ROBUST PRINCIPAL COMPONENT ANALYSIS WITH MATRIX FACTORIZATION

Yongyong Chen and Yicong Zhou

Department of Computer and Information Science
University of Macau, Macau 999078, China
YongyongChen.cn@gmail.com, yicongzhou@umac.mo

ABSTRACT

Traditional robust principle component analysis (RPCA) has a high computational cost because RPCA needs to calculate the singular value decomposition of large matrices. To address this issue, this paper proposes a matrix-factorization-based RPCA (MFRPCA) model. MFRPCA has high computation efficiency while improving the robustness and flexibility of traditional RPCA using a non-convex low-rank approximation. Experiment results on challenging datasets demonstrate superior performance of MFRPCA compared with several advanced low-rank reconstruction methods.

Index Terms— Robust PCA, Matrix factorization, Non-convex regularizer, Background subtraction.

1. INTRODUCTION

Low-rank matrix approximation (LRMA), aiming to reconstruct a low-rank matrix L from the noisy data $D \in R^{m\times n}$, has gained increasing interest in the societies of computer vision, image and video processing. For example, the moving objects detection as a basic operation in video analysis aims at subtracting the "foreground" such as pedestrians and cars from a set of video named "background". Principal component analysis (PCA) as a classical LRMA method is brittle to non-Gaussian noise (outliers). To address this shortcoming, Candès *et al.* [1] developed robust PCA (RPCA), which can be expressed in (1)

$$\min_{L,S} rank(L) + \lambda\|S\|_0, \;\; s.t. \;\; D = L + S. \tag{1}$$

Where L is a low-rank matrix, while S is a sparse matrix such as outliers. For example, L represents the recovered hyperspectral images (HSI) as shown in Section 3.3 and S denotes the impulse noise, stripes, dead lines, and many others. However, solving (1) is NP-hard and intractable. The most popular choice is to utilize the nuclear norm and l_1-norm instead of using the rank function and l_0-norm, leading to the following convex optimization problem:

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1, \;\; s.t. \;\; D = L + S, \tag{2}$$

where $\|L\|_* = \sum_i \sigma_i$ is the nuclear norm; σ_i is the i-th singular value of the matrix L. $\|S\|_1$ is the l_1-norm.

Various approaches have been developed for solving (2). Examples include singular value thresholding [2], exact and inexact augmented lagrange method [3], alternating direction method of multipliers [4], fast alternating linearization method [5]. However, these approaches inevitably need to calculate the singular value decomposition (SVD). Therefore, with the increase of matrix dimension, computation complexity exponentially increases. On the other hand, because nuclear norm treats each singular value equally, this leads to overshrinking the rank component. The main reason is that larger singular values of an input data matrix quantify the main information of its underlying principal directions. To address this issue, several recent studies have investigated non-convex sparsity-inducing regularizers, such as weighted nuclear norm [6], capped norm [7], weighted Schatten p-norm [8], log-determinant penalty [9], γ-norm [10]. Their promising performance has been validated by extensive experiments on background /foreground separation, face image shadow removal, image inpainting, image alignment, and multispectral/hyperspectral image denoising.

Another way to LRMA is low-rank matrix factorization [11–13], *i.e.*, $L = UV^T$, where $U \in R^{m\times r}$, $V \in R^{n\times r}$ and $r \ll \min\{m, n\}$. The low rank property is based on the fact that $rank(L) = rank(UV^T) \leq \min\{rank(U), rank(V)\}$. Considering the assumption that data noise may be modeled as a mixture of Gaussians, Zhao *et al.* proposed a generative RPCA model under the Bayesian framework. One bottleneck of low-rank matrix factorization methods is that they usually need to precisely predefine r, which is hard in practice.

Inspired by the promising performance of the non-convex sparsity-inducing regularizer [10] and the low computational cost of the low-rank matrix factorization technology, we propose a novel matrix-factorization-based RPCA model (MFRPCA).

