# Hyperspectral Image Restoration Via Total Variation Regularized Low-Rank Tensor Decomposition

Yao Wang , Jiangjun Peng, Qian Zhao , Yee Leung, Xi-Le Zhao , and Deyu Meng , *Member, IEEE*

*Abstract*—Hyperspectral images (HSIs) are often corrupted by a mixture of several types of noise during the acquisition process, e.g., Gaussian noise, impulse noise, dead lines, stripes, etc. Such complex noise could degrade the quality of the acquired HSIs, limiting the precision of the subsequent processing. In this paper, we present a novel tensor-based HSI restoration approach by fully identifying the intrinsic structures of the clean HSI part and the mixed noise part. Specifically, for the clean HSI part, we use tensor Tucker decomposition to describe the global correlation among all bands, and an anisotropic spatial–spectral total variation regularization to characterize the piecewise smooth structure in both spatial and spectral domains. For the mixed noise part, we adopt the $\ell_1$ norm regularization to detect the sparse noise, including stripes, impulse noise, and dead pixels. Despite that TV regularization has the ability of removing Gaussian noise, the Frobenius norm term is further used to model heavy Gaussian noise for some real-world scenarios. Then, we develop an efficient algorithm for solving the resulting optimization problem by using the augmented Lagrange multiplier method. Finally, extensive experiments on simulated and real-world noisy HSIs are carried out to demonstrate the superiority of the proposed method over the existing state-of-the-art ones.

*Index Terms*—Hyperspectral image (HSI), low-rank tensor decomposition, mixed noise, total variation (TV).

## I. INTRODUCTION

HYPERSPECTRAL imaging employs an imaging spectrometer to collect hundreds of spectral bands ranging from ultraviolet to infrared wavelengths for the same area on the surface of the Earth. It has a wide range of applications including environmental monitoring, military surveillance, mineral exploration, to name but a few examples [1], [2]. Due to various factors, e.g., thermal electronics, dark current, and stochastic error of photocounting in an imaging process, hyperspectral images (HSIs) are inevitably corrupted by severe noise during the acquisition process. Therefore, HSI denoising is an important research topic, which has received much attention in the past decades.

A natural way for restoring HSIs is to regard each band as a gray-level image and then apply traditional two-dimensional (2-D) or one-dimensional denoising methods to remove noise band-by-band. See, e.g., [3]–[6]. However, this kind of method ignores the correlations among the spectral bands or spatial pixels, and thus usually could not provide satisfactory results. To take account of such correlations, a variety of studies have been conducted over the years. For example, Othman and Qian [7] proposed a hybrid spatial–spectral wavelet shrinkage method to take advantage of the dissimilarity of the signal regularity in both the spatial and spectral domains of the HSIs. Zhong and Wang [8] designed a multiple spectral-band conditional random field method, which can simultaneously model and utilize the spatial and spectral dependencies in a unified probabilistic framework. In [9], an efficient HSI denoising procedure is proposed by designing a spectral–spatial adaptive total variation (TV) model, where the spectral noise differences and spatial information differences are both considered. Additionally, several advanced techniques in traditional image processing, including the nonlocal similarity [10], wavelet shrinkage [11], anisotropic diffusion [12], have been adopted for HSI restoration in recent years.

Actually, all the aforementioned methods are designed to remove only one or two types of noise, i.e., Gaussian noise, impulse noise, or hybrid Gaussian–impulse noise. In real-world scenarios, however, HSIs are usually corrupted by a combination of several different types of noise during the acquisition process, e.g., Gaussian noise, impulse noise, dead lines, stripes, etc. Although several methods based on low-rank matrix modeling have been proposed for removing mixed noise in HSIs, e.g., [13]–[17], modeling an HSI cube as a matrix is not able to utilize the finer spatial–spectral information, leading to suboptimal restoration results in some heavy noisy situations (see Fig. 1 for an example). It can easily be observed that, for the HSI band polluted by severe noise (e.g., band 108), the state-of-the-art low-rank matrix method, e.g., TV-regularized low-rank matrix factorization (LRTV) proposed by [17], could not provide satisfactory restoration as for the moderately noisy HSI band (e.g., band 103).

Y. Wang, J. Peng, and Q. Zhao are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yao.s.wang@gmail.com; andrew.pengjj@gmail.com; timmy.zhaoqian@gmail.com).

Y. Leung is with the Institute of Future Cities, The Chinese University of Hong Kong, Shatin, Hong Kong, SAR (e-mail: yeeleung@cuhk.edu.hk).

