng, H. Yu, and S. Yan. Online robust pca via stochastic optimization. In *NIPS*, pages 404–412, 2013.

[He *et al.*, 2005a] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *IEEE Transaction on PAMI*, 27(3):328–340, 2005.

[He *et al.*, 2005b] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang. Neighborhood preserving embedding. In *ICCV*, pages 1208–1213, 2005.

[Kasiviswanathan *et al.*, 2012] S. P. Kasiviswanathan, H. Wang, A. Banerjee, and P. Melville. Online l1-dictionary learning with application to novel document detection. In *NIPS*, pages 2267–2275, 2012.

[Ke and Kanade, 2005] Q. Ke and T. Kanade. Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *CVPR*, pages 739–746, 2005.

[Kim *et al.*, 2015] E. Kim, M. Lee, and S. Oh. Elastic-net regularization of singular values for robust subspace learning. In *CVPR*, pages 915–923, 2015.

[Kong *et al.*, 2011] D. Kong, C. Ding, and H. Huang. Robust nonnegative matrix factorization using l21-norm. In *CIKM*, pages 673–682, 2011.

[Kwak, 2008] N. Kwak. Principal component analysis based on l1-norm maximization. *IEEE Transaction on PAMI*, 30(9):1672–1680, 2008.

[Lee and Seung, 1999] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[Lee *et al.*, 2005] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transaction on PAMI*, 27(5):684–698, 2005.

[Liu and Yan, 2011] G. Liu and S. Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *ICCV*, pages 1615–1622, 2011.

[Liu *et al.*, 2010] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. *ICML*, pages 663–670, 2010.

[Luo *et al.*, 2011] D. Luo, F. Nie, C. Ding, and H. Huang. Multi-subspace representation and discovery. In *ECML PKDD*, pages 405–420, 2011.

[Ma *et al.*, 2009] S. Ma, D. Goldfarb, and L. Chen. Fixed point and bregman iterative methods for matrix rank minimization. *CoRR*, abs/0905.1643, 2009.

[Nie *et al.*, 2014] F. Nie, J. Yuan, and H. Huang. Optimal mean robust pricipal component analysis. In *ICML*, pages 1062–1070, 2014.

[Peng *et al.*, 2010] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low rank decomposition for linearly correlated images. In *CVPR*, pages 763–770, 2010.

[Wright *et al.*, 2009a] J. Wright, A. Ganesh, S. Rao, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *NIPS*, pages 2080–2088, 2009.

[Wright *et al.*, 2009b] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sprare representation. *IEEE Transaction on PAMI*, 31(2):210–227, 2009.

[Xu *et al.*, 2012] H. Xu, C. Caramanis, and S. Sanghavi. Robust pca via outlier pursuit. *IEEE Transaction on Information Theorey*, 58(5):3047–3064, 2012.

[Yu *et al.*, 2012] Linbin Yu, Miao Zhang, and Chris Ding. An efficient algorithm for l1-norm principal component analysis. In *ICASSP*, pages 1377–1380, 2012.

[Zhang *et al.*, 2011] L. Zhang, Z. Chen, M. Zheng, and X. He. Robust nonnegative matrix factorization. *Frontiers of Electrical and Electronic Engineering in China*, 6(2):192–200, 2011.

[Zhang *et al.*, 2012] D. Zhang, Y. Hu, J. Ye, X. Li, and X. He. Matrix completion by truncated nuclear norm regularization. In *CVPR*, pages 2192–2199, 2012.

[Zhao and Cham, 2011] C. Zhao and X. G. Wang W. K. Cham. Background subtraction via robust dictionary learning. *EURASIP Journal on Image and Video Processing*, 2011.