Denoising of Hyperspectral Images Using Nonconvex Low Rank Matrix Approximation

Yongyong Chen, Yanwen Guo, Yongli Wang, Dong Wang, Chong Peng, and Guoping He, *Member, IEEE*

***Abstract*— Hyperspectral image (HSI) denoising is challenging not only because of the difficulty in preserving both spectral and spatial structures simultaneously, but also due to the requirement of removing various noises, which are often mixed together. In this paper, we present a nonconvex low rank matrix approximation (NonLRMA) model and the corresponding HSI denoising method by reformulating the approximation problem using nonconvex regularizer instead of the traditional nuclear norm, resulting in a tighter approximation of the original sparsity-regularised rank function. NonLRMA aims to decompose the degraded HSI, represented in the form of a matrix, into a low rank component and a sparse term with a more robust and less biased formulation. In addition, we develop an iterative algorithm based on the augmented Lagrangian multipliers method and derive the closed-form solution of the resulting subproblems benefiting from the special property of the nonconvex surrogate function. We prove that our iterative optimization converges easily. Extensive experiments on both simulated and real HSIs indicate that our approach can not only suppress noise in both severely and slightly noised bands but also preserve large-scale image structures and small-scale details well. Comparisons against state-of-the-art LRMA-based HSI denoising approaches show our superior performance.**

***Index Terms*— Denoising, hyperspectral image (HSI), nonconvex low rank approximation (NonLRMA).**

I. Introduction

HYPERSPECTRAL images (HSIs) are widely used in various applications, including environmental studies, biomedical imaging, and military surveillance [1]–[4], due to digital sensing technology advance. In these applications, however, some difficulties are exposed. First, the size of the ground HSIs is increasing dramatically, resulting in "the curse of dimensionality." Second, caused by photon shot noise, malfunctioning arrays in camera sensors, faulty memory locations in hardware, and transmission in a noisy channel [5], HSIs are often unavoidably corrupted by several types of noises, including Gaussian noise, impulse noise, deadlines, and stripes [6]. These pose great challenges to further applications, such as classification, target detection, and HSI unmixing [7]–[12]. It is, therefore, of vital importance to design an HSI denoising method, which can efficiently process HSI data and, meanwhile, effectively remove these different types of noises. To achieve this, the sparsity property of HSIs, which states that high-dimensional HSI data often lie in a low-dimensional subspace, should be sufficiently exploited.

Various techniques have been proposed for image denoising. Along them, the global correlation along spectrum (GCS) and the nonlocal self-similarity across space (NSS) as prior information have been widely used by them. An HSI contains dozens or even hundreds of bands, and GCS means that images across different spectra are highly correlated. Mathematically, by representing an HSI as a matrix, GCS implies the low rank property of the data matrix. The NSS refers to the fact that many full-band patches similar to a given local full-band patch can be found [13]. This property has been used [14]–[16] to denoise the HSI. As stated in [6] and [17], the performance of these methods often suffers from the fact that the intrinsic properties are inevitably neglected [13]. An HSI essentially can be viewed as a 3-D array. Based on this observation, tensor-based multilinear data analysis is proposed [18]–[20]. For example, to take account of the spectral redundancy, the lower rank tensor approximation algorithm, which jointly removes noise and reduces spectral dimensions, is proposed in [20]. The robust low-rank tensor recovery algorithm proposed in [19] is similar to tensor recovery via multi-linear augmented Lagrange multiplier [21] except that it adopts the $l_{2,1}$ norm. A novel tensor-based denoising method is proposed in [13] where GCS and NSS as prior knowledge are exploited. Most tensor-based HSI denoising models in the literature mainly take two forms: CANDECOMP/PARAFAC (CP) decomposition and Tucker decomposition. However, these methods may suffer from several issues. On the one hand, it is NP-hard to compute the CP rank [22]. As a result, the best low rank CP approximation of a tensor is unknown. On the other hand, Tucker decomposition is not unique. Besides, the application of a core tensor and an n-mode tensor product would incur information loss of spatial details [23].

Recently, the low rank models in which one aims to recover a low rank matrix representing the global information and a sparse one denoting the local information from the observation have attracted considerable attention. Besides,

Manuscript received January 19, 2017; revised April 9, 2017; accepted May 15, 2017. Date of publication June 27, 2017; date of current version August 25, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61373059, Grant 61672279, Grant 41505022, and Grant 11626143, in part by National Science Foundation of Jiangsu Province under Grant BK20150016, and in part by the Fundamental Research Funds for the Central Universities under Grant 020214380034/4-3. *(Corresponding author: Yanwen Guo.)*

Y. Chen and Y. Wang are with the College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China.

Y. Guo is with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China (e-mail: ywguo@nju.edu.cn).

D. Wang is with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China.

C. Peng is with the Department of Computer Science, Southern Illinois University Carbondale, Carbondale, IL 62901 USA.

G. He is with the Shandong Academy of Science, Jinan 250014, China.

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

Digital Object Identifier 10.1109/TGRS.2017.2706326

0196-2892 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Wright *et al.* [24] and Candès *et al.* [25] have shown that under incoherence assumptions, the solution can be exactly obtained with a high probability despite significant corruption. Compared with the classical principal component analysis (PCA), which is sensitive to outliers and gross corruptions (non-Gaussian noise), the low rank models are robust to various types of noises. Inspired by the fact that the low rank property generally holds for clean HSI, denoising methods based on low rank decomposition [6], [26]–[30] lexicographically vectorize all images or similar patches to yield 2-D data, which can then be decomposed into a low rank matrix representing the clean images (patches) and a sparse matrix denoting non-Gaussian noise.

Albeit the success of low rank models in theoretical research and practical applications, there are still some drawbacks in convex relaxation. First, the nuclear norm adds all singular values together rather than treating them equally as the rank function does. This implies that larger singular values are penalized more heavily than smaller ones. Furthermore, the nuclear norm is essentially an l_1 norm of singular values. As is well known, l_1 norm generally results in a biased estimator due to its shrinkage effect [31]–[33]. Second, previous theoretical analysis is generally based on a strong assumption that the underlying matrix satisfies incoherence property, which may not be guaranteed in practical scenarios. Moreover, the convex relaxation has poor convergence rate as the matrix dimension grows. To date, using nonconvex function instead of the l_1 norm as the surrogates of the l_0 norm has received much attention. Specifically, the popular nonconvex regularizers include minimax concave penalty [32], l_p norm ($0 < p < 1$) [34], log-sum penalty [35], log-determinant penalty [36], truncated nuclear norm (TNN) [37], and the capped l_1 penalty [38].

Some nonconvex low rank matrix approximation (NonLRMA) models have been proposed for HSI denoising. But they are different from our work. Zhang *et al.* [6] introduced a mixed HSI restoration model, which applies low rank matrix recovery (LRMR) to each image patch to simultaneously remove Gaussian noise, impulse noise, deadlines, and stripes. To further remove mixed noise, the nonconvex formulation GoDec [39], which alternates between updating low rank matrix via low rank projection and estimating sparse matrix using a hard thresholding operator, is used. To reduce complexity, a bilateral random projection method is employed to avoid computing singular value decomposition (SVD), leading to significant increase of processing speed. However, GoDec needs to first estimate the rank and cardinality, which are not easy-to-obtain in practice. That is mainly because the rank of each patch is usually relatively small and generally unknown. Another limitation is that it cannot guarantee a unique solution, since different projection may yield different results [40]. Similar to [6], He *et al.* [27] developed the noise-adjusted iterative LRMA to account for different noise intensities in different bands, whereas similar issues in [6] are introduced by utilizing the randomized SVD method to solve the nonconvex formulation. In addition, the rank also needs to be predefined, which implies that the intrinsic rank information cannot be automatically captured. Some other denoising approaches [17], [41], [42] were developed.

Our Work: To overcome the aforementioned issues, this paper presents the NonLRMA and the corresponding HSI denoising method to more accurately extract the low rank component from the degraded HSI, which may be contaminated by various types of noises. Our key observation is that those larger singular values should be less penalized. This is quite reasonable, since for natural images, larger singular values are more important than smaller ones as they imply energy of the major components, and, by contrast, the smaller ones mainly correspond to noise. To achieve this, a novel nonconvex regularizer, instead of the traditional nuclear norm widely used by previous methods, is presented in our formulation, such that the low rank property of HSIs is better exploited. Note that using a nonconvex penalty function to capture the low rank information is also motivated by the conventional compressive sensing recovery [43] and the recent literature [28], [36], [44], [45].

In summary, our contributions are threefold.

1) We leverage NonLRMA to denoise HSIs, which may be corrupted by mixed noises, not just Gaussian noise assumed by most previous methods, and study the general solver of this problem.
2) A novel nonconvex regularizer is proposed in our NonLRMA. Unlike the traditional nuclear norm, which shrinks all singular values with the same amount, the new regularizer achieves more sparse approximation of singular values and automatically captures the intrinsic rank information. Furthermore, different from [6] and [39], NonLRMA does not require the rank information to be known beforehand.
3) An effective, robust, yet easy-to-implement iterative algorithm based on the augmented Lagrangian multipliers (ALM) method is developed. In theory, we prove that our algorithm is able to converge easily to a stable point of the objective function.

Compared with many state-of-the-art HSI denoising methods, extensive experiments on both simulated and real HSIs demonstrate that our method works well in simultaneously removing the Gaussian noise, impulse noise, deadlines, and stripes, while at the same time preserving image details.

The remainder of this paper is organized as follows. In Section II, the low rank model for HSI denoising is analyzed in detail. Then, we formally introduce the proposed NonLRMA model and the optimization method we develop in Section III. We evaluate the performance of our proposed method in Section IV and conclude this paper in Section V.

II. Background

Due to the significant correlations of the spectra of neighboring pixels, the Casorati matrix often presents the low rank property. That is to say, the subspace spanned by the spectra can be well regularized by the low rank constraint. For better visualization, Fig. 1 shows the low rank property of HSIs.

How closely the noise model fits the real noise environment is crucial to the performance of denoising. Traditionally, most existing HSI denoising methods only consider the additive white Gaussian noise, which is not true in practice. Noise in HSIs normally includes Gaussian noise, impulse noise,

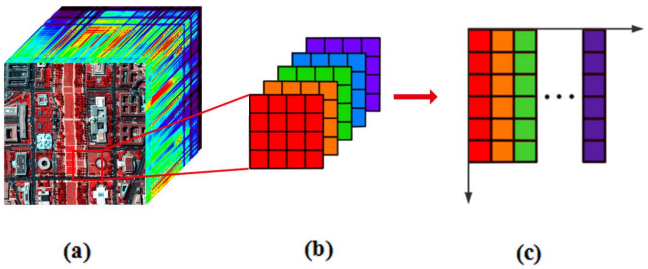


Fig. 1. Illustration of the low rank property. (a) Original Washington DC Mall image. (b) Patch extracted from the Washington DC Mall image. (c) Corresponding Casorati matrix.

deadlines, and stripes, and is usually a mixture of them. To account for this, the degradation model of an HSI can be formulated as

$$\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{S} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{Y}$ is the degraded HSI, $\boldsymbol{L}$ represents the clean HSI, $\boldsymbol{S}$ denotes the impulse noise, deadlines, strips, and so on, and $\boldsymbol{N}$ is the Gaussian noise. Here, $\boldsymbol{Y}, \boldsymbol{L}, \boldsymbol{S}$, and $\boldsymbol{N}$ are the third-order tensors with the same size of $M \times N \times K$, where M, N, and K stand for the width, the height of the image, and the number of spectral bands, respectively. For a subcube of size $q \times q \times K$ centered at the pixel (i, j), a matrix $Y \in \mathbb{R}^{q^2 \times K}$ with a spatial mode and a spectral mode via lexicographical rearrangement can be obtained (the process shown in Fig. 1). Consequently, the corresponding noise degradation model can be converted into the following matrix form:

$$Y = L + S + N \tag{2}$$

where matrices L, S, and N are with the same size as Y. Clearly, L has a low rank. Note that patches are usually overlapped. To deal with the joint regions of neighboring patches and hide possible seams, the pixel value in each band of HSI is then determined by taking the average of the restored patches at this pixel.