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/016/2015/A1 and by the Research Committee at University of Macau under Grant MYRG2016-00123-FST.

2. PROPOSED MFRPCA

In this section, we introduce the proposed MFRPCA model, in which the low-rank property is realized by a matrix factorization scheme and a non-convex sparsity-inducing regularizer is used to improve the robustness and flexibility of the traditional nuclear norm.

2.1. Problem formulation

The nuclear norm-based algorithms always suffer from high computation cost issue due to computing SVDs in each iteration. We adopt the matrix factorization approach to model the low-rank prior of the underlying clean data. Therefore, we write the observed data matrix D in model (1) as $D = UV^T + S$, where S represents the sparse error matrix, yielding the following optimization problem:

$$\begin{aligned} &\min_{U,V,S} \|S\|_1 \\ &s.t.\ \ D = UV^T + S. \end{aligned} \tag{3}$$

For the model in (3), we should give the exact value of rank r. However, in real applications, the prior information on r is not present. To overcome this dilemma, we introduce a non-convex low-rank regularizer defined in our previous work [10]. And following the idea in [14], we enforce an equality constraint $U^TU = I$, where I is an identity matrix, to facilitate the uniqueness of solution. Hereby, we obtain a much smaller-scale minimization problem:

$$\begin{aligned} &\min_{U,V,S} \|S\|_1 + \lambda\|V\|_\gamma \\ &s.t.\ \ D = UV^T + S,\ \ U^TU = I, \end{aligned} \tag{4}$$

where $\|V\|_\gamma = \sum_i \left(1 - e^{-\sigma_i(V)/\gamma}\right), \gamma > 0$ is a particular non-convex sparsity-inducing regularizer. λ is a trade-off parameter to control the contribution between low-rank and sparse priors. Compared with the convex model in (2), the main advantage of our model in (4) lies in the fact that the convex model in (2) is converted into a small-scale matrix minimization problem, resulting in lower computational cost. Another merit is that, as shown in our experimental results in Section 3, our proposed model needs to know only a upper bound of the true rank r by exploiting the non-convex regularizer. Whereas, the challenge of model in (4) is that the sub-gradient method is no longer applicable due to the non-convexity of $\|V\|_\gamma$.

2.2. MFRPCA algorithm

Here, we solve our proposed model in (4) using the efficient augmented Lagrangian multipliers (ALM) [3] method. The partial augmented Lagrangian function of model in (4) is:

$$\begin{aligned} \mathcal{L}(U,V,S,\Pi;\rho) = &\|S\|_1 + \lambda\|V\|_\gamma + \\ &\langle \Pi, D - UV^T - S\rangle + \frac{\rho}{2}\|D - UV^T - S\|_F^2, \end{aligned} \tag{5}$$

where $\Pi \in R^{m\times n}$ is the Lagrangian multiplier corresponding to constraint $D = UV^T + S$ and ρ is a positive scalar. Here, an efficient alternating direction strategy was adopted to iteratively minimize each variable by fixing all others as follows:

$$U_{k+1} = \arg\min_{U^TU=I} \mathcal{L}(U, V_k, S_k, \Pi_k; \rho_k); \tag{6}$$

$$V_{k+1} = \arg\min_{V} \mathcal{L}(U_{k+1}, V, S_k, \Pi_k; \rho_k); \tag{7}$$

$$S_{k+1} = \arg\min_{S} \mathcal{L}(U_{k+1}, V_{k+1}, S, \Pi_k; \rho_k); \tag{8}$$

$$\Pi_{k+1} = \Pi_k + \rho_k(D - U_{k+1}V_{k+1}^T - S_{k+1}); \tag{9}$$

$$\rho_{k+1} = \min\{\beta * \rho_k, \rho_{max}\}. \tag{10}$$

where U_k denotes U in the k-th iteration, and β is set to 1.618 to further facilitate the convergence speed.

