Low-Rank Approximation via the Generalized Reweighted Iterative Nuclear Norm

1st Yan Huang
State Key Laboratory of Millimeter Waves
Southeast University,
Nanjing, China
yellowstone0636@hotmail.com

2nd Lan Lan
National Lab of Radar Signal Processing
Xidian University
Xi'an, China
lanlan_xidian@foxmail.com

3rd Lei Zhang
School of Electronics Engineering,
Sun Yat-Sen University
Guangzhou, China
zhanglei_0330@126.com

***Abstract*—The low-rank approximation problem has recently attracted wide concern due to its excellent performance in real-world applications such as image restoration, traffic monitoring, and face recognition. Compared with the classic nuclear norm, the Schatten-p norm is stated to be a closer approximation to restrain the singular values for practical applications in the real world. However, Schatten-p norm minimization is a challenging non-convex, non-smooth, and non-Lipschitz problem. In this paper, inspired by the reweighted ℓ_1 norm for compressive sensing, the generalized iterative reweighted nuclear norm (GIRNN) algorithm is proposed to approximate Schatten-p norm minimization. By involving the proposed algorithms, the problem becomes more tractable and the closed solutions are derived from the iteratively reweighted subproblems. Numerical experiments for the practical matrix completion (MC) problem and robust principal component analysis (RPCA) problem are illustrated to validate the superior performance of both algorithms over some common state-of-the-art methods.**

***Index Terms*—Low-rank approximation problem, matrix completion (MC), robust principal component analysis (RPCA), generalized iterative reweighted nuclear norm (GIRNN)**

I. Introduction

In recent years, low-rank approximation [1]–[3] has been widely used in real-world applications. For example, the image of a natural scene can be considered as a low-rank matrix, and we can exploit the low-rank property for image restoration [1]–[3]. In additional, clusters of human facial images can be used to reconstruct and classify numerical face datasets for face recognition [4], [5]. Also, the video frames captured by a surveillance camera are obviously low rank, so the foreground (e.g., moving object) can be detected from the background environment [6]. The famous Netflix collaborative filtering dataset, which contains over 100 million ratings on more than 10,000 movies, is believed to be low rank due to the fact that most customers' ratings in this big dataset are affected by a few common factors [7].

The low-rank approximation problem can be relaxed to the sparse recovery problem of singular values. However, it is difficult to solve the low-rank approximation problem directly since the rank function minimization is NP hard. Usually, we relax the rank minimization function to a convex problem [4], [8]. But the Frobenius norm needs to be restrained under the low-rank matrix subspace. In consideration of the shortcoming of the nuclear norm, a weighted nuclear norm minimization (WNNM) algorithm was proposed to approximate the rank function, with excellent performance on image denoising [1]. However, low-rank approximation for real natural scenes is a little different from the sparse recovery problem. An augmented Lagrange multiplier (ALM) method is proposed with the Newton method to search for the optimal solution [14]. However, the method is time-consuming.

Inspired by the iteratively reweighted ℓ_1 norm scheme, in this paper, we propose an iteratively reweighted algorithm, the generalized iterative reweighted nuclear norm (GIRNN). By strict formulations and using the well-known von Neumann's trace inequality, we relax the proposed weighted low-rank approximation problem of the observed matrix to the weighted sparse recovery problem of the singular values. Then the closed solutions of the iterative algorithms are given with careful derivations. Finally, we apply the proposed methods to the matrix completion (MC) problem and robust principal component analysis (RPCA) with real world datasets to validate their outperforming performance compared with some state-of-the-art algorithms.