In both theoretical and computational aspects, Wright *et al.* [24] and Candès *et al.* [25] presented a convex optimization framework called robust PCA (RPCA) for LRMA. This formulation and its extensions have been successfully used in HSI denoising [6], [26], [28], [29]. RPCA is formulated as

$$\min_{L,S} \ \text{Rank}(L) + \lambda \|S\|_0 \ \ \text{s.t.} \ \ Y = L + S \tag{3}$$

where $\lambda > 0$ is a regularization parameter and $\|S\|_0$ denotes the number of nonzero entries in matrix S. Due to the discrete nature of the rank and the l_0 quasi-norm, solving problem (3) is NP-hard and intractable. A common approach is to use the convex relaxation substituting the rank function and the l_0 norm with the nuclear norm and l_1 norm, respectively, resulting in the following convex optimization problem:

$$\min_{L,S} \ \|L\|_* + \lambda \|S\|_1 \ \ \text{s.t.} \ \ Y = L + S \tag{4}$$

where $\|L\|_* = \sum_i \sigma_i(L)$ and $\sigma_i(L)$ is the ith singular value of the matrix L; $\|S\|_1 = \sum_{i,j} |S_{i,j}|$. Generally speaking, the nuclear norm is a loose approximation of the rank function [28], [36], [44]. Inspired by the low rank property of HSIs, Zhang *et al.* [6] adopted the GoDec model [39] and used iterative hard thresholding to solve the following formulation:

$$\min_{L,S} \ \|Y - L - S\|_F \ \ \text{s.t.} \ \ \text{rank}(L) \le r, \text{card}(S) \le k \tag{5}$$

where r and k denote the upper bound of the rank of L and the cardinality of S, respectively. $\|\cdot\|_F$ is the Frobenius norm. Note that, both L and S need to be known beforehand by this method.

Though great achievements have been made, most previous HSI denoising methods based on low rank approximation suffer from the following aspects. First, the nuclear norm adds all singular values together, which indicates that larger singular values are penalized more heavily than smaller ones, resulting in overshrinking the rank component. Second, the nuclear norm-based approaches generally have poor convergence rate due to many iterations. Third, the rank information needs to be predefined by some methods, which is impractical in practice.

III. Proposed Nonconvex Low Rank Matrix Approximation Model

In this section, we develop a novel NonLRMA model and derive the corresponding HSI denoising algorithm. The key to the proposed method is to use the γ-norm, which can address the aforementioned drawbacks of the nuclear norm. We also show how to optimize the proposed model effectively based on the ALM method. The remainder of this section will elaborate that with our nonconvex low rank regularizer, the rank information is not required to be known beforehand as some previous methods have done [6]. Furthermore, our low rank model works well in capturing the intrinsic rank information.

A. Nonconvex Regularizer and NonLRMA Model

Inspired by the remarkable performance of nonconvex regularizers used for signal and image processing [34]–[36], [44], we introduce the nonconvex matrix rank approximation, which is defined as

$$\text{rank}(L) \approx \|L\|_\gamma = \sum_{i=1}^{\min\{q^2, K\}} \left(1 - e^{-\sigma_i(L)/\gamma}\right) \tag{6}$$

where $\gamma > 0$. Here, we refer to (6) as γ-norm.

Some matrix rank relaxations have been proposed in the literature as shown in Fig. 2. For example, to alleviate the deficiency of the convex relaxation (4), Lingala *et al.* [46] proposed the Schatten p-norm, which could date back to 1993 [34]. Especially, the recent work [28] has the similar idea, which is called the weighted Schatten p-norm. From a mathematical point of view, our nonconvex low-rank regularizer is completely different from the Schatten p-norm. On the other hand, the corresponding HSI denoising approach [28] needs to manually determine the weight set $\{\omega_i\}$, whereas our algorithm does not. From Fig. 2, it is easy to observe that the nuclear norm, the log-determinant function, as well as the Schatten p-norm deviate significantly from 1 when the singular values are greater than 1, indicating that they overshrink the rank component. By contrast, our γ-norm (denoted by the black curve) matches well with the true rank, meaning

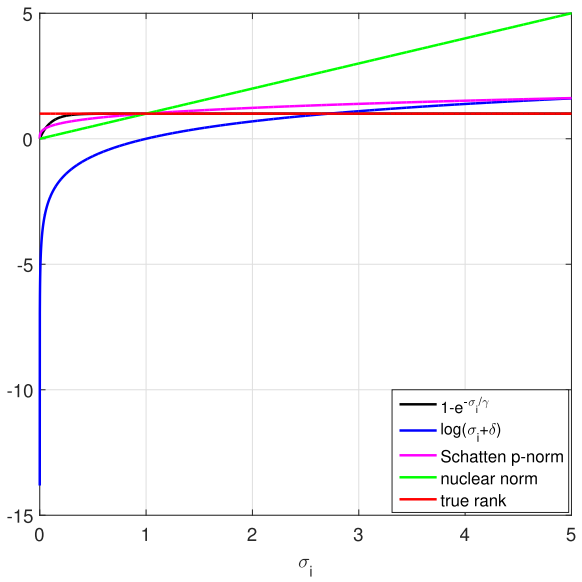


Fig. 2. Approximation of the rank function using different functions. γ, δ, and p are set to 0.1, 10^{-6}, and 0.3, respectively, and the true rank is 1 for nonzero σ_i.

that the γ-norm approximates the rank function better than them. Specifically, the γ-norm is a pseudonorm, but with nice properties as stated by the following proposition.

Proposition 1: Given $L \in \mathbb{R}^{q^2 \times K}$, the γ-norm defined in (6) satisfies the following properties.

1) $\lim_{\gamma \to 0} \|L\|_\gamma = \text{rank}(L)$.
2) $\|L\|_\gamma$ is unitarily invariant, i.e., $\|L\|_\gamma = \|ULV\|_\gamma$ for any orthonormal $U \in \mathbb{R}^{q^2 \times q^2}$ and $V \in \mathbb{R}^{K \times K}$.
3) Positive definiteness: $\|L\|_\gamma \geq 0$ for any $L \in \mathbb{R}^{q^2 \times K}$ and $\|L\|_\gamma = 0$ if and only if $L = \mathbf{0}$.

Proof:

1) For $\phi(\sigma_i(L)) = 1 - e^{-\sigma_i(L)/\gamma}$, we have $\lim_{\gamma \to 0} \phi(\sigma_i(L)) = \begin{cases} 0 & \sigma_i(L) = 0 \\ 1 & \sigma_i(L) \neq 0 \end{cases}$ which consequently reaches the conclusion.
2) Let $(\cdot)^T$, $(\cdot)^{-1}$, and λ denote the transpose, inverse, and eigenvalue of a matrix, separately. E is a unit matrix. Then, the eigenpolynomial $|(ULV)^T ULV - \lambda E| = |V^T L^T U^T ULV - \lambda E| = |V^T L^T LV - \lambda E| = |V^{-1} L^T LV - \lambda E| = |L^T L - \lambda E|$. This means that matrices $(ULV)^T ULV$ and $L^T L$ have the same eigenvalues. As a result, ULV and L have the same singular values. Hence, the conclusion is true.
3) For any $\sigma_i(L)$, we have $\sigma_i(L) \geq 0$. It is, therefore, obvious that $\phi(\sigma_i(L)) \geq 0$ and $\|L\|_\gamma \geq 0$. Finally, the conclusion holds for the reason that $\phi(\sigma_i(L))$ takes 0 if and only if $L = \mathbf{0}$. □

Consequently, we consider the nonconvex γ-norm defined in (6) as the regularization term, Following [47], we adopt the $l_{2,1}$ norm to model the sparse component. Based on the aforementioned two aspects, our proposed NonLRMA model for HSI denoising is formulated as

$$\min_{L,S} \quad \|L\|_\gamma + \lambda \|S\|_{2,1} \quad \text{s.t.} \quad \|Y - L - S\|_F \leq \delta \tag{7}$$

where $\|S\|_{2,1} = \sum_{j=1}^{K} (\sum_{i=1}^{q^2} S_{ij}^2)^{1/2}$ and δ is a constant related to the standard deviation of the independent and identically distributed Gaussian noise N. The NonLRMA model (7) can be regarded as an extension of (4). Whereas compared with (4), the challenge of solving NonLRMA model lies in the fact that a global solution is difficult to obtain due to the nonconvexity of (7). As a result, the subgradient method is no longer applicable. In Section III-B, we prove that each subproblem of (7) has a closed-form solution despite of its nonconvexity.

B. ALM-NonLRMA Optimization

It is easy to deduce that the augmented Lagrangian function for problem (7) is

$$\mathcal{L}(L, S, \Lambda; \rho) = \|L\|_\gamma + \lambda \|S\|_{2,1} + \langle \Lambda, Y - L - S \rangle + \frac{\rho}{2} \|Y - L - S\|_F^2 \tag{8}$$

where $\rho > 0$ is called the penalty parameter, $\Lambda \in \mathbb{R}^{q^2 \times K}$ is the Lagrangian multiplier, and $\langle \cdot \rangle$ denotes the standard trace inner product, i.e., $\langle A, B \rangle = \text{trace}(A^T B)$. Following the idea of the Gauss–Seidel method, we can update L and S alternatively by keeping one of them fixed at its latest value and then update the Lagrange multiplier, that is:

$$L_{k+1} = \arg\min_L \mathcal{L}(L, S_k, \Lambda_k; \rho_k) \tag{9}$$

$$S_{k+1} = \arg\min_S \mathcal{L}(L_{k+1}, S, \Lambda_k; \rho_k) \tag{10}$$

$$\Lambda_{k+1} = \Lambda_k + \rho_k (Y - L_{k+1} - S_{k+1}) \tag{11}$$

$$\rho_{k+1} = \min\{\beta * \rho_k, \rho_{\max}\} \tag{12}$$

where L_k denotes L in the kth iteration and β is set to 1.1. With simple manipulation, we have the following iterative scheme:

$$L_{k+1} = \arg\min_L \|L\|_\gamma + \frac{\rho_k}{2} \|L - \left(Y - S_k + \frac{\Lambda_k}{\rho_k}\right)\|_F^2 \tag{13}$$

$$S_{k+1} = \arg\min_S \lambda \|S\|_{2,1} + \frac{\rho_k}{2} \|S - \left(Y - L_{k+1} + \frac{\Lambda_k}{\rho_k}\right)\|_F^2. \tag{14}$$

Here, we refer to subproblems (13) and (14) as L-subproblem and S-subproblem, separately. Obviously, the above-mentioned iterative scheme exploits the separable property of our objective function.