Update for U: In particular, for solving variable U, the sub-problem is formulated as:

$$U_{k+1} = \arg\min_{U^TU=I} \frac{\rho_k}{2}\|UV_k^T - (T - S_k)\|_F^2, \tag{11}$$

where $T = D + \Pi_k/\rho_k$. The closed-form solution to (11) can be obtained using the classical Orthogonal Procrustes problem [15]. Suppose $(T - S_k)V = A\Sigma B^T$ is SVD of matrix $(T - S_k)V$, the optimal solution to sub-problem in (11) is

$$U_{k+1} = AB^T. \tag{12}$$

Update for V: To solve the variable V, the optimization problem is formulated as:

$$V_{k+1} = \arg\min_{V} \|V\|_\gamma + \frac{\rho_k}{2\lambda}\|V - (T - S_k)^TU_{k+1}\|_F^2. \tag{13}$$

Note that the γ-norm is concave, the sub-problem in (13) is non-convex. We utilize the Difference of Convex programming [16] to efficiently solve (13).

Lemma 1 *The sub-gradient of $\|V\|_\gamma$ is*

$$\partial\|V\|_\gamma = \{A_V diag(l) B_V^T\}, \tag{14}$$

where $l_i = e^{\frac{-\sigma_i(V)}{\gamma}}/\gamma$, the columns of A_V and B_V are the left and right singular matrices of V, respectively.

Based on Lemma 1, the sub-problem in (13) can be changed as

$$V_{k+1} = \arg\min_{V}\ \langle \partial\|V_k\|_\gamma, V\rangle + \frac{\rho_k}{2\lambda}\|V - (T - S_k)^TU_{k+1}\|_F^2 \tag{15}$$

Then, the optimal solution to (15) is

$$V_{k+1} = (T - S_k)^TU_{k+1} - \lambda\partial\|V_k\|_\gamma/\rho_k. \tag{16}$$

Update for S: To solve variable S, the optimization problem is formulated as:

$$S_{k+1} = \arg\min_{S} \|S\|_1 + \frac{\rho_k}{2}\|S - (T - U_{k+1}V_{k+1}^T)\|_F^2.$$

It has a closed-form solution S_{k+1} via the element-wise shrinkage-thresholding operator [17], *i.e.*,

$$[S_{k+1}]_{i,j} = \max\{|[W_k]_{i,j}| - 1/\rho_k, 0\} sign([W_k]_{i,j}) \quad (17)$$

where $W_k = T - U_{k+1}V_{k+1}^T$.

Algorithm 1 shows the detail procedures of MFRPCA.

Algorithm 1 MFRPCA algorithm

Input: The observed data matrix D, parameters λ, ρ_0, β.
Initialize: λ, β, ϵ, ρ_0, S_0, V_0, Π_0, k.

```
while not converged do
    Compute T = D + Π_k/ρ_k;
    Update U_{k+1} by (12);
    Update V_{k+1} by (16);
    Compute W_k = T − U_{k+1}V_{k+1}^T;
    Update S_{k+1} by (17);
    Update Π_{k+1} by (9);
    Update ρ_{k+1} by (10);
    Check the convergence condition
        ‖D − U_{k+1}V_{k+1}^T − S_{k+1}‖_F ≤ ϵ ∗ ‖D‖_F.
end while
```

Output: The low-rank matrix $L = U_{k+1}V_{k+1}^T$.

2.3. Complexity analysis

The computational complexity of MFRPCA is dominated by updating variables: U and V. Updating U and V have a running time of $\mathcal{O}(mr^2 + mnr)$ and $\mathcal{O}(nr^2 + mnr)$, respectively. To update S and Π all need $\mathcal{O}(mn)$ cost per iteration. Therefore the total complexity of MFRPCA is $\mathcal{O}(mr^2+mnr)$ $(m \geq n)$.

3. EXPERIMENTAL RESULTS

The MATLAB (2012a) code was implemented on a personal lenovo laptop with a 2.3GHz and 4GB of memory.