X.-L. Zhao is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: xlzhao122003@163.com).

D. Meng is with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China, and also with Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China (e-mail: dymeng@xjtu.edu.cn).

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

Band 103    Band 108



(a)    (b)

(c)    (d)
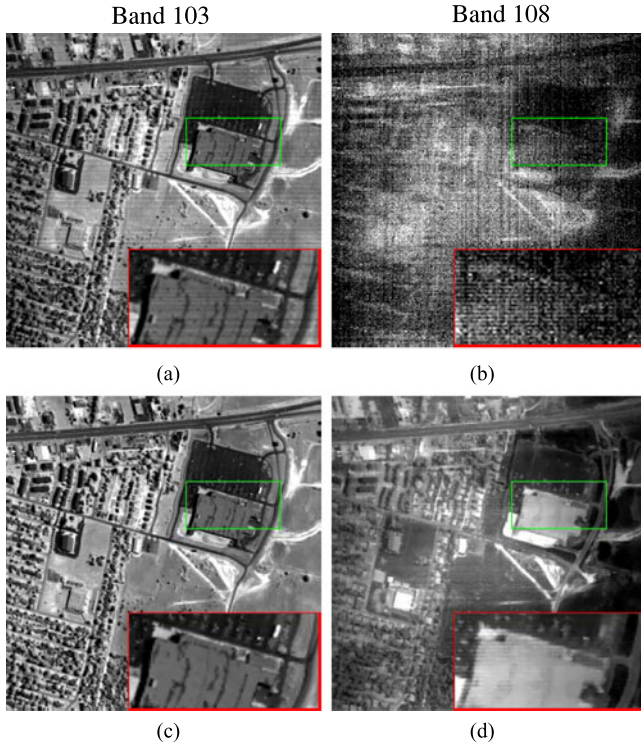
Fig. 1. Typical restoration example on the HYDICE urban data: (a) and (b) the original band; and (c) and (d) denoised results by LRTV.



Fig. 2. Schematic diagrams of the popular matrix procedures and our tensor procedure for HSI restoration.

Recently, a number of studies, both practically and theoretically, have demonstrated the advantages of direct tensor modeling techniques over matricization techniques in dealing with high-order tensor data. See, e.g., [18]–[22]. Motivated by such studies, we propose in this paper a novel mixed noise removal approach using direct tensor modeling techniques to fully exploit the spatial–spectral priors underlying the clean HSI part and characterize the intrinsic structures of the heavy noise part. Actually, the idea of exploiting the spatial–spectral information has been widely considered in other HSI processing problems, e.g., the works [23]–[25] for HSI classification task and the works [26] and[27] for band selection task. However, as far as we know, few studies have been conducted to exploit the spatial–spectral information using a direct tensor modeling idea for HSI restoration. To highlight our contributions, we shall go over related work on HSI restoration using different low-rank modeling techniques and contrast our innovations with the existing literature.

### A. Related Work

In the past few years, various approaches based on low-rank matrix approximation (LRMA) have been proposed for HSI restoration and can be represented as state-of-the-art techniques. Motivated by the idea of a robust principle component analysis (RPCA) [28], Zhang et al. [13] explored the low-rank property by lexicographically ordering a patch of the HSI into a 2-D matrix and modeled the non-Gaussian noise (including impulse noise, dead lines, and stripes) as a sparse part. The so-called
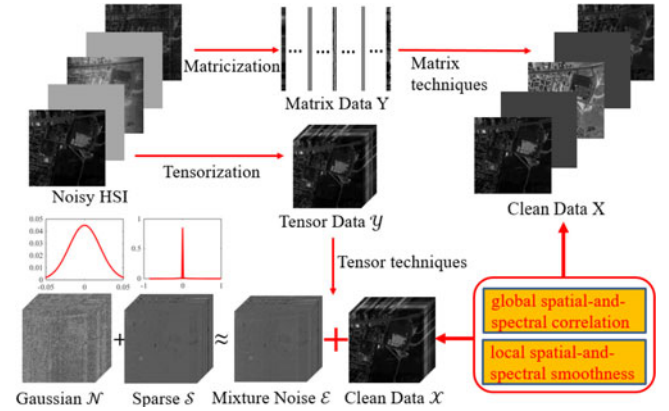
"go decomposition" (GoDec) [29] algorithm was used to enforce rank $r$ and cardinality $k$ constraints for the low-rank and sparse parts, respectively. Considering that the noise intensity in different bands is different, a noise-adjusted iterative LRMA method on the basis of patchwise randomized singular value decomposition was proposed in [15]. Encouraged by the powerfulness of TV regularization in various image restoration tasks, He et al. [17] integrated the band-by-band TV regularization into a rank-constrained RPCA model to explore the spatial piecewise smoothness of the HSI and, as a result, enhanced the capability of the LRMA technique for HSI restoration. Similarly, Wu et al. [30] proposed an HSI mixed denoising model by combining the band-by-band TV regularization with the widely used weighted nuclear norm minimization (WNNM) [31], [32]. To give better approximation to the low-rank assumption of HSI data, a weighted Schatten $p$-norm regularization was introduced by Xie et al. [33] into the LRMA framework. To simultaneously exploit both the local similarity within an HSI patch and the nonlocal similarity across patches in a group, a novel group low-rank representation model was considered in [16]. As stated above, though this kind of LRMA methodology has been an increasingly useful technique in HSI restoration, it fails to fully exploit the prior knowledge on the intrinsic structures of the HSI cube after vectorizing the HSI bands.