Let $\sigma_1^k \geq \sigma_2^k \geq \cdots \geq \sigma_s^k$ represent the singular values of L_k with $s = \min\{q^2, K\}$ and $\nabla\phi(\sigma_i^k)$ denote the gradient of ϕ at σ_i^k. Let $f(L) = (1/2)\|L - D_k\|_F^2$ with $D_k = Y - S_k + (\Lambda_k/\rho_k)$. It is easy to prove that the gradient of $f(L)$ is Lipschitz continuous by setting the Lipschitz constant being 1. From Fig. 2, we can observe that the nonconvex regularizer is continuous, concave, smooth, differentiable, and monotonically increasing on $[0, +\infty)$, and its gradient is nonnegative and monotonically decreasing. Considering the nonascending order of singular values and according to the antimonotone property of gradient of our nonconvex function, we have

$$0 \leq \nabla\phi(\sigma_1^k) \leq \nabla\phi(\sigma_2^k) \leq \cdots \leq \nabla\phi(\sigma_s^k) \tag{15}$$

$$\phi(\sigma_i(L)) \leq \phi(\sigma_i^k) + \nabla\phi(\sigma_i^k)(\sigma_i(L) - \sigma_i^k). \tag{16}$$

The above-mentioned result (16) can be easily proved by the supergradient definition of the concave function.

Following (16), we solve the following relaxation problem:

$$L_{k+1} = \arg\min_{L} \frac{1}{\rho_k}\sum_{i=1}^{s}\phi\left(\sigma_i^k\right) + \nabla\phi\left(\sigma_i^k\right)\left(\sigma_i(L)-\sigma_i^k\right) + f(L)$$
$$= \arg\min_{L} \frac{1}{\rho_k}\sum_{i=1}^{s}\nabla\phi\left(\sigma_i^k\right)\sigma_i(L) + \frac{1}{2}\|L - D_k\|_F^2. \quad (17)$$

Then, following [44], the optimum solution of L-subproblem can be efficiently obtained by resorting to recently proposed generalized weight singular value thresholding (WSVT) [48], as shown in Lemma 1. Solving (17) is equivalent to computing the proximity operator of the weighted nuclear norm. It is worth noting that the weight $\nabla\phi(\sigma_i^k)$ in problem (17) is automatically calculated. However, the weight in the weighted Schatten p-norm [28] must be given in advance. More specifically, L is a low rank matrix representing the clean image.

Lemma 1: For any $(1/\rho_k) > 0$, the given data $D_k = Y - S_k + (\Lambda_k/\rho_k)$, and $0 \le \nabla\phi(\sigma_1^k) \le \nabla\phi(\sigma_2^k) \le \cdots \le \nabla\phi(\sigma_s^k)$, a globally optimal solution L^* to problem (17) is given by the WSVT [48]

$$L^* = U\mathcal{S}_{\frac{\nabla\phi}{\rho_k}}(\Sigma)V^T \quad (18)$$

where $D_k = U\Sigma V^T$ is the SVD of D_k and $\mathcal{S}_{(\nabla\phi/\rho_k)}(\Sigma) = \text{Diag}\{\max(\Sigma_{ii} - (\nabla\phi(\sigma_i^k)/\rho_k), 0)\}$.

The solution of S-subproblem is given by Lemma 2.

Lemma 2: Given a matrix $W_k = Y - L_{k+1} + (\Lambda_k/\rho_k)$, the S-subproblem has an optimal solution S^* and its jth column is [49]

$$S^*(:,j) = \begin{cases} \dfrac{\|W_k(:,j)\|_2 - \dfrac{\lambda}{\rho_k}}{\|W_k(:,j)\|_2} W_k(:,j), & \text{if } \dfrac{\lambda}{\rho_k} < \|W_k(:,j)\|_2; \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

It is well known that $l_{2,1}$-norm is a sparsity-inducing norm defined as the l_1-norm of the columns of the sparse noise matrix S. Embodying the specific structures in HSIs, the norm is vital to deal with sample-specific corruptions and outliers, such as stripes and deadlines.

Algorithm 1 shows the pseudocode of our specifically designed algorithm for optimizing the NonLRMA model (7).

C. Convergence Analysis

We theoretically analyze the convergence of the proposed algorithm with the main result presented in Theorem 1. To introduce Theorem 1, we first give Lemmas 3–5, which provide some sufficient conditions under which Theorem 1 holds.

Lemma 3: $\mathcal{L}(L, S, \Lambda; \rho)$ is monotonically decreasing with respect to variable L.

Proof: In iteration $k+1$, L_{k+1} is a global solution to problem (17). We then have

$$\frac{1}{\rho_k}\sum_{i=1}^{s}\nabla\phi(\sigma_i^k)\sigma_i^{k+1} + \frac{1}{2}\|L_{k+1} - D_k\|_F^2 \le \frac{1}{\rho_k}\sum_{i=1}^{s}\nabla\phi(\sigma_i^k)\sigma_i^k + \frac{1}{2}\|L_k - D_k\|_F^2. \quad (20)$$

Algorithm 1 ALM-NonLRMA for HSI denoising

Input: Noisy HSI Y.
Initialize: $\gamma,\ \epsilon,\ \lambda,\ \beta,\ \rho_0,\ S_0 = \Lambda_0 = \mathbf{0},\ k = 0$.
1: **while** not converged **do**
2: Compute $D_k = Y - S_k + \frac{\Lambda_k}{\rho_k}$;
3: Update L_{k+1} by (18);
4: Compute $W_k = Y - L_{k+1} + \frac{\Lambda_k}{\rho_k}$;
5: Update S_{k+1} by (19);
6: Update Λ_{k+1} by (11);
7: Update ρ_{k+1} by (12);
8: Check the convergence condition
9: $\|Y - L_{k+1} - S_{k+1}\|_F \le \epsilon * \|Y\|_F$.
10: **end while**
11: Compute the Gaussian noise $N = Y - L_k - S_k$;
Output: Denoised HSI L_k.

It can be transformed to

$$\frac{1}{2}\|L_{k+1} - D_k\|_F^2 - \frac{1}{2}\|L_k - D_k\|_F^2 \le \frac{1}{\rho_k}\sum_{i=1}^{s}\nabla\phi(\sigma_i^k)(\sigma_i^k - \sigma_i^{k+1}). \quad (21)$$

Second, summing (16) for $i = 1, 2, \ldots, s$

$$\sum_{i=1}^{s}(\phi(\sigma_i^{k+1}) - \phi(\sigma_i^k)) \le \sum_{i=1}^{s}\nabla\phi(\sigma_i^k)(\sigma_i^{k+1} - \sigma_i^k). \quad (22)$$

Now, we add (21) and (22) together

$$\mathcal{L}(L_{k+1}, S_k, \Lambda_k; \rho_k) - \mathcal{L}(L_k, S_k, \Lambda_k; \rho_k) \le 0. \quad (23)$$

This reaches the conclusion that $\mathcal{L}(L, S, \Lambda; \rho)$ is monotonically decreasing with respect to variable L. □

Lemma 4: The sequence $\{\Lambda_k\}$ is bounded.

Proof: The optimal S_{k+1} needs to satisfy the first-order optimality condition, that is

$$0 \in \partial_S\mathcal{L}(L_{k+1}, S, \Lambda_k; \rho_k)|_{S_{k+1}} = \partial_S(\lambda\|S\|_{2,1})|_{S_{k+1}} - \Lambda_k - \rho_k(Y - L_{k+1} - S_{k+1}) = \partial_S(\lambda\|S\|_{2,1})|_{S_{k+1}} - \Lambda_{k+1}. \quad (24)$$

It can easily be proved that $\partial_S(\lambda\|S\|_{2,1})|_{S_{k+1}}$ is bounded, since

$$\partial_S(\|S\|_{2,1})|_{S_{k+1}} = \begin{cases} 0, & \text{if } \|S_{k+1}(:,j)\|_2 = 0 \\ \dfrac{S_{k+1}(:,j)}{\|S_{k+1}(:,j)\|_2}, & \text{otherwise.} \end{cases}$$

Thus, $\{\Lambda_k\}$ is bounded. □

Lemma 5: Sequences $\{L_k\}$ and $\{S_k\}$ are bounded if $\sum_{i=1}^{\infty}(\rho_i + \rho_{i+1}/(\rho_i)^2) < \infty$.

Proof: With simple manipulation, we have

$$\mathcal{L}(L_k, S_k, \Lambda_k; \rho_k) = \mathcal{L}(L_k, S_k, \Lambda_{k-1}; \rho_{k-1}) + \frac{\rho_k - \rho_{k-1}}{2}\|Y - L_k - S_k\|_F^2 + \langle\Lambda_k - \Lambda_{k-1}, Y - L_k - S_k\rangle = \mathcal{L}(L_k, S_k, \Lambda_{k-1}; \rho_{k-1}) + \frac{\rho_k + \rho_{k-1}}{2(\rho_{k-1})^2}\|\Lambda_k - \Lambda_{k-1}\|_F^2. \quad (25)$$

The above-mentioned equation can be obtained by using (11). Then

$$\begin{aligned}&\mathcal{L}(L_{k+1}, S_{k+1}, \Lambda_k; \rho_k)\\ &\quad\le \mathcal{L}(L_{k+1}, S_k, \Lambda_k; \rho_k)\\ &\quad\le \mathcal{L}(L_k, S_k, \Lambda_k; \rho_k)\\ &\quad= \mathcal{L}(L_k, S_k, \Lambda_{k-1}; \rho_{k-1}) + \frac{\rho_k + \rho_{k-1}}{2(\rho_{k-1})^2}\|\Lambda_k - \Lambda_{k-1}\|_F^2.\end{aligned} \tag{26}$$

Since S_{k+1} is a global solution to (14), we get the first inequation. The second inequation is given by Lemma 3. Iterating over the inequality chain (26) for k times, we have

$$\begin{aligned}&\mathcal{L}(L_{k+1}, S_{k+1}, \Lambda_k; \rho_k)\\ &\quad\le \mathcal{L}(L_1, S_1, \Lambda_0; \rho_0) + \sum_{i=1}^{k}\frac{\rho_i + \rho_{i-1}}{2(\rho_{i-1})^2}\|\Lambda_i - \Lambda_{i-1}\|_F^2\\ &\quad\le \mathcal{L}(L_1, S_1, \Lambda_0; \rho_0) + C\sum_{i=1}^{k}\frac{\rho_i + \rho_{i-1}}{2(\rho_{i-1})^2}\end{aligned} \tag{27}$$

where C is an upper bound of $\{\|\Lambda_i - \Lambda_{i-1}\|_F^2\}$. Under the condition that $\sum_{i=1}^{\infty}(\rho_i + \rho_{i+1}/(\rho_i)^2) < \infty$, all terms on the right-hand side of the above-mentioned inequality are bounded; thus, $\mathcal{L}(L_{k+1}, S_{k+1}, \Lambda_k; \rho_k)$ is upper bounded. Add $(1/2\rho_k)\|\Lambda_k\|_F^2$ to the augmented Lagrangian function (8)

$$\begin{aligned}&\mathcal{L}(L_{k+1}, S_{k+1}, \Lambda_k; \rho_k) + \frac{1}{2\rho_k}\|\Lambda_k\|_F^2\\ &= \|L_{k+1}\|_\gamma + \lambda\|S_{k+1}\|_{2,1} + \frac{\rho_k}{2}\left\|Y - L_{k+1} - S_{k+1} + \frac{\Lambda_k}{\rho_k}\right\|_F^2.\end{aligned}$$

According to Lemma 4, the left-hand side is bounded. Thus, each term on the right-hand side is bounded; then, $\{L_k\}$ and $\{S_k\}$ are bounded. □

Till now, we have shown the boundedness of $\{L_k, S_k, \Lambda_k\}$. Next, we will present the main theoretical result in Theorem 1.