3.1. Experimental settings and implementation details

To test the performance of our proposed method, we consider the following two applications:

(1) background extraction in surveillance videos; In subsection 3.2, our proposed method was compared with the open source state-of-the-art RPCA-based approaches: IALM [3], NcRPCA [18], GoDec [19], and RBF [14];

(2) hyperspectral image denoising. In subsection 3.3, we conduct experiments for hyperspectral image denoising task with approaches: VBM3D [20], LRMR [21], NAILRMA [22]. Note that these methods, especially LRMR and NAILRMA, represent the state-of-the-art HSI denoising methods.

We download the codes from the authors' websites using their default settings of parameters.

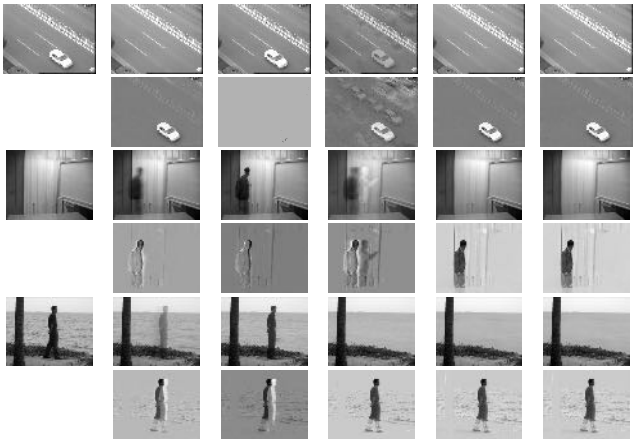

Fig. 1: Background extraction in three surveillance videos: Car, Curtain, and Water surface. The first column: the original frames of Car, Curtain, and Water surface, respectively. Second column to the last column: estimated backgrounds and foregrounds by IALM, NcRPCA, GoDec, RBF, and MFRPCA, respectively.

3.2. Background extraction in surveillance videos

Background extraction in surveillance videos is an important application of RPCA and a basic task in computer vision. The parameters in our method are listed as follows: $\gamma = 0.05$, $\beta = 1.618$, $\rho_0 = 0.01$, $\lambda = 20$. Since the background is static or approximately static, the rank of the background should be 1. In order to estimate the robustness of our method, we set $r = 5$ which is an upper bound of the true rank 1. In this subsection, we evaluate all methods on three challenging datasets [1], as shown in Table 1.

The visual results of representative frames in three different surveillance videos are shown in Fig. 1. It is easy to see that for all the testing data, MFRPCA is able to produce clear background, as well as reconstructing a complete foreground even under prominently embedded foreground moving objects. However, the competing methods either generate more or less artifacts/residuals in the background or cannot completely detect the moving objects. As shown in the last two rows in Fig. 1, backgrounds obtained by IALM and NcRPCA still remain the shadow of the moving man. Similar observation could be drawn from other two datasets.

We also provide some quantitative assessment results of different background extraction approaches in Table 1, in which RelErr is defined as $\frac{\|D-L-S\|_F}{\|D\|_F}$. As can be seen, the number of iterations (Iter) of MFRPCA is much less than these of other methods. Meanwhile, MFRPCA is the fastest

[1] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

algorithm. This is due to the matrix factorization scheme and non-convex low-rank regularizer. Note that RBF has the similar visual performance with our proposed method, however, the running time of RBF is two times longer than that of MFRPCA. Although we set $r = 5$, the low-rank matrix L recovered by MFRPCA is still with the true rank 1, indicating that our proposed method is more robust than other competing methods.