Despite the efficiency of tensor methods in removing Gaussian noise or hybrid Gaussian–impulse noise in HSIs, e.g., [34]–[39], to the best of our knowledge, only two studies [40], [41] based on tensor techniques have been conducted to remove mixed noise in HSIs. Specifically, in [40], the tensor nuclear norm introduced by Liu et al. [18] was used to acquire low-rank property of the HSI cube, and the mixed $\ell_2/\ell_1$ norm was used to impose sparsity property of the outliers and non-Gaussian noise; the work [41] integrated the structure tensor TV [42] into the WNNM model and showed that it outperforms the band-by-band TV-regularized WNNM method [30], because of the utilization of finer spatial structure information.

Basically, our work is related to the aforementioned works; however, there are significant differences between our work and the others. Fig. 2 illustrates the framework of the popular matrix

modeling idea, especially LRMA, and our direct tensor modeling idea in dealing with the noisy HSI cubes. It can be seen that the matricization techniques should preliminarily vectorize all the HSI bands at the cost of losing spatial structures of the HSI cube, whereas direct tensor modeling techniques could more faithfully deliver the underlying information of the HSI cube without destroying the spatial structures. Despite their connection and similarity, compared with the work in [40], we adopt an anisotropic spatial–spectral TV to further exploit the local piecewise smoothness in both spatial and spectral domains; compared with the work in [41], low-rank Tucker decomposition is used to further characterize the spatial correlation in each HSI band, and the proposed spatial–spectral TV is simpler than the structure tensor TV. Additionally, the use of the Frobenius norm term for dealing with very large Gaussian noise was not considered in [40], [41]. A detailed motivation of our work can be found in Section III.

### B. Our Contributions

In this paper, we mainly focus on HSI mixed noise removal. The main contributions of this paper are summarized as follows.

1) The low-rank Tucker decomposition is applied to separate the clean HSI from the raw observation corrupted by complex noise, which can be capable of fully exploiting the global spatial-and-spectral correlation among the HSI bands. The classical higher order orthogonality iteration (HOOI) algorithm is used to achieve such Tucker decomposition efficiently, without bringing extra computational cost.
2) A designed spatial–spectral total variation (SSTV) regularizer is incorporated into the low-rank Tucker decomposition framework. It is shown that the SSTV regularization has the ability of enhancing the spatial information, and thus preserves the spectral signatures of the HSIs very well.
3) The well-known augmented Lagrange multiplier (ALM) method is used for solving the TV-regularized low-rank tensor decomposition model. And extensive studies on both simulated and real datasets are carried out to illustrate that the proposed method clearly improves the HSI restoration results over some other popular techniques, in terms of both the quantitative evaluation and the visual comparison.

### C. Organization of the Paper

This paper is organized as follows. To facilitate our presentation, we first introduce some notations and preliminaries of tensors in Section II. In Section III, the anisotropic SSTV-regularized low-rank tensor decomposition model and its motivations are introduced. We then develop an efficient ALM algorithm for solving the proposed model. In Section IV, extensive experiments on both simulated and real datasets are carried out to illustrate the merits of our model. We then conclude this paper with some discussions on future research in Section V.