Theorem 1: Let $\{L_k, S_k, \Lambda_k\}$ be the sequence generated by Algorithm 1 and $\{L^*, S^*, \Lambda^*\}$ be an accumulation point. Then, $\{L^*, S^*\}$ is a stationary point of the original problem (7) as long as $\sum_{i=1}^{\infty}(\rho_i + \rho_{i+1}/(\rho_i)^2) < \infty$ and $\lim_{k\to\infty}\rho_k(S_k - S_{k+1}) = 0$.

Proof: From Lemmas 4 and 5, we know that the sequence $\{L_k, S_k, \Lambda_k\}$ is bounded. By Bolzano–Weierstrass theorem, the sequence has at least one accumulation point, denoted as $\{L^*, S^*, \Lambda^*\}$. Without loss of generality, we assume that $\{L_k, S_k, \Lambda_k\}$ converges to $\{L^*, S^*, \Lambda^*\}$. Next, we prove that this accumulation point is a stationary point of problem (7).

From (11) and the boundedness of $\{\Lambda_k\}$, we have $\lim_{k\to\infty} Y - L_{k+1} - S_{k+1} = \lim_{k\to\infty}(\Lambda_{k+1} - \Lambda_{k+1}/\rho_k) = 0$; then

$$Y = L^* + S^*. \tag{28}$$

By the first-order optimality condition, we have

$$\begin{aligned}&\partial_L\mathcal{L}(L, S_k, \Lambda_k; \rho_k)|_{L_{k+1}}\\ &\quad= \partial_L(\|L\|_\gamma)|_{L_{k+1}} - \Lambda_k - \rho_k(Y - L_{k+1} - S_k)\\ &\quad= \partial_L(\|L\|_\gamma)|_{L_{k+1}} - \Lambda_{k+1} + \rho_k(S_k - S_{k+1})\\ &\quad= 0.\end{aligned}$$

Due to the assumption of $\lim_{k\to\infty}\rho_k(S_k - S_{k+1}) = 0$, we have

$$\partial_L(\|L^*\|_\gamma) - \Lambda^* = 0. \tag{29}$$

Following (24), we get

$$\partial_S(\lambda\|S^*\|_{2,1}) - \Lambda^* = 0. \tag{30}$$

From (28)–(30), we can conclude that $\{L^*, S^*, \Lambda^*\}$ satisfies the Karush—Kuhn–Tucker conditions of $\mathcal{L}(L, S, \Lambda; \rho)$. $\{L^*, S^*\}$ is thus a stationary point of the original problem (7). □

D. Connection With Existing Work

We further emphasize the connection and difference between the proposed method and some existing approaches in the literature for LRMA, including the TNN [37] and iteratively reweighted nuclear norm (IRNN) [44] for matrix completion.

1) It is a common observation that those larger singular values quantify the information of underlying principal directions. The TNN, however, discards the sum of those larger singular values. Furthermore, this approach needs outer iterations and additional SVD computation, resulting in extra, often unnecessary computation load and storage. On the other hand, it is hard to determine the number of singular values.
2) IRNN uses a family of nonconvex surrogates to approximate the rank function, leading to more accurate results for LRMR. However, it remains unclear which choice is optimal for the particular nonconvex function and how to determine the additional tuning parameters. In this paper, we will give some strategies for better recovery, which will be discussed in Section IV. Moreover, IRNN mainly deals with matrix completion, which aims to recover a low rank matrix from partial observations without considering the non-Gaussian noise, or, for each image, it needs some pixels not corrupted by non-Gaussian noise to ensure successful LRMR.

In addition, TNN and IRNN apply nonconvex low rank models to the whole image data, without considering the local similarity. On the other hand, the noise intensities in different bands of an HSI may differ much. Thus, applying TNN or IRNN to each band uniformly would inevitably incur blur or the loss of local details. These issues inspire us to develop a patch-based algorithm, facilitating the detection of mixed noise.

IV. Experimental Results and Discussion

We conduct extensive experiments to verify the effectiveness of our proposed NonLRMA method and show the results in comparison with some existing competing HSI denoising approaches: GoDec-based LRMR [6], the video block matching 3-D filtering (VBM3D) [14], noise-adjusted LRMA (NAILRMA) [27], and weighted Schatten p-norm-based LRMA (WSN-LRMA) [28]. Note that the last three are recently developed state-of-the-art HSI denoising methods and belong to the low rank-based scheme. All the codes are provided by the authors. Before denoising, the gray values of each band are scaled to the interval [0, 1], and they are

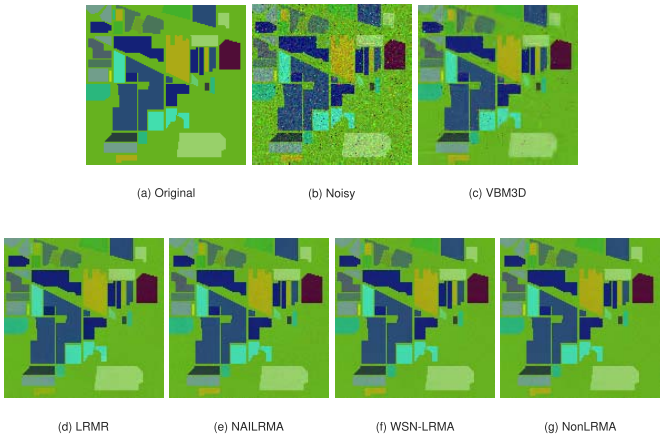


Fig. 3. Restoration results on synthetic data: moderate noise level. (a) Original false-color image (R:6, G:88, B:221). (b) Noisy image. The resorted image obtained by (c) VBM3D, (d) LRMR, (e) NAILRMA, (f) WSN-LRMA, and (g) NonLRMA. The figure is better viewed in a zoomed-in PDF.

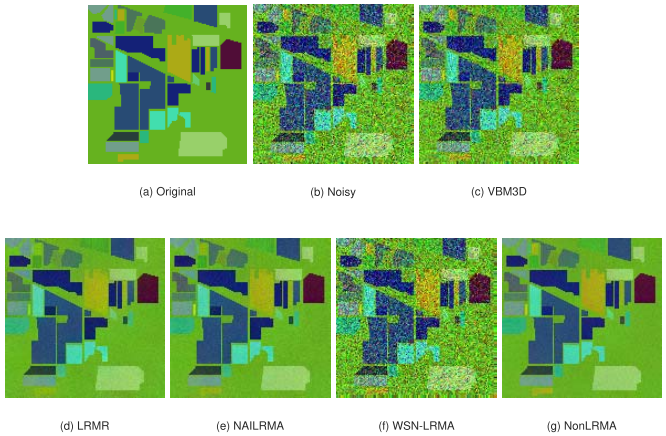


Fig. 4. Restoration results on synthetic data: severe noise level. (a) Original false-color image (R:6, G:88, B:221). (b) Noisy image. The resorted image obtained by (c) VBM3D, (d) LRMR, (e) NAILRMA, (f) WSN-LRMA, and (g) NonLRMA. The figure is better viewed in a zoomed-in PDF.

stretched back to the original level after denoising. All the experiments were run in MATLAB R2015a on a laptop with a CORE i7 3.40-GHz CPU and 16-GB memory. For the purpose of reproductivity, we provide our source code at https://www.researchgate.net/publication/315725239_Code_of_NonLRMA.

A. Simulated Data Experiment for Mixed-Noise Removal

In this section, we first compare all the methods on synthetic data, which is generated by the ground truth of the Indian Pines data set [50] in a similar way to [29]. The size of the synthetic image is $145 \times 145 \times 224$, and the reflectance values of all the voxels in the HSI are linearly mapped to [0, 1]. More detailed description of the synthetic data can be found in [29].

In this experiment, two typical kinds of noises are added to all the bands of the ground truth, which are briefly described as follows.

1) Since, in practice, noise intensity in different bands is often different, two degrees of zero-mean Gaussian noise are added. To simulate the level of moderate noise, the signal-to-noise ratio (SNR) value of each band varies from 15 to 25 dB randomly, and the mean SNR value of all the bands is 20.43 dB. For the level of severe noise, the SNR value of each band varies from 45 to 55 dB randomly, and the mean SNR value of all the bands is 49.75 dB.
2) Salt-and-pepper impulse noise is added to all the bands with 20% pixels of each band being corrupted. The impulse noise intensity of each band varies from 0.0196 to 0.0784, and the mean intensity is 0.0492.

Three commonly used measures are adopted for quantitative comparison, including peak SNR (PSNR), structural similarity (SSIM) [51] measuring spatial similarity between the denoised HSI and the original one, and feature similarity (FSIM) [52] emphasizing perceptual consistency with the original HSI. In general, the larger these three measures are, the closer the denoised HSI is to the ground truth. Finally, the mean value of each measure for the whole HSI is calculated and denoted as MPSNR, MSSIM, and MFSIM, separately. In this simulated experiment, we first adopt the same parameter setting to the original papers [6], [28] and then manually adjust them for better visual effects. The parameter setting of our method is shown in Table III.

The statistical data reporting the performance of all the methods are shown in Table I. The highest MPSNR, MSSIM, and MFSIM values are highlighted in bold, and the second-best results for each quality index are underlined. From Table I, we can see that our proposed method achieves the highest MPSNR under both moderate and severe noise levels. It achieves 0.812- and 1.499-dB improvement over the NAILRMA method, respectively. For the moderate noise level, in comparison with LRMR, our proposed method yields comparable or even slightly better results in terms of MSSIM and MFSIM. For the severe noise level, it yields best results over other competing methods, which validates the effectiveness of the nonconvex regularizer for removing mixed noise. In addition, some typical bands of denoised HSIs obtained by different methods are shown in Figs. 3 and 4. Fig. 3 displays the denoising results of different methods in the moderate noise case. Obviously, some residual impulse noise can still be observed in the restored image by VBM3D as shown in Fig. 3(c), while the results by LRMR, NAILRMA, WSN-LRMA, and NonLRMA are visually similar, which agrees with the quantitative comparison in Table I. For the severe noise level, the denoising results of different approaches are shown in Fig. 4. The results of VBM3D and WSN-LRMA cannot remove the mixed noise. LRMR and NAILRMA remove the noise and preserve the spectral information to some extent, but the Gaussian noise cannot be completely removed. In Fig. 4(f), however, our proposed method achieves the best visual performance in comparison with the restoration results by other methods. For quantitative comparison on details, we also calculate the PSNR, SSIM, and FSIM values of each band as shown in Figs. 5 and 6. We can observe that in almost all the bands, the PSNR, SSIM, and FSIM values of the proposed method (the black curves) are higher than those of the other methods under both moderate and severe noise

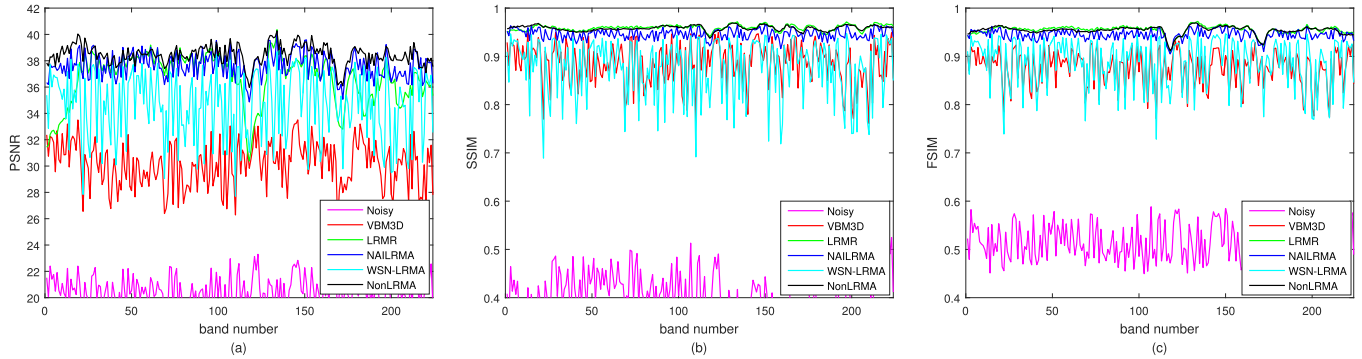


Fig. 5. PSNR, SSIM, and FSIM values of each band of the experimental results: moderate noise level. (a) PSNR values. (b) SSIM values. (c) FSIM values.