II. Problem Formulation and Preliminary Works

As introduced in Section I, the low-rank approximation problem aims to recover a low-rank matrix from the observed dataset. It is first formulated with the rank function, that is,

$$\begin{aligned} &\min_{\mathbf{X}} \ \operatorname{rank}(\mathbf{X}) \\ &\text{s.t.} \quad \|\mathbf{Y}-\mathbf{X}\|_F^2 < \delta, \end{aligned} \tag{1}$$

where $\operatorname{rank}(\cdot)$ is the rank function, $\|\cdot\|_F$ is the Frobineus norm, δ is the bound of Gaussian noise power, and $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$ are the low-rank matrix and observed matrix, respectively. The rank function is a nonconvex problem, and the optimization problem in (1) is an NP-hard problem. Most previous studies relaxed the rank operation to the nuclear norm, which is a convex function that restricts the low rank of the matrix. Then the nuclear norm minimization (NNM) problem is formulated as

$$\begin{aligned} &\min_{\mathbf{X}} \ \|\mathbf{X}\|_* \\ &\text{s.t.} \quad \|\mathbf{Y}-\mathbf{X}\|_F^2 < \delta, \end{aligned} \tag{2}$$

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

where $\|\cdot\|_*$ is the nuclear norm. The NNM problem can be rewritten as the unconstraint optimization problem, that is,

$$\min_{\mathbf{X}} \ \lambda\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2, \tag{3}$$

where λ is a positive user parameter. The problem in (3) is easy to solve by SVT [8] with the global solution. For another perspective, the nuclear norm is actually the ℓ_1 norm of the singular values, i.e., $\|\mathbf{X}\|_*=\|\sigma(\mathbf{X})\|_1$, where $\sigma(\mathbf{X})$ is the singular value vector of $\mathbf{X}$, and $\|\cdot\|_1$ is the ℓ_1 norm. Viewed this way, the low-rank minimization problem is the sparse recovery problem for the singular values. For natural images, the large singular values stay in the dominant position and need to be protected because the noise cannot affect the large singular values easily [1]. Also the ℓ_1 norm cannot guarantee the sparse solution perfectly when the entries are very different from each other. Therefore, it is necessary to approximate the low-rank objective function in another precise way.

III. Reweighted Low-rank Scheme

Since the low-rank approximation of a matrix is equal to the sparse recovery of its singular values, our goal of low-rank approximation can be reformulated to finding the true sparse solution of the singular values. In some previous studies, the reweighted norm is proposed to recover the real sparsity of vector [11]. However, the above reweighted method cannot be exploited directly for low-rank approximation. First, in [11], the weighted ℓ_1 norm minimization problem is formulated as

$$\min_{\mathbf{x}} \ \lambda\sum_{i=1}^{l} w_i x_i + \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2, \tag{4}$$

where $\mathbf{x},\mathbf{y} \in \mathbb{R}^l$ and w_i is the weight of the ith entry of $\mathbf{x}$. Hence similar to (4), the reweighted nuclear norm minimization problem can be cast as

$$\min_{\mathbf{X}} \ \lambda \sum_{i=1}^{\min\{m,n\}} w_i\sigma_i(\mathbf{X}) + \ \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2. \tag{5}$$

where note that the singular values are in a non-increasing order, i.e., $\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq ... \geq \sigma_{\min\{m,n\}}(\mathbf{X})$. And we apply this order notation for all the following denotations of singular values. For the weighted nuclear norm problem in (5), we can get the closed form solution according to the weighted singular value thresholding [16], [17].

$$\mathbf{X} = \mathbf{U}S_{\lambda w}(\mathbf{\Sigma}_{\mathbf{Y}})\mathbf{V}^T, \tag{6}$$

It is critical to emphasize again that the weights should be in a non-descending order to make the solution feasible. Therefore, the weighted nuclear norm is able to reconstruct the real rank of the matrix, i.e., the ℓ_0 norm of the singular value vector, with the weights $w_i = c(\sigma_i(\mathbf{X})+\varepsilon)^{-1}$, where c is a positive constant, and ε is a sufficiently small positive number, such as 10^{-16}, to avoid dividing by zero when the corresponding singular value is zero. By introducing the above weight, an iterative reweighted method can be exploited to solve the problem. When the solution reaches convergence, the objective function approaches the ℓ_0 norm of the singular value vector, i.e., the rank of the matrix. This method is widely known as the weighted nuclear norm minimization (WNNM) method [2]. Inspired by the weighted ℓ_1 norm for compressive sensing, we propose a generalized iterative reweighted nuclear norm (GIRNN) method:

$$\mathbf{X}^{(k)} = \mathbf{U}S_{\lambda w^{(k-1)}}(\mathbf{\Sigma}_{\mathbf{Y}})\mathbf{V}^T, \tag{7}$$

$$w_i^{(k)} = c\Big(\sigma_i\Big(\mathbf{X}^{(k)}\Big)+\varepsilon\Big)^{p-1}, \tag{8}$$

where $0 \leq p < 1$, the superscript (k) denotes the kth iteration, and $k \geq 1$. The initial value of weights can be assigned by the singular values of observed matrix $\mathbf{Y}$, i.e., $w_i^{(0)} = c(\sigma_i(\mathbf{Y})+\varepsilon)^{p-1}$. When $p=0$, the weight is degenerated as that of the WNNM. The updated rule of the GIRNN method can be solved by alternatively optimizing the optimal $\mathbf{X}$ and updating the weight until the convergence.

IV. Numerical Experiments

In this section, we evaluate the effectiveness of our proposed GIRNN algorithm for solving low-rank approximation problems, like the matrix completion (MC) problem and robust principal component analysis (RPCA) problem. We also compare our proposed algorithm with several state-of-the-art methods, each of which represents a kind of low-rank approximation problem.

A. Matrix Completion (MC) Problem

The MC problem seeks to recover the low-rank matrix from a corrupted signal. The MC problem of the proposed GIRNN model can be formulated as

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{S}} \ \lambda \sum_{i=1}^{\min\{m,n\}} w_i\sigma_i(\mathbf{X}) \\ &\text{s.t.} \ \ P_{\mathbf{\Omega}}(\mathbf{Y}) = P_{\mathbf{\Omega}}(\mathbf{X}) \end{aligned} \tag{9}$$

where $\mathbf{\Omega}$ is a binary support indicator matrix of the same size as $\mathbf{Y}$, and zeros in $\mathbf{\Omega}$ indicate missing entries in the observation matrix. $\mathrm{P}_{\mathbf{\Omega}}(\mathbf{Y}) = \mathbf{\Omega}\odot\mathbf{Y}$ is the element-wise matrix multiplication (Hardamard product) between the support matrix $\mathbf{\Omega}$ and the variable Y. By introducing an auxiliary variable E, (9) can be recast as

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{S}} \ \lambda \sum_{i=1}^{\min\{m,n\}} w_i\sigma_i(\mathbf{X}) \\ &\text{s.t.} \ \ \mathbf{Y} = \mathbf{X}+\mathbf{E}, \\ &\qquad P_{\mathbf{\Omega}}(\mathbf{E}) = 0 \end{aligned} \tag{10}$$

Equation (10) can be efficiently solved by the alternating direction method of multipliers (ADMM) to alternatively optimize the low-rank matrix $\mathbf{X}$ and error matrix $\mathbf{E}$ as

$$\min_{\mathbf{X}} \ \lambda \sum_{i=1}^{\min\{m,n\}} w_i\sigma_i(\mathbf{X}) + \frac{\mu}{2}\left\|\mathbf{Y}-\mathbf{X}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}\right\|_F^2, \tag{11}$$

$$\begin{aligned} &\min_{\mathbf{E}} \ \tfrac{\mu}{2}\left\|\mathbf{Y}-\mathbf{X}-\mathbf{E}+\tfrac{\mathbf{Y}_1}{\mu}\right\|_F^2 \\ &\text{s.t.} \quad P_{\mathbf{\Omega}}(\mathbf{E}) = 0, \end{aligned} \tag{12}$$

TABLE I
RUNNING TIMES OF THE WNNM, SQN, PSSV, AND THE PROPOSED GIRNN METHOD FOR FIG. 1.

Method	WNNM	SQN	PSSV	GIRNN
Time (s)	18.58	28.35	12.73	7.73

where $\mathbf{Y_1}$ is the Lagrangian multiplier. Equation (11) is easy to solve via the GIRNN algorithm and the closed form solution of (12) is found by its subgradient.