TABLE I
NOTATIONS

| Notations | Explanations |
|---|---|
| $\mathcal{X}, \mathbf{X}, \mathbf{x}, x$ | Tensor, matrix, vector, scalar. |
| $\mathbf{X}_{(n)}$ or $\mathcal{X}_{(n)}$ | Mode-$n$ matricization of tensor $\mathcal{X} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$, obtained by arranging the mode-$n$ fibers as the columns of the resulting matrix of size $\Re^{I_n \times \prod_{k \neq n} I_k}$. |
| $(r_1, r_2, \ldots, r_N)$ | Multilinear rank, where $r_n = \text{rank}(\mathbf{X}_{(n)})$, $n = 1, 2, \ldots, N$. |
| $\langle \mathcal{X}, \mathcal{Y} \rangle$ | Inner product of tensor $\mathcal{X}$ and $\mathcal{Y}$. |
| $\|\mathcal{X}\|_F$ | Frobenius norm of tensor $\mathcal{X}$. |
| $\mathcal{Y} = \mathcal{X} \times_n \mathbf{U}$ | Mode-$n$ multiplication of $\mathcal{X}$ and $\mathbf{U}$ with the matrix representation $\mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)}$. |

## II. NOTATIONS AND PRELIMINARIES

It is well known that a tensor can be treated as a multi-index numerical array, and its order is defined as the number of its modes or dimensions. A real-valued tensor of order $N$ is denoted by $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and its entries by $x_{i_1, i_2, \ldots, i_N}$. We then can consider an $N \times 1$ vector $x$ as a tensor of order one, and an $N \times M$ matrix $\mathbf{X}$ as a tensor of order two. Following [43], we shall provide a brief introduction on tensor algebra.

The inner product of two same-sized tensors $\mathcal{X}$ and $\mathcal{Y}$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{i_1, i_2, \ldots, i_N} x_{i_1, i_2, \ldots, i_N} \cdot y_{i_1, i_2, \ldots, i_N}$. Then, the corresponding Frobenius norm is defined as $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. The so-called mode-$n$ matricization of a tensor $\mathcal{X}$ is denoted as $\mathbf{X}_{(n)}$, where the tensor element $(i_1, i_2, \ldots, i_N)$ is mapped to the matrix element $(i_n, j)$ satisfying $j = 1 + \sum_{k=1, k \neq n}^{N}(i_k - 1)J_k$ with $J_k = \prod_{m=1, m \neq n}^{k-1} I_m$. The mode-$n$ multiplication of a tensor $\mathcal{X}$ with a matrix $\mathbf{U}$ is denoted by $\mathcal{X} \times_n \mathbf{U}$ and elementwise, we have $(\mathcal{X} \times_n \mathbf{U})_{i_1, \ldots, i_{n-1} j i_{n+1} \ldots, i_N} = \sum_{i_n} x_{i_1, i_2, \ldots, i_N} \cdot u_{j, i_n}$. The multilinear rank is defined as an array $(r_1, r_2, \ldots, r_N)$ where $r_n = \text{rank}(\mathbf{X}_{(n)})$, $n = 1, 2, \ldots, N$.

The tensor notations used in this work are summarized in Table I. Interested readers are referred to [43] for a more detailed introduction.

## III. HSI RESTORATION VIA TV-REGULARIZED LOW-RANK TENSOR DECOMPOSITION

### A. Motivation

In many real situations, the observed HSI data are contaminated by a mixture of different kinds of noise [13], [15], [17]. As a result, a noisy HSI cube denoted by a three-order tensor $\mathcal{Y} := \{\mathbf{Y}^1, \mathbf{Y}^2, \ldots, \mathbf{Y}^B\}$, where each matrix $\mathbf{Y}^i \in \mathbb{R}^{h \times w}$ ($i = 1, 2, \ldots, B$) represents the $i$th band, with the height $h$ and width $w$, and $B$ denotes the number of bands, can be described as