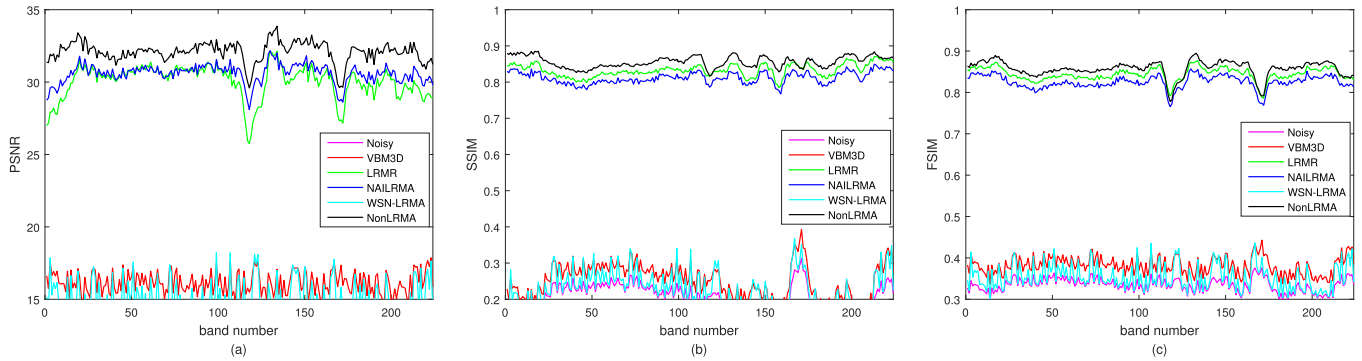


Fig. 6. PSNR, SSIM, and FSIM values of each band of the experimental results: severe noise level. (a) PSNR values. (b) SSIM values. (c) FSIM values.

TABLE I

QUANTITATIVE EVALUATION OF THE DIFFERENT RESTORATION ALGORITHMS WITH THE SYNTHETIC DATA

	Moderate Noise Level					
	Noisy	VBM3D	LRMR	NAILRMA	WSN-LRMA	NonLRMA
MPSNR(dB)	20.526	30.129	36.657	37.643	34.687	**38.455**
MSSIM	0.405	0.885	0.955	0.947	0.876	**0.959**
MFSIM	0.514	0.875	0.952	0.945	0.887	**0.954**
	Severe Noise Level					
	Noisy	VBM3D	LRMR	NAILRMA	WSN-LRMA	NonLRMA
MPSNR(dB)	14.026	16.151	30.035	30.589	14.925	**32.088**
MSSIM	0.200	0.250	0.833	0.812	0.226	**0.856**
MFSIM	0.334	0.381	0.844	0.825	0.354	**0.857**

levels, meaning that our proposed method outperforms other methods for mixed-noise removal.

Figs. 7 and 8 visualize the spectral signatures before and after restoration. Under a moderate noise level, Fig. 7 presents the spectral signatures of the pixel (100, 100). It can be easily seen that the curve shape of the spectral signatures acquired by the proposed NonLRMA method [see Fig. 7(g)] closely resembles the shape of the spectral signatures of the noise-free spectrum [Fig. 7(a)], which further verifies our effectiveness. Similar conclusion can be reached in Fig. 8, which compares the shapes of spectral signatures in the case of a severe noise level.

Another typical kind of noise in HSIs is stripes, mainly due to inconsistent responses between different detectors. To test the performance on stripe removal, we experiment on the Hyperspectral Digital Imagery Collection Experiment (HYDICE) image of the Washington DC Mall [53]. Due to page limitation, a subimage of size $256\times256\times11$ is selected here. We randomly add periodical stripes to five different bands. From the restoration results shown in Fig. 9, we can easily observe that most competing methods cannot remove stripes completely. By contrast, the proposed method obtains comparative visual results in removing periodical stripes as shown in Fig. 9(g). General speaking, NonLRMA employs the γ-norm instead of the traditional nuclear norm to regularize the spectrum component and adopts the $l_{2,1}$ norm to model the sparse component, providing remarkably discriminative ability on mixed noise.

B. Real Data Experiments

In this section, we conduct experiments on real data, including HYDICE urban and EO-1 Hyperion Australia data

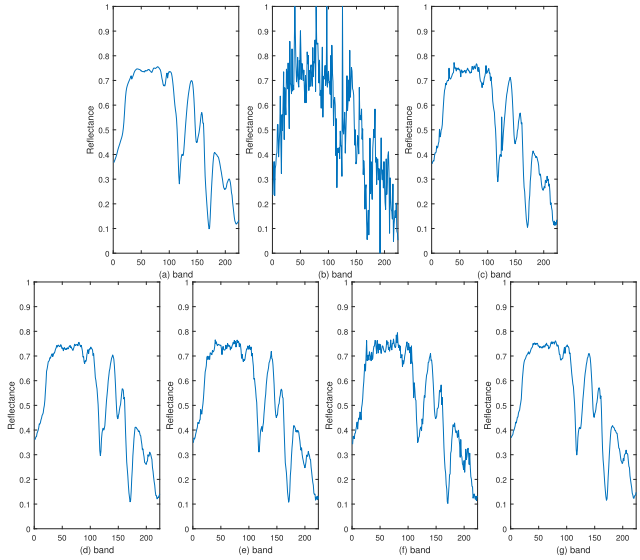


Fig. 7. Spectrum of pixel (100, 100) in the restoration results: moderate noise level. (a) Original. (b) Noisy. (c) VBM3D. (d) LRMR. (e) NAILRMA. (f) WSN-LRMA. (g) NonLRMA. The figure is better viewed in a zoomed-in PDF.

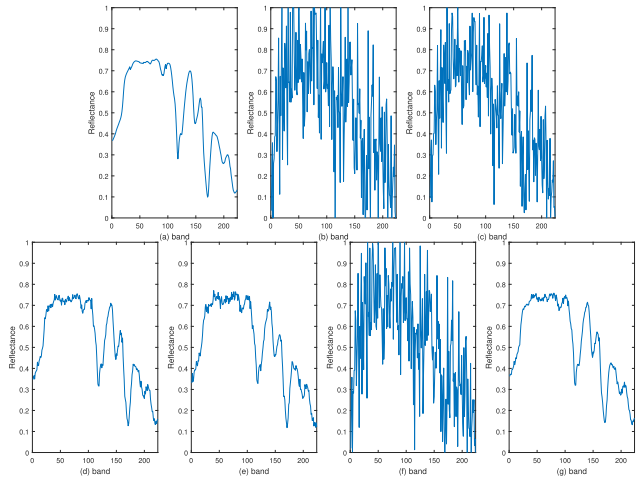


Fig. 8. Spectrum of pixel (100, 100) in the restoration results: severe noise level. (a) Original. (b) Noisy. (c) VBM3D. (d) LRMR. (e) NAILRMA. (f) WSN-LRMA. (g) NonLRMA. The figure is better viewed in a zoomed-in PDF.

sets, to investigate the performance of the proposed method. Detailed experimental settings and results on these data are given as follows.

1) HYDICE Urban Data Set: The HYDICE urban images [54] of size $307 \times 307 \times 210$ are corrupted by stripes, deadlines, atmosphere, water absorption, and other unknown noise (see Figs. 11 and 12). Unlike [6] and [28] where several bands polluted by atmosphere and water absorption are removed, we select all images as the test data, which makes the denoising task more challenging. We follow the experimental settings stated in [6], [28], and [55]. For the benchmark VBM3D method, the noise variation is set to 12 as provided in the paper [6]. The patch size, step size, rank of L, and the cardinality of S in LRMR are set to 20, 4, 4, and 4000, separately. For NAILRMA, the corresponding parameters are set to 20, 4, 7, and 0.15, respectively. The parameters in

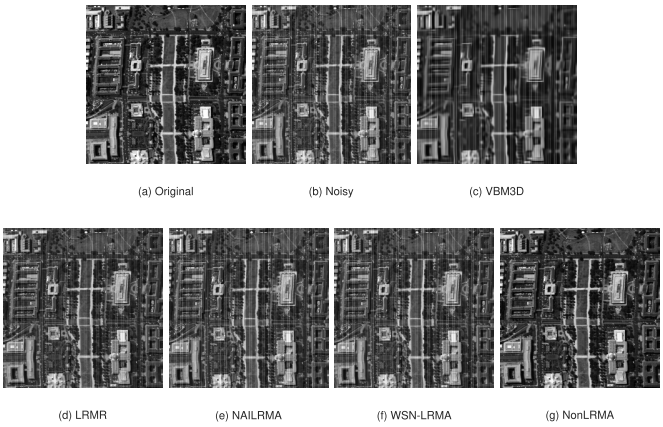


Fig. 9. Restoration results on synthetic data: Washington DC Mall. (a) Original image. (b) Noisy image. The resorted image obtained by (c) VBM3D, (d) LRMR, (e) NAILRMA, (f) WSN-LRMA, and (g) NonLRMA. The figure is better viewed in a zoomed-in PDF.

WSN-LRMA including the weights for penalizing low rank and sparse constraint, power, number of iterations, step size, and patch size are set to 0.007, 1.2, 0.7, 3, 4, and 20, respectively. For this test, we report the parameters of NonLRMA in Table III.

Figs. 10–12 show the denoising results on bands 87, 207, and 108 on the HYDICE urban data set. In general, it is observed that our method can simultaneously remove the Gaussian noise, impulse noise, deadlines, and stripes from the noisy images, and preserve underlying image structures well. For the slight noise mixed by Gaussian noise and stripes as shown in Fig. 10, LRMR and NAILRMA can remove the mixed noise, but VBM3D tends to oversmooth the result. In addition, the WSN-LRMA method is also able to obtain nearly the best visual representation. Nevertheless, in Fig. 11, where the noise is mixed by Gaussian noise, deadlines, and stripes simultaneously, it can be seen that there still exist lots of noises in the results of VBM3D, LRMR, and NAILRMA. WSN-LRMA achieves relatively better results, but also introduces artifacts into the final results. As shown in Fig. 12, other competing methods perform badly when the HSI is heavily polluted by mixed noise. Our method, by contrast, can recover most details; this is mainly because the nonconvex regularizer is a tighter approximation to the original matrix rank function.