We test the image inpainting problem on a real world picture, which is a classic MC problem. We compare our proposed method, the GIRNN algorithm, with the WNNM method [1], partial sum singular value (PSSV) method [21], and the Schatten-q quasi norm (SQN) method [14]. All of the above methods are solved via the ADMM algorithms, and we pre-set the same values of the parameters r and $\lambda = \sqrt{2mn}$ for all the methods, where r is estimated via the method in [2]. The cross validation set is one figure, and we use the same value for all the testing images.

The inpainting results of the above methods on the 'sunset' image with 80% random mask are presented in Fig. 1. It can be seen that the proposed GIRNN algorithm has the highest average PSNRs of the presented state-of-the-art methods. The enlarged details show that the proposed algorithms can recover the sunset more clearly and precisely than others. In addition, among all the above methods, the GIRNN method is the most efficient one as shown in Table I and the caption of Fig. 1. All the Matlab programs were performed on a laptop with Intel i7-6820 HQ CPU and its core frequency is 2.6 GHz. All the algorithms stop when the residual errors are less than 10^{-3}. For the matrix completion problem, the main computational costs of WNNM, PSSV, GIRNN algorithms are the singular value decomposition (SVD) for the large-dimension matrix in every iteration. The efficiency of the SQN algorithm is restricted not only by the SVD but also by the Newton's method. Therefore, the computational complexities of all the above methods, except for the SQN algorithm, are similar for each iteration.

The values of p in the updated rules are critical to the final performance. The WNNM can be seen as the special case of the proposed GIRNN method with p equal to 0. But the WNNM method is restricted by the lost observations and can hardly recover the image in a high quality. In Fig. 2, we plot the average PSNRs of the proposed GIRNN algorithm for seven figures versus different values of p under 85% random missing pixels. Not surprisingly, the recovered PSNRs are lower when p is greater than zero and less than 0.5. And the best PSNRs are always located around the range from 0.7 to 0.9, which matches the analysis in [22] about natural images.

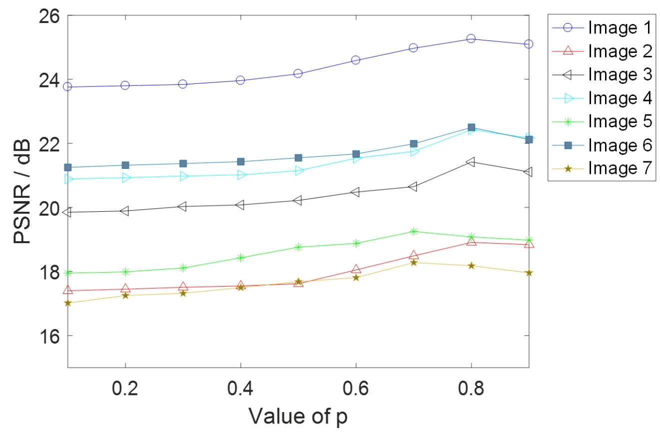

Fig. 2. Average PSNRs of the proposed GIRNN algorithm for eight figures versus different value of p under 85% random missing pixels.

B. Robust Principal Component Analysis (RPCA) Problem

Consider the following optimization models

$$\min_{\mathbf{X},\mathbf{S}} \lambda \sum_{i=1}^{\min\{m,n\}} w_i \sigma_i (\mathbf{X}) + \|\mathbf{S}\|_p^p \quad \text{s.t.} \;\; \mathbf{Y} = \mathbf{X} + \mathbf{S} \tag{13}$$