$$\mathcal{Y} = \mathcal{X} + \mathcal{E} \qquad (1)$$

where $\mathcal{X}$ and $\mathcal{E}$ are with the same size of $\mathcal{Y}$, which represent the clean HSI cube and the mixed noise term, respectively. Now, the objective of HSI restoration is to estimate $\mathcal{X}$ from the observed $\mathcal{Y}$ by exploiting the structures of the clean HSI $\mathcal{X}$ and the noise terms $\mathcal{E}$.
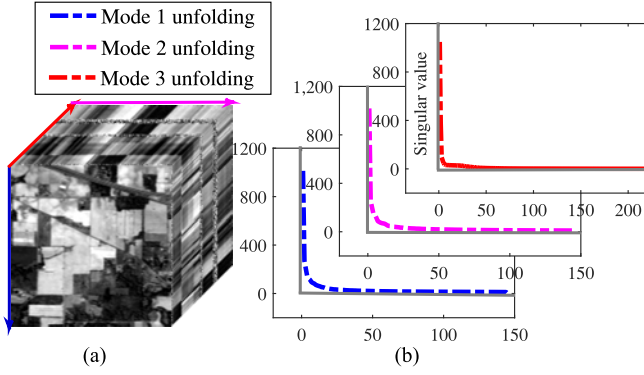
Fig. 3. Global spatial-and-spectral correlation in the HSI cube. (a) An HSI, (b) Singular Values of unfolding matrices.
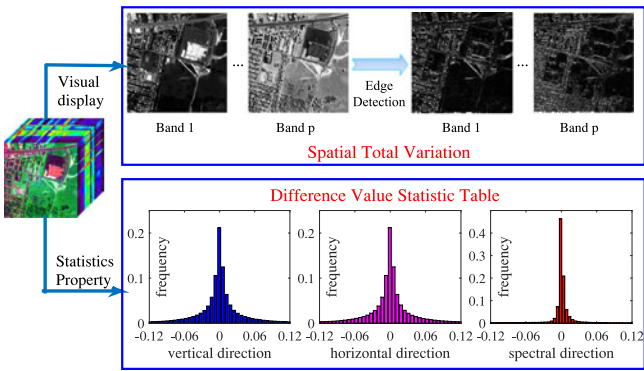


Fig. 4. Property of total variation of HSI in both spatial and spectral modes.

Following the studies of [17] and [33], we divide the noise term $\mathcal{E}$ into two subterms, as shown in Fig. 2, i.e., the Gaussian noise term $\mathcal{N}$ and the sparse noise term $\mathcal{S}$ including stripes, impulse noise, and dead pixels, leading to the following degradation model:

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} + \mathcal{S}. \tag{2}$$

As such, the Frobenius norm and the $\ell_1$ norm can be naturally used to model the noise terms $\mathcal{N}$ and $\mathcal{S}$, respectively.

It is well known that each spectral signature can be represented by a linear combination of a small number of pure spectral endmembers, which means that the mode-3 matricization $\mathbf{X}_{(3)}$ can be factorized as $\mathbf{X}_{(3)} = \mathbf{MD}^T$, where $\mathbf{D} \in \mathbb{R}^{B \times r}$ is the so-called endmember matrix, and $\mathbf{M} \in \mathbb{R}^{wh \times r}$ is regarded as the abundance matrix. As stated in [13], the number of endmembers $r$ is relatively small, that is, $r \ll B$ or $r \ll wh$, which means that only a small fraction of the singular values is greater than zero, as shown in the singular value curve of $\mathbf{X}_{(3)}$ in Fig. 3. Besides, like other real images, there have been certain correlations in both spatial modes, as also shown in Fig. 3, where the singular value curves of $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$ have obvious decaying trends. Therefore, we can utilize low-rank Tucker decomposition to characterize the aforementioned spatial–spectral correlation.

TV regularization has been widely used to explore the spatial piecewise smooth structure for tackling the HSI restoration task. See, e.g., [17], [30], [44]. One can see in Fig. 4 an illustrative example of edge detection for demonstrating the spatial smooth-

ness of one HSI. Actually, as depicted in the third row of Fig. 4, there also exits a strong local smooth structure of an HSI along its spectral mode. That is, most of the difference values between adjacency bands in the spectral domain are nearly equal to zero, and the number of zeros is clearly larger than that for the spatial domain. Considering the fact that the commonly used band-by-band TV regularizer neglects such spectral smoothness, it could be useful to design a new SSTV regularizer to explore the piecewise smooth structure in both spatial and spectral domains.

### B. Low-Rank Tensor Decomposition With Anisotropic SSTV

As previously stated, exploiting the prior knowledge is a key consideration for HSI mixed noise removal. Based on the discussion of in Section III-A, by combining the low-rank and TV properties in both spatial and spectral modes, we introduce a *TV-regularized low-rank tensor decomposition* (LRTDTV for short) model as follows:

$$\min_{\mathcal{X},\mathcal{S},\mathcal{N}} \tau \|\mathcal{X}\|_{\mathrm{SSTV}} + \lambda \|\mathcal{S}\|_1 + \beta \|\mathcal{N}\|_F^2$$

$$\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N},$$

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \mathbf{U}_i^T \mathbf{U}_i = \mathbf{I} \, (i = 1, 2, 3) \tag{3}$$