Table 1: Quantitative results on three real datasets

Data (size)	Algorithm	Rank(L)	$\frac{\|S\|_0}{mn}$	RelErr	Iter	Time(s)
Car ($240 \times 320 \times 24$)	IALM	8	0.644	9.47e-4	16	4.66
	NcRPCA	5	0.624	7.13e-4	58	27.02
	GoDec	5	0.977	1e-6	101	50.57
	RBF	5	0.717	7.95e-4	24	3.54
	MFRPCA	1	0.772	7.98e-4	12	1.92
Curtain ($128 \times 160 \times 633$)	IALM	100	0.745	6.33e-4	20	54.60
	NcRPCA	5	0.823	7.62e-4	41	42.84
	GoDec	5	0.219	1.04e-2	101	155.65
	RBF	2	0.901	8.72e-4	25	13.35
	MFRPCA	1	0.845	9.08e-4	12	6.53
Water surface ($128 \times 160 \times 633$)	IALM	211	0.733	8.67e-4	20	139.83
	NcRPCA	5	0.909	2.38e-4	46	76.45
	GoDec	5	0.139	2.29e-2	101	224.82
	RBF	1	0.837	9.78e-4	23	26.91
	MFRPCA	1	0.855	8.05e-4	12	10.93

3.3. Hyperspectral image denoising

HSI denoising has aroused increasing attention on various fields, including environmental studies, military surveillance, and biomedical imaging. However, HSIs are inevitably corrupted by Gaussian, impulse noise, stripes, dead lines, and many others [21]. Therefore, it is very important to remove the HSI noise. Fortunately, the denoising methods based on low-rank matrix approximation [10, 21, 22] have achieved promising performance in HSI denoising.

In this experiment, we select a real-world HSI dataset (EO-1 Hyperion Australia dataset) [2] to investigate the performance of our proposed method. The size of original image is $3858 \times 256 \times 242$. Here, we use only a subregion of size $400 \times 200 \times 150$ in our experiment. And the parameters of our method are: $\gamma = 0.05,\ \beta = 1.5,\ \rho_0 = 0.005,\ \lambda = 1$.

The restoration results of three typical bands of EQ-1 Hyperion Australia dataset are shown in Fig. 2. The original bands, see the first column of Fig. 2, are contaminated by the mixture of Gaussian noise, impulse noise, stripes, and dead lines. It is easy to see that MFRPCA could effectively remove the mixed noise, and meanwhile, preserve the essential structures of HSIs. All other competing methods could reduce the mixed noise to a certain level. VBM3D suffers from oversmoothing the results and fails to restore images with heavy noise. LRMR and NAILRMA can partially remove noise. When handing the dead lines, all the restored results by VBM3D, LRMR, and NAILRMA (see the third row) are presented ghosting shadow more or less at band 123. Only the proposed method achieves a desirable result. The experimental results show that our algorithm has a great advantage of removing stripes and dead lines.

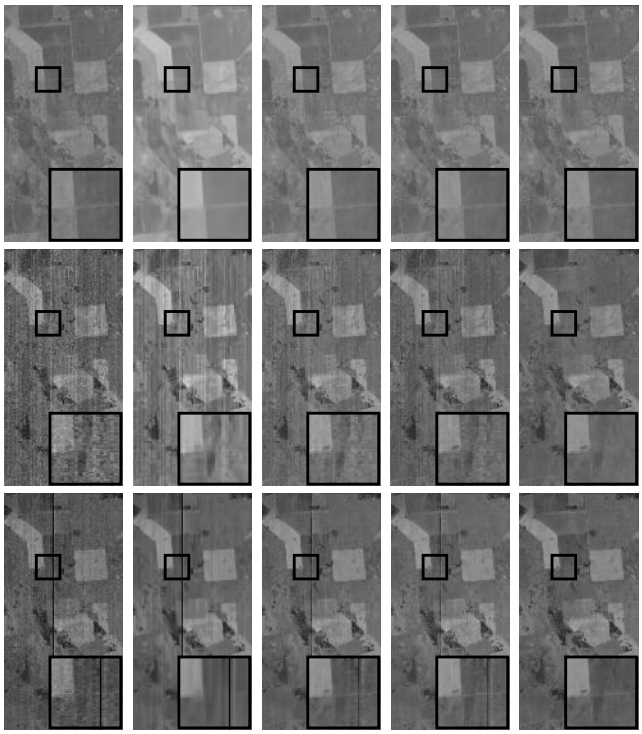

Fig. 2: Denoising results of EO-1 Hyperion Australia dataset. The three rows from top to bottom are the bands located at 83, 88 and 123; the five columns from left to right are the original bands with mixed noise and the restored bands obtained by: VBM3D, LRMR, NAILRMA, and MFRPCA, respectively.