for the GIRNN algorithm, where $\|\mathbf{S}\|_p = (\sum |S_{ij}|^p)^{\frac{1}{p}}$ is the ℓ_p norm. The penalty terms of sparsity are restricted by the ℓ_p norm. In order to solve the problems in (13) efficiently, we use the ADMM to alternatively optimize the low-rank $\mathbf{X}$ and sparse $\mathbf{S}$. By tackling the low-rank matrix, it is easy to update $\mathbf{X}$ via the GIRNN method. We find that updating $\mathbf{X}$ just once in the inner loop is sufficient to generate an accurate solution for all the experiments. This approach, called the inexact GIRNN method, can reduce the computational burden. Then, to update the sparse term $\mathbf{S}$, it is optimized by solving the following problem:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathbf{S}\|_p^p + \frac{\mu}{2} \left\| \mathbf{Y} - \mathbf{X} - \mathbf{S} + \frac{\mathbf{Y}_2}{\mu} \right\|_F^2, \tag{14}$$

where $\mathbf{Y}_2$ is the Lagrangian multipliers. The lp norm minimization problem in (14) can be iteratively solved by the generalized shrinkage-thresholding (GST) operator under $0 \leq p < 1$ [23]. In numerical experiments below, the GST operator is demonstrated to be highly efficient and accurate.

1) Text Removal

We compare the proposed methods with the other RPCA or equivalent models, including the traditional RPCA model [4] (nuclear norm & ℓ_1 norm), WNNM model [1] (weighted nuclear norm & ℓ_1 norm), LpSq model [14] (Schatten q-norm & ℓ_p norm), PSSV model [21] (truncated nuclear norm & ℓ_1 norm), and the greedy go decomposition (GreGoDec) model [6] (exact rank and cardinality constraint).

Next we apply the above methods on the real natural image in Fig. 3 to further validate the effectiveness of the proposed methods. The rank is set to 60 for all the methods and λ is the same as above. Limited by the pages, we omit the text image

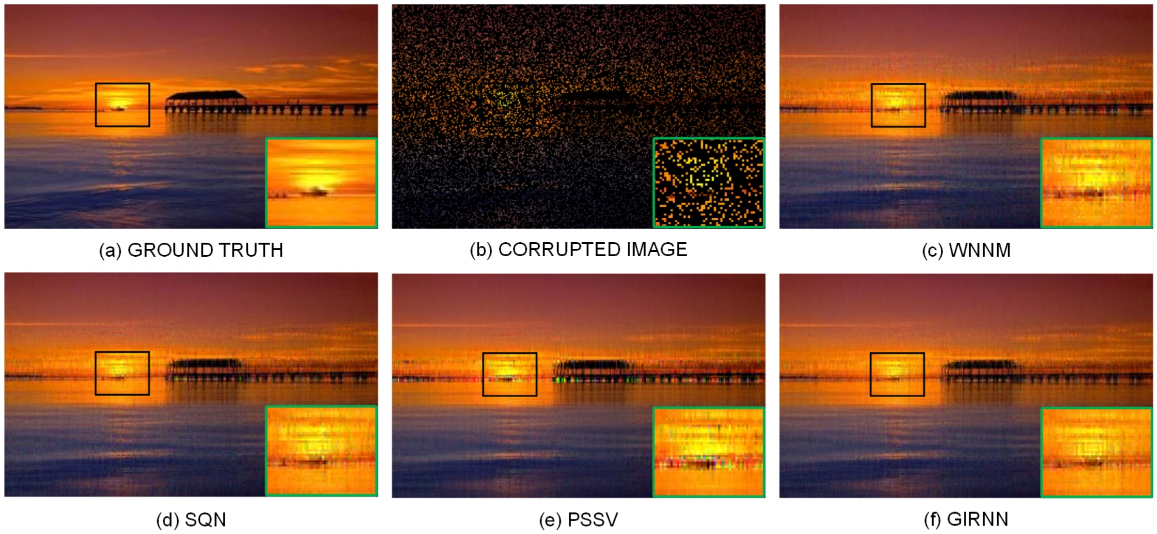

Fig. 1. Image inpainting results of different methods with 80% random missing pixels. (a) Ground truth. (b) Corrupted image. (c) WNNM. PSNR = 27.96 dB. Time = 18.58 s. (d) SQN. PSNR = 28.76 dB. Time = 28.35 s. (e) PSSV. PSNR = 27.48 dB. Time = 12.73 s. (f) GIRNN. PSNR = 29.24 dB. Time = 7.73 s.

removed by the above methods. All the methods can recover the image without visible text, except for the GreGoDec algorithm. Even though the proposed methods cannot recover the ground truth as clearly as the toy image, it can be seen that the proposed methods still outperform the other state-of-the-art methods. For example, one can see a blurry bird on the top of the tree in Fig. 3 (h). The other methods miss the bird in the recovered images.