For a more detailed comparison, we show the vertical and horizontal mean profiles of band 207 in this data set in Figs. 13 and 14, and the horizontal and vertical axes represent the column (resp. row) number and the corresponding mean digital number value in Fig. 13 (resp. Fig. 14), respectively. It is clearly observed that the curves of the original image fluctuate significantly due to the existence of mixed noise. From Fig. 13(c)–(e), it is observed that LRMR, NAILRMA, and WSN-LRMA have removed stripes in the vertical direction, but failed to remove stripes in the horizontal direction. As shown in Fig. 11, our proposed method provides evidently smoother curve, indicating that the stripes in the vertical and horizontal directions have been removed more effectively. Specially, VBM3D, LRMR, and NAILRMA fail to remove the thick vertical white line in the 261th column, as shown in the highlighted red rectangle in Fig. 13(b)–(d).

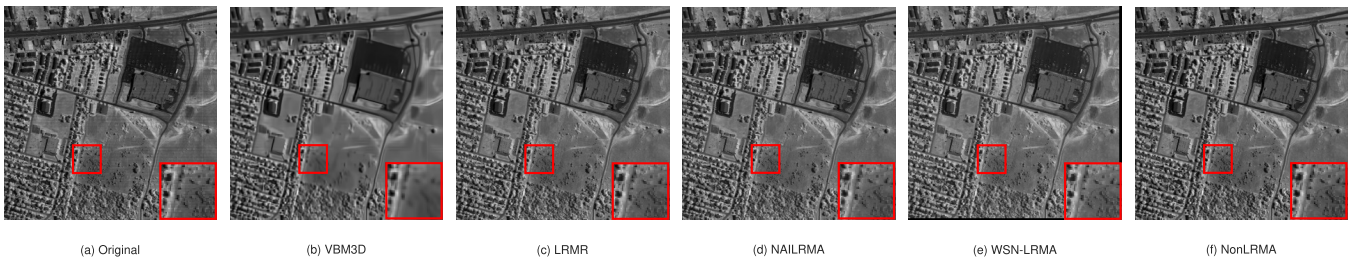


Fig. 10. Restoration results on HYDICE urban data set: slight noise band. (a) Original image located at the 87th band. Resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

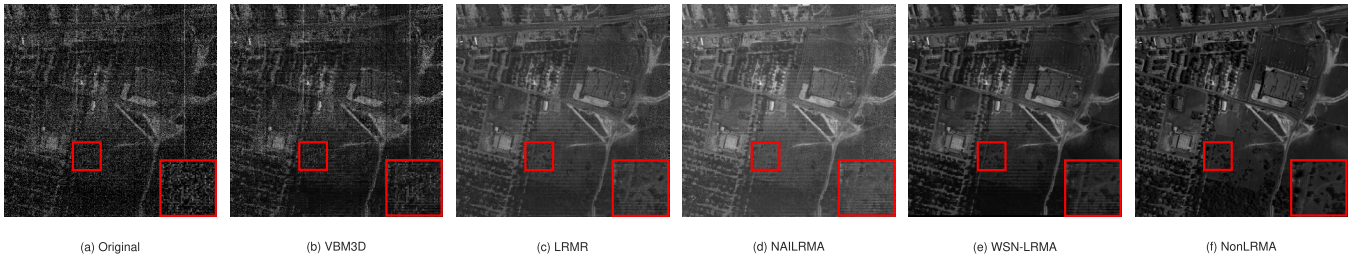


Fig. 11. Restoration results on HYDICE urban data set: moderate noise band. (a) Original image located at the 207th band. Resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

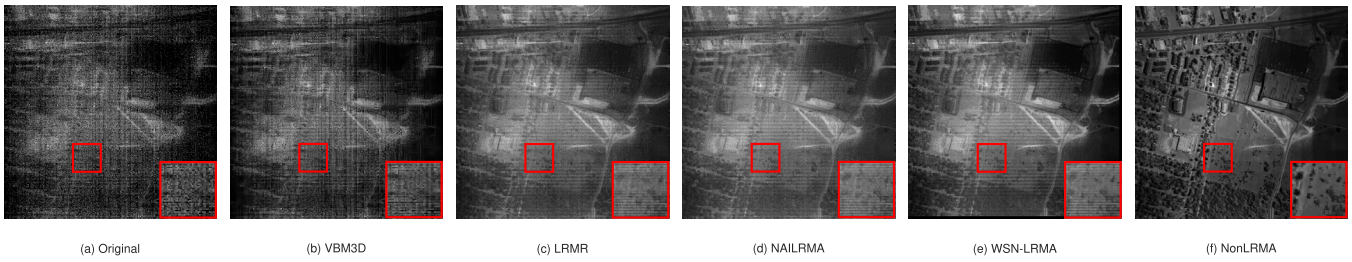


Fig. 12. Restoration results on HYDICE urban data set: severe noise band. (a) Original image located at the 108th band. Resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

2) EO-1 Hyperion Australia Data Set: The Hyperion image [56] of size $3858 \times 256 \times 242$ is captured on December 4, 2010. Similar to [28], after removing the overlapping bands between visual near-infrared and shortwave infrared ranges, only a subregion of size $400 \times 200 \times 150$ is used in our experiment.

Some representative denoising results are shown in Figs. 15–17. It is observed that VBM3D fails to remove heavy noise and suffers from oversmoothing the results on all these HSIs. As shown in Figs. 15 and 16, LRMR and NAILRMA remove part of stripes, but deadlines are left; hence, the denoising performance is far from satisfactory. Similar observation was also reported in [28]. An intuitive explanation is that the GoDec-based LRMR using bilateral random projection instead of computing SVD cannot guarantee a unique solution, since different random projections may yield different results. Moreover, LRMR and NAILRMA need the upper bound of matrix rank to be estimated as prior, which, unfortunately, is usually difficult to achieve. Such failures, again, reveal the challenge in denoising HSIs. It is promising to observe that WSN-LRMA and NonLRMA have recovered images with significantly improved visual quality. Though much better than VBM3D, LRMR, and NAILRMA, compared with NonLRMA, WSN-LRMA still leaves some stripes in the resulting images, which is not as expected. For example, as shown in Figs. 15–17, the superior performance of the proposed method can be easily observed in the highlighted regions, where our method removes all types of noises and stripes while others do not. To better illustrate this, we show the vertical and horizontal mean profiles of band 51 in Figs. 18 and 19, respectively. From Figs. 18 and 19, especially in the highlighted regions, it is easy to see that NonLRMA has the best performance whereas WSN-LRMA does not remove all stripes both vertically and horizontally. Moreover, WSN-LRMA is more computationally expensive, which will be clear in later subsection IV-C. These observations strongly reveal the effectiveness of our NonLRMA by adopting γ-norm.

C. Model Analysis

We first discuss how to set parameters and analyze the sensitivity of parameters in our model, including the regularization parameter λ, penalty parameter ρ_0, parameter γ, and patch size q. We then report the running time for the HSI denoising methods on two real-world HSIs. Please note that the experiments used for parameter setting are based on the simulated data under the moderate noise level.

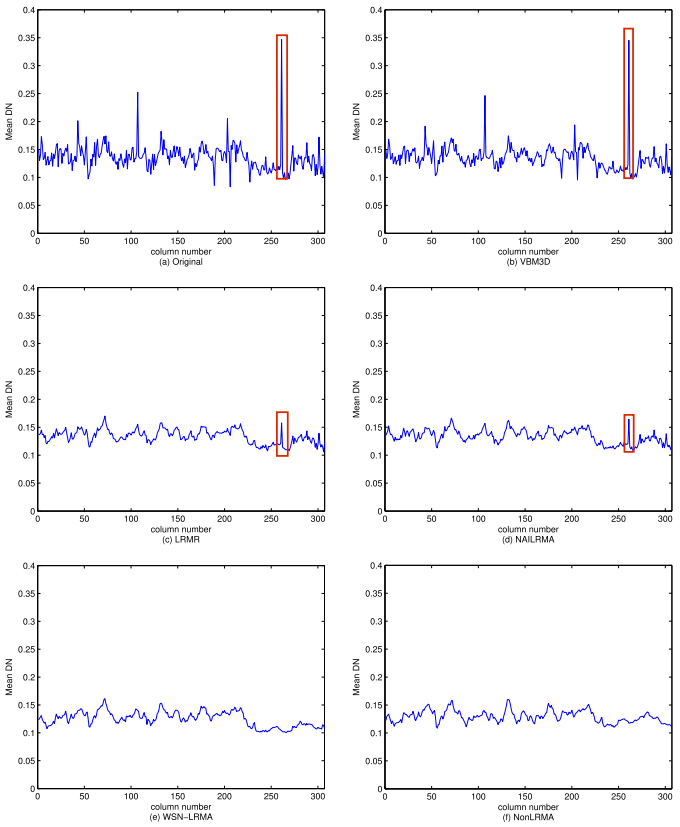


Fig. 13. Vertical mean profiles of band 207 in HYDICE urban data set. (a) Original. (b) VBM3D. (c) LRMR. (d) NAILRMA. (e) WSN-LRMA. (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

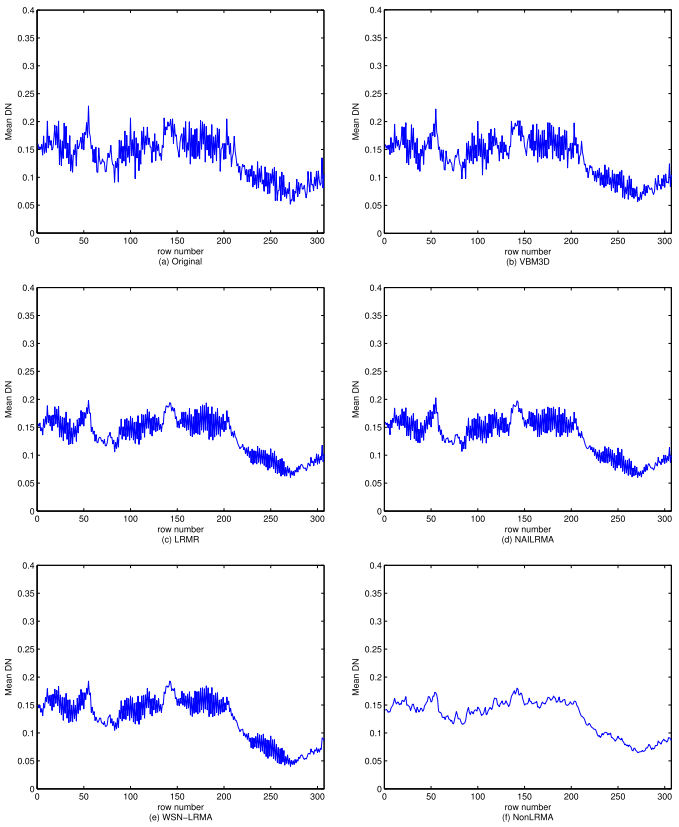


Fig. 14. Horizontal mean profiles of band 207 in HYDICE urban data set. (a) Original. (b) VBM3D. (c) LRMR. (d) NAILRMA. (e) WSN-LRMA. (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

1) Regularization Parameter λ: The regularization parameter λ in (7) balances the influences of the rank of L and the sparsity of S. As shown in Fig. 20, which presents the MPSNR, MSSIM, and MFSIM values with changing λ, it can be observed that when λ is selected from the interval [35, 50], NonLRMA achieves the optimal performance.