4. CONCLUSION

In this paper, we have proposed a novel matrix factorization-based RPCA (MFRPCA) model. Compared with the nuclear norm-based approaches, MFRPCA uses the matrix factorization scheme to facilitate the computation speed, and preserves the low-rank property of the underlying data. To further improve the robustness of the traditional nuclear norm, MFRPCA adopts a non-convex sparsity-inducing regularizer to give an upper-bound of the true rank. To validate the effectiveness of MFRPCA, we have tested them on two important applications: background subtraction in surveillance videos and hyperspectral image denoising. Experiment results demonstrated the efficiency, effectiveness, and robustness of MFRPCA.

[2] http://remote-sensing.nci.org.au/

5. REFERENCES

[1] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11, 2011.

[2] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[3] Zhouchen Lin, Minming Chen, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[4] Xiaoming Yuan and Junfeng Yang, "Sparse and low rank matrix decomposition via alternating direction method," *Pacific Journal of Optimization*, vol. 9, no. 1, 2009.

[5] Donald Goldfarb, Shiqian Ma, and Katya Scheinberg, "Fast alternating linearization methods for minimizing the sum of two convex functions," *Mathematical Programming*, pp. 1–34, 2013.

[6] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *International journal of computer vision*, vol. 121, no. 2, pp. 183–208, 2017.

[7] Qian Sun, Shuo Xiang, and Jieping Ye, "Robust principal component analysis via capped norms," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 311–319.

[8] Yuan Xie, Shuhang Gu, Yan Liu, Wangmeng Zuo, Wensheng Zhang, and Lei Zhang, "Weighted schatten p-norm minimization for image denoising and background subtraction," *IEEE transactions on image processing*, vol. 25, no. 10, pp. 4842–4857, 2016.

[9] Zhao Kang, Chong Peng, and Qiang Cheng, "Robust pca via nonconvex rank approximation," in *Data Mining (ICDM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 211–220.

[10] Yongyong Chen, Yanwen Guo, Yongli Wang, Dong Wang, Chong Peng, and Guoping He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5366–5380, 2017.

[11] Naiyan Wang, Tiansheng Yao, Jingdong Wang, and Dit-Yan Yeung, "A probabilistic approach to robust matrix factorization," in *European Conference on Computer Vision*. Springer, 2012, pp. 126–139.

[12] Deyu Meng and Fernando De La Torre, "Robust matrix factorization with unknown noise," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1337–1344.

[13] Qian Zhao, Deyu Meng, Zongben Xu, Wangmeng Zuo, and Lei Zhang, "Robust principal component analysis with complex noise," in *International Conference on Machine Learning*, 2014, pp. 55–63.

[14] Fanhua Shang, Yuanyuan Liu, James Cheng, and Hong Cheng, "Recovering low-rank and sparse matrices via robust bilateral factorization," in *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 2014, pp. 965–970.

[15] Peter H Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[16] Pham Dinh Tao and Le Thi Hoai An, "Convex analysis approach to dc programming: Theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[17] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[18] Praneeth Netrapalli, UN Niranjan, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain, "Non-convex robust pca," in *Advances in Neural Information Processing Systems*, 2014, pp. 1107–1115.

[19] Tianyi Zhou and Dacheng Tao, "Godec: Randomized low-rank and sparse matrix decomposition in noisy case.," in *International Conference on Machine Learning*, 2011, pp. 33–40.

[20] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[21] Hongyan Zhang, Wei He, Liangpei Zhang, Huanfeng Shen, and Qiangqiang Yuan, "Hyperspectral image restoration using low-rank matrix recovery," vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[22] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3050–3061, 2015.