2)Background Extraction and Pedestrian Detection

The background extraction problem is a classic RPCA problem, because the background is almost invariant for all the frames of a surveillance video and the pedestrian moves in each frame, which can be explained as the low-rank background and a sparse moving pedestrian. In this subsection, we test the proposed methods in Fig. 4 on the popular surveillance video: Bootstrap dataset (http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html). We select 200 frames for the experiment and set the rank as 3 for all the methods. As a result, all the methods can detect the pedestrians correctly. However, the LpSq, GreGoDec, RPCA, and PSSV methods detect the reflected light on the floor as the sparse part (shown in the red rectangles). Also, the backgrounds in the yellow rectangles are falsely extracted from the pedestrians. It is clear that the decomposition results of the proposed methods are better than those of the other methods.

V. Conclusion

In this paper, the generalized iterative reweighted nuclear norm (GIRNN) algorithm was proposed to approximate the Schatten-p norm for low-rank approximation in some practical applications. We formulated the closed-form solution for the iterative reweighted algorithm. The proposed algorithm outperformed other selected state-of-the-art algorithms in the practical applications such as image inpainting, pedestrian detection, and text removal. In the future, we will work on the fast implementation of the iterative reweighted algorithm to make them more efficient while remaining their current excellent performance.

References

[1] S. Gu, L. Zhang, W. Zuo, et al, “Weighted nuclear norm minimization with application to image denoising,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2862-2869.

[2] F. Shang, J. Cheng , Y. Liu , et al, “Bilinear factor matrix norm minimization for robust pca: Algorithms and applications,” IEEE transactions on pattern analysis and machine intelligence, 2017.

[3] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2233-2246, Nov. 2012.

[4] E. Cands, X. Li, Y. Ma, et al, “Robust principal component analysis?,” Journal of the ACM (JACM), 2011, 58(3): 11.

[5] G. Liu, Z. Lin, S. Yan, et al, “”Robust recovery of subspace structures by low-rank representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171-184.

[6] T. Zhou, D. Tao, “Greedy bilateral sketch, completion & smoothing,” in International Conference on Artificial Intelligence and Statistics. JMLR. org, 2013.

[7] N. Srebro, R. Salakhutdinov, “Collaborative filtering in a non-uniform world: Learning with the weighted trace norm,” in Advances in Neural Information Processing Systems. 2010: 2056-2064.

[8] J. Cai, E. Cands, Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM Journal on Optimization, 1956, 20(4): 2010.

[9] X. Peng, C. Lu, Z. Yi, et al, “Connections between nuclear-norm and frobenius-norm-based representations,” IEEE transactions on neural networks and learning systems, 2017.

[10] H. Zhang, Z. Yi, and X. Peng, “fLRR: Fast low-rank representation using frobenius-norm,” Electron. Lett., vol. 50, no. 13, pp. 936-938, Jun. 2014.

[11] E. Candes, M. Wakin, and S. Boyd, “Enhancing sparsity by reweighted l1 minimization,” J. Fourier Anal. Appl., vol. 14, no. 5, pp. 877-905, Dec. 2008.

[12] S. Wang, L. Zhang, and Y. Liang, “Nonlocal spectral prior model for low-level vision,” in Proc. Asian Conf. Comput. Vis., 2013, pp. 231-244.

[13] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, “Fast and accurate matrix completion via truncated nuclear norm regularization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 9, pp. 2117-2130, Sep. 2013.