2) Penalty Parameter ρ_0: Subproblem (13) can be treated as an extension of the Moreau–Yosida proximity operator. ρ_k is a parameter controlling the tradeoff between the low rank constraint and the violation of fidelity, with smaller values of ρ_k associated with mapped points near the minimum and larger values leading to a smaller movement toward the minimum. Therefore, a varying sequence of ρ_k is used in (12) for acceleration, instead of keeping the penalty parameter ρ unchanged. Intuitively, the value of ρ_k should increase during iteration. ρ_0 is chosen from $[10^{-2}, 5 * 10^{-1}]$.

3) Parameter γ: As stated in Proposition 1, we have the proposition: $\lim_{\gamma \to 0} \|L\|_\gamma = rank(L)$. Theoretically, the smaller the γ value is, the tighter the γ-norm approximates the original rank function. To empirically testify how the γ value affects the denoising performance, we vary γ value and show the performance in Fig. 21. It is observed that when γ is set in the interval $[7 * 10^{-3}, 1.2 * 10^{-2}]$, NonLRMA performs the best.

4) Patch Size q: Given the optimal values of λ, ρ_0, and γ, we investigate the influence of the patch size q. The MPSNR, MSSIM, and MFSIM values of the results obtained by our proposed NonLRMA method with different patch sizes are reported in Table II. Based on Table II, we set q to 20 in our simulated data experiments. On real data sets, we set q to 50 in these experiments for faster convergence with a step size being 16.

Overall, the parameter settings of our proposed NonLRMA method used for simulated and real data experiments are presented in Table III.

5) Running Time: The running time of different approaches on two real-world HSIs is reported in Table IV. VBM3D[1] is the fastest approach among all the methods in all cases. However, its denoising performance is limited. Although the visual qualities of WSN-LRMA are similar to ours, Table IV demonstrates that NonLRMA is at least six times faster than WSN-LRMA. This is mainly due to the fact that fewer iterations are required by our algorithm to converge to the final solution. In addition, though our method consumes more time than NAILRMA, we provide better results.

D. Discussions

Most above-mentioned methods we compare with are recently developed methods in the low-rank modeling framework and represent state of the art due to their effectiveness and efficiency [28], [29]. Comparisons with such methods

[1]VBM3D is implemented with complied C++ mex-function and uses parallelization for acceleration, while other approaches are implemented using MATLAB.

Fig. 15. Restoration results on Hyperion Australia data set: slight noise band. (a) Original image located at the 61st band. Resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA.

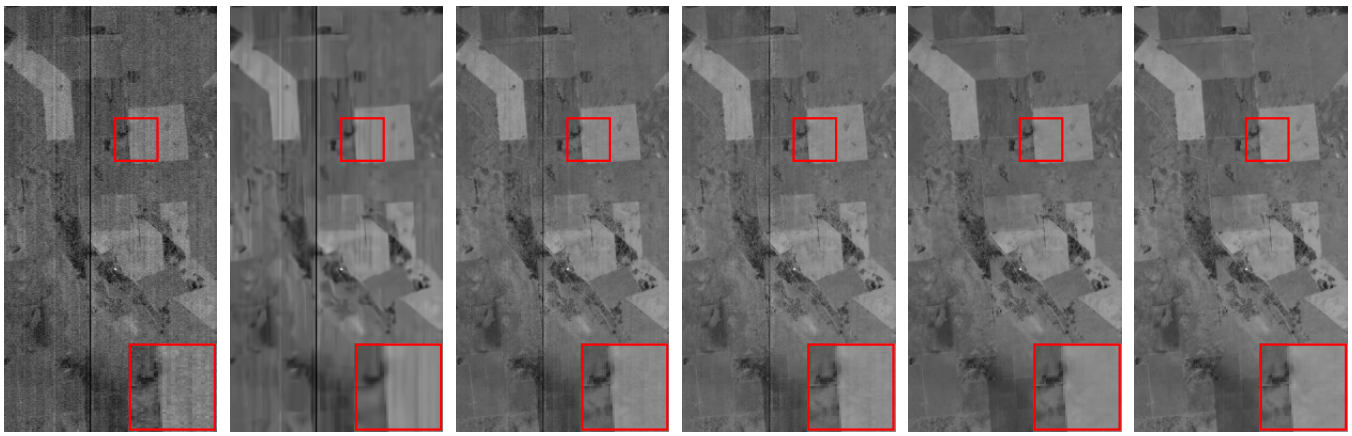


Fig. 16. Restoration results on Hyperion Australia data set: moderate noise band. (a) Original image located at the 123th band. The resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA.

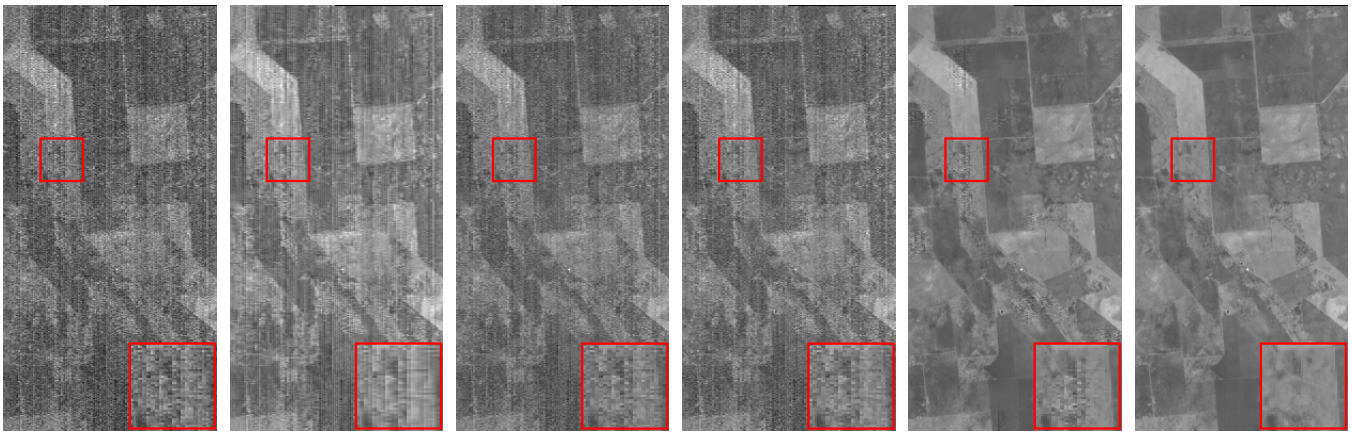


Fig. 17. Restoration results on Hyperion Australia data set: severe noise band. (a) Original image located at the 51st band. Resorted image obtained by (b) VBM3D, (c) LRMR, (d) NAILRMA, (e) WSN-LRMA, and (f) NonLRMA.

provide convincing evidence that supports the effectiveness of NonLRMA. As fully exploited in Section IV, NonLRMA can not only remove the mixture of Gaussian noise, impulse noise, deadlines, and stripes, but also recover most details of the dropped bands that are severely corrupted by mixed noises, with both qualitative and quantitative improvements as shown

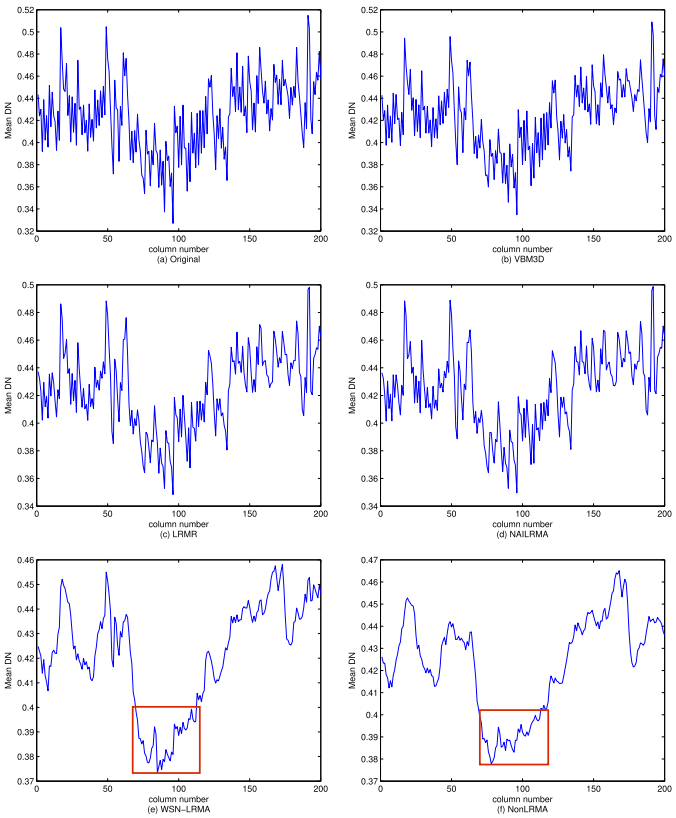


Fig. 18. Vertical mean profiles of band 51 in Hyperion Australia data set. (a) Original. (b) VBM3D. (c) LRMR. (d) NAILRMA. (e) WSN-LRMA. (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

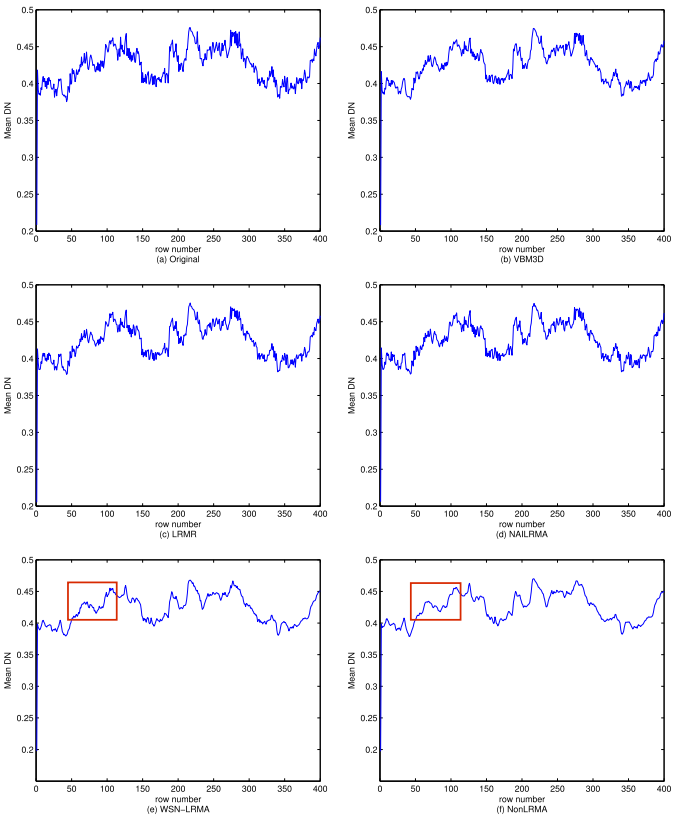


Fig. 19. Horizontal mean profiles of band 51 in Hyperion Australia data set. (a) Original. (b) VBM3D. (c) LRMR. (d) NAILRMA. (e) WSN-LRMA. (f) NonLRMA. The figure is better viewed in a zoomed-in PDF.