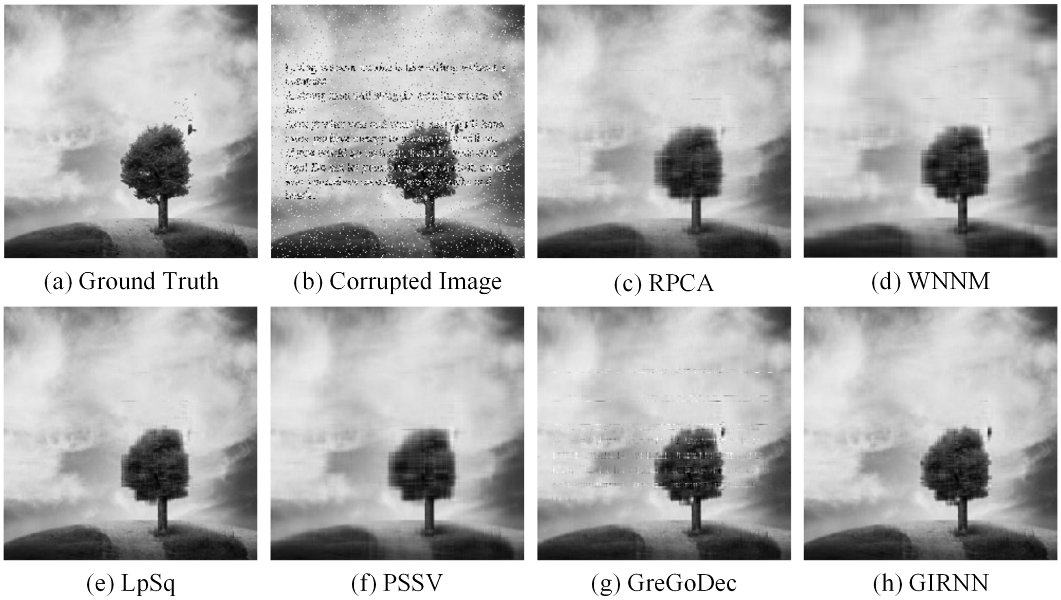

Fig. 3. Removal results of the real natural image. (a) Ground truth. (b) Input image. (c) Traditional RPCA. F-measure: 0.9325, RMSE: 0.1131. (d) WNNM. F-measure: 0.9329, RMSE: 0.1158. (e) LpSq. F-measure: 0.9371, RMSE: 0.0971. (f) PSSV. F-measure: 0.9266, RMSE: 0.1130. (g) GreGoDec. F-measure: 0.9161, RMSE: 0.1264. (h) GIRNN. F-measure: 0.9444, RMSE: 0.0812.

Fig. 4. Background extraction and pedestrian detection for different algorithms on the Bootstrap dataset. The extracted backgrounds are on the top and the detected pedestrians are at the bottom.

[14] F. Nie, H. Wang, X. Cai, et al, "Robust matrix completion via joint schatten p-norm and lp-norm minimization," in Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, 2012: 566-574.

[15] C. Lu, Z. Lin, and S. Yan, "Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization," IEEE Trans. Image Processing, vol. 24, no. 2, pp. 646-654, Feb. 2015.

[16] C. Lu, J. Tang, S. Yan, et al, "Generalized nonconvex nonsmooth low-rank minimization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 4130-4137.

[17] C. Lu, C. Zhu, C. Xu, et al, "Generalized Singular Value Thresholding," AAAI. 2015: 1805-1811.

[18] L. Mirsky, "A trace inequality of John von Neumann," Monatsh. Math., vol. 79, pp. 303-306, 1975.

[19] J. von Neumann, Some matrix-inequalities and metrization of matrix-space, Tomsk Univ. Rev, vol. 1, pp. 286-300, 1937.

[20] D. Rhea, "The case of equality in the Von Neumann trace inequality," preprint, 2011.

[21] T. Oh, Y. Tai, J. Bazin, H. Kim, and I. Kweon, "Partial sum minimization of singular values in robust PCA: Algorithm and applications," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 4, pp. 744-758, Apr. 2016.

[22] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 1033-1041.

[23] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 217-224.

[24] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 597-610, Mar. 2013.