TABLE II
QUANTITATIVE EVALUATION OF NONLRMA WITH DIFFERENT PATCH SIZES

Patch Size: q	10	12	16	20	24	30
MPSNR(dB)	20.526	20.526	38.374	38.455	38.406	37.999
MSSIM	0.405	0.405	0.962	0.959	0.955	0.947
MFSIM	0.514	0.514	0.957	0.954	0.951	0.945

TABLE III
PARAMETERS SETTING IN OUR ALGORITHM

HSIs	γ	λ	ρ_0	q	step size
Synthetic data	10^{-2}	50	$3 * 10^{-2}$	20	8
Urban dataset	10^{-2}	10	$5 * 10^{-2}$	50	16
Australia dataset	10^{-2}	10	$5 * 10^{-2}$	50	16

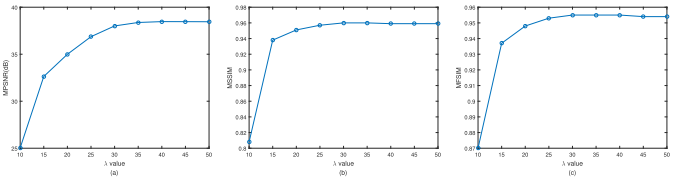


Fig. 20. Sensitivity analysis of parameter λ (λ from 10 to 50). (a) Change in the MPSNR value. (b) Change in the MSSIM value. (c) Change in the MFSIM value. The figure is better viewed in a zoomed-in PDF.

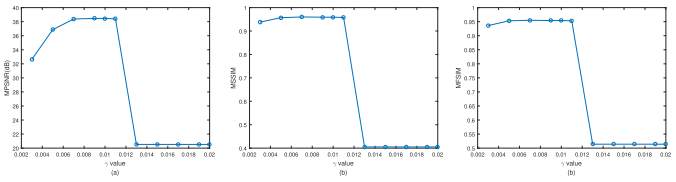


Fig. 21. Sensitivity analysis of parameter γ (γ from 3×10^{-3} to 2×10^{-2}). (a) Change in the MPSNR value. (b) Change in the MSSIM value. (c) Change in the MFSIM value. The figure is better viewed in a zoomed-in PDF.

in Table I and Fig. 4. Similar observations can be found on real data experiments conducted in Section IV-B. It has been empirically verified that NonLRMA can achieve promising performance with a wide range of parameter settings. Moreover, Table IV shows that NonLRMA is much faster than WSN-LRMA. Considering the superior performance as well as its insensitivity to parameters, NonLRMA has great potential in real world applications.

V. CONCLUSION

We have presented the NonLRMA model by applying the particular nonconvex surrogate of l_0 norm on the singular values to approximate the rank function instead of the nuclear norm for the reconstruction of corrupted HSIs. This model targets at recovering a low rank matrix representing the clean HSI, as well as simultaneously removing the Gaussian noise,

TABLE IV
COMPARISON OF RUNNING TIME (IN HOURS) FOR THE HSI DENOISING APPROACHES ON TWO DIFFERENT REAL-WORLD HSIs

Model	Urban dataset	Australia dataset
VBM3D	0.058	0.032
LRMR	0.59	0.31
NAILRMA	0.064	0.035
WSN-LRMA	2.45	1.39
NonLRMA	0.26	0.19

impulse noise, deadlines, and stripes. In addition, we develop an easy-to-implement iterative algorithm based on the ALM method. The extensive experiments on different simulated and real HSIs demonstrate that our approach is robust to both Gaussian and impulse noises, as well as deadlines and stripes.

Undeniable, the proposed algorithm still has room for improvement in running time. How to design fast and accurate HSI denoising approaches is our future work. Additionally, an HSI is essentially a third-order tensor, and we plan to consider sparsity on a tensor level in the future.

ACKNOWLEDGMENT

The authors would like to thank the Editor and anonymous reviewers for providing the insightful comments and suggestions, which greatly helped improve the quality of this paper.

REFERENCES

[1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[2] S. A. Lewis *et al.*, "Using hyperspectral imagery to estimate forest floor consumption from wildfire in boreal forests of Alaska, USA," *Int. J. Wildland Fire*, vol. 20, no. 2, pp. 255–271, Jan. 2011.

[3] K. C. Tiwari, M. K. Arora, and D. Singh, "An assessment of independent component analysis for detection of military targets from hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 5, pp. 730–740, Oct. 2011.

[4] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[5] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[6] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[7] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surv. Land Inf. Sci.*, vol. 62, no. 2, p. 115, Jun. 2002.

[8] C. Li, Y. Ma, X. Mei, J. Ma, and C. Liu, "Hyperspectral image classification with robust sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 13, no. 5, pp. 641–645, Mar. 2016.

[9] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[10] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[11] D. Cerra, R. Müller, and P. Reinartz, "Noise reduction in hyperspectral images through spectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 109–113, Jan. 2014.

[12] C. Jiang, H. Zhang, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image denoising with a combined spatial and spectral weighted hyperspectral total variation model," *Can. J. Remote Sens.*, vol. 42, no. 1, pp. 53–72, Apr. 2016.

[13] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1692–1700.

[14] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[15] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[16] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2272–2279.

[17] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

[18] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

[19] C. Li, Y. Ma, J. Huang, X. Mei, and J. Ma, "Hyperspectral image denoising using the robust low-rank tensor recovery," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 9, pp. 1604–1612, Sep. 2015.

[20] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[21] H. Tan, B. Cheng, J. Feng, G. Feng, and Y. Zhang, "Tensor recovery via multi-linear augmented Lagrange multiplier method," in *Proc. 6th Int. Conf. Image Graph.*, Aug. 2011, pp. 141–146.

[22] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.

[23] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.

[24] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2009, pp. 2080–2088.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

[26] H. Song, G. Wang, and K. Zhang, "Hyperspectral image denoising via low-rank matrix recovery," *Remote Sens. Lett.*, vol. 5, no. 10, pp. 872–881, Oct. 2014.

[27] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[28] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted schatten p-norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.

[29] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[30] M. Ye, Y. Qian, and J. Zhou, "Multitask sparse nonnegative matrix factorization for joint spectral–spatial hyperspectral imagery denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2621–2639, May 2015.

[31] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.

[32] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.

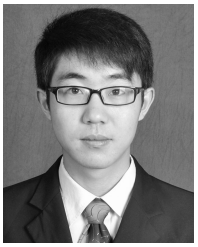
[33] Y. Chen, Y. Wang, M. Li, and G. He, "Augmented Lagrangian alternating direction method for low-rank minimization via non-convex approximation," *Signal, Image and Video Processing*, 2017, pp. 1–8.

[34] L. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.

[35] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Nov. 2007.

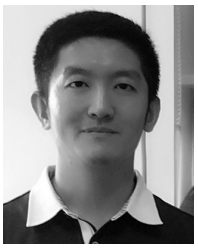
[36] Z. Kang, C. Peng, and Q. Cheng, "Robust PCA via nonconvex rank approximation," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2015, pp. 211–220.

[37] 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.
[38] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *J. Mach. Learn. Res.*, vol. 11, pp. 1081–1107, Mar. 2010.
[39] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, Jan. 2011, pp. 33–40.
[40] M. Rahmani and G. Atia, "A subspace learning approach for high dimensional matrix decomposition with efficient column/row sampling," in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 1206–1214.
[41] X. Lu, Y. Yuan, and P. Yan, "Sparse coding for image denoising using spike and slab prior," *Neurocomputing*, vol. 106, no. 6, pp. 12–20, Apr. 2013.
[42] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Hyperspectral image recovery employing a multidimensional nonlocal total variation model," *Signal Process.*, vol. 111, pp. 230–248, Jun. 2015.
[43] Y. Wang, P. Liu, Z. Li, T. Sun, C. Yang, and Q. Zheng, "Data regularization using Gaussian beams decomposition and sparse norms," *J. Inverse Ill-Posed Problems*, vol. 21, no. 1, pp. 1–23, Feb. 2013.
[44] C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 829–839, Feb. 2016.
[45] X. Liu, X. Lu, H. Shen, Q. Yuan, Y. Jiao, and L. Zhang, "Noise removal from hyperspectral image with joint spectral–spatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.
[46] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.
[47] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[48] S. Gaïffas and G. Lecué. (2011). "Weighted algorithms for compressed sensing and matrix completion." [Online]. Available: https://arxiv.org/pdf/1107.1638.pdf
[49] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 569–592, 2009.
[50] [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html
[51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[52] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[53] *HYDICE Image of Washington DC Mall*, accessed on Apr. 10, 2016. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html
[54] *HYDICE Urban Images*, accessed on May 15, 2016. [Online]. Available: http://www.tec.army.mil/hypercube
[55] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising with a spatial–spectral view fusion strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2314–2325, May 2014.
[56] *EO-1 Hyperion Australia Images*, accessed on May 2, 2016. [Online]. Available: http://remote-sensing.nci.org.au/



Yongyong Chen received the B.S. degree in information and computing sciences from the Shandong University of Science and Technology, Qingdao, China, in 2014, where he is currently pursuing the M.S. degree with the College of Mathematics and Systems Science.

He is currently an Exchange Student with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China. His research interests include (nonconvex) low-rank and sparse matrix/tensor decomposition models, with applications to image processing, magnetic resonance imaging, and computer vision.

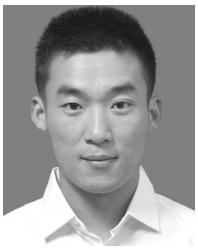


Yanwen Guo received the Ph.D. degree in applied mathematics from the State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, China, in 2006.

He was a Visiting Professor with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2006 and 2009, and the Department of Computer Science, The University of Hong Kong, Hong Kong, in 2008, 2012, and 2013. He was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Champaign, IL, USA, from 2013 to 2015. He is currently a Full Professor with the National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing, China. His research interests include image and video processing, vision, and computer graphics.

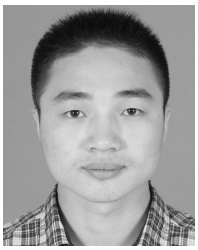
Yongli Wang received the B.S. degree in operations research and control theory from the Shandong University of Science and Technology, Qingdao, China, in 2001, and the Ph.D. degree in applied mathematics from Shanghai Jiao Tong University, Shanghai, China, in 2006.

Her research interests include optimization theory and method, low-rank representation, image processing, and distributed computation.



Dong Wang received the M.S. degree from the College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Traffic and Transportation, Institute of Systems Science, Beijing Jiaotong University, Beijing, China.

His research interests include multiagent systems, traffic planning, mathematical programming, and image processing.



Chong Peng received the bachelor's degree in statistics from Qingdao University, Qingdao, China, in 2012. He is currently pursuing the Ph.D. degree in computer science with Southern Illinois University, Carbondale, IL, USA.

He has published more than 20 research papers in prestigious journals and prominent conferences, including the IEEE Conference on Computer Vision and Pattern Recognition, IEEE International Conference on Data Engineering, Industrial Conference on Data Mining, and so on. His research interests include machine learning, data mining, information retrieval, and image processing.

Guoping He received the B.S. degree in computational mathematics from the Shandong University of Science and Technology, Qingdao, China, in 1982, and the M.S. and the Ph.D. degrees in operations research and control theory from the Chinese Academy of Sciences, Beijing, China, in 1988 and 1995, respectively.

He is currently the Vice President with the Shandong Academy of Sciences, Jinan, China. He has over 100 research papers and is the holder of eight foundations. His research interests include nonlinear optimization theory, numerical computation, and data mining.