where $\mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ is the Tucker decomposition with core tensor $\mathcal{C}$ and factor matrices $\mathbf{U}_i s$ of rank $r_i s$, and the SSTV term $\|\mathcal{X}\|_{\mathrm{SSTV}}$ is defined as

$$\|\mathcal{X}\|_{\mathrm{SSTV}} := \sum_{i,j,k} w_1 |x_{i,j,k} - x_{i,j,k-1}| + w_2 |x_{i,j,k} - x_{i,j-1,k}|$$

$$+ w_3 |x_{i,j,k} - x_{i-1,j,k}| \tag{4}$$

where $x_{i,j,k}$ is the $(i, j, k)$th entry of $\mathcal{X}$, and $w_j \, (j = 1, 2, 3)$ is the weight along the $j$th mode of $\mathcal{X}$ that controls its regularization strength and should be carefully tuned as we shall discuss in Section IV-C. It is not hard to see that (3) is a nonconvex optimization problem due to the nonconvexity of the Tucker decomposition. Thus, it would be possible to find a good local solution by using the popular ALM method [45] that we shall show in the next section.

It is worth noting that the proposed model can fully capture the spatial and spectral information of the HSI, and thus is expected to have a strong ability in mixed noise removal. Specifically, the Tucker decomposition could make use of the spectral similarity of all the pixels and certain correlations in both spatial modes. Once the sparse noise term, including impulse noise, dead lines, and stripes, is detected by the $\ell_1$ norm term, the designed SSTV norm term is used to characterize the piecewise smooth structure in both spatial and spectral domains and, as a result, helps to remove the Gaussian noise. The Frobenius norm term is further used to model Gaussian noise and can enhance the performance of the proposed model in some heavy Gaussian noise situations.

## C. Optimization Procedure

By introducing some auxiliary variables, we rewrite (3) into the following equivalent minimization problem:

$$\min_{\mathcal{C}, \mathbf{U}_i, \mathcal{X}, \mathcal{F}, \mathcal{S}, \mathcal{N}} \tau \|\mathcal{F}\|_1 + \lambda \|\mathcal{S}\|_1 + \beta \|\mathcal{N}\|_F^2$$

$$\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}, \mathcal{X} = \mathcal{Z}, D_w(\mathcal{Z}) = \mathcal{F},$$

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \mathbf{U}_i^T \mathbf{U}_i = \mathbf{I} \quad (5)$$

where $D_w(\cdot) = [w_1 \times D_h(\cdot); w_2 \times D_v(\cdot); w_3 \times D_t(\cdot)]$ is the so-called weighted three-dimensional difference operator and $D_h$, $D_v$, $D_t$ are the first-order difference operators with respect to three different directions of the HSI cube. Based on the ALM methodology, the above-mentioned problem (5) can be transformed into minimizing the following augmented Lagrangian function:

$$L(\mathcal{X}, \mathcal{S}, \mathcal{N}, \mathcal{Z}, \mathcal{F}, \Gamma_1, \Gamma_2, \Gamma_3) = \tau \|\mathcal{F}\|_1 + \lambda \|\mathcal{S}\|_1 + \beta \|\mathcal{N}\|_F^2$$

$$\langle \Gamma_1, \mathcal{Y} - \mathcal{X} - \mathcal{S} - \mathcal{N} \rangle + \langle \Gamma_2, \mathcal{X} - \mathcal{Z} \rangle$$

$$+ \langle \Gamma_3, D_w(\mathcal{Z}) - \mathcal{F} \rangle + \frac{\mu}{2} \left( \|\mathcal{Y} - \mathcal{X} - \mathcal{S} - \mathcal{N}\|_F^2 \right.$$

$$+ \|\mathcal{X} - \mathcal{Z}\|_F^2 + \|D_w(\mathcal{Z}) - \mathcal{F}\|_F^2 \right) \quad (6)$$

under the constraints $\mathcal{X} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ and $\mathbf{U}_i^T \mathbf{U}_i = \mathbf{I}$, where $\mu$ is the penalty parameter and $\Gamma_i$ ($i = 1, 2, 3$) are the Lagrange multipliers. Therefore, we can alternatively optimize the augmented Lagrangian function (6) over one variable while fixing the others. Specifically, in the $(k + 1)$th iteration, variables involved in the model (3) can be updated as follows:

1) *Update $\mathcal{C}, \mathbf{U}_i, \mathcal{X}$:* Extracting all terms containing $\mathcal{X}$ from the augmented Lagrangian function (6), we need to solve

$$\min_{\substack{\mathbf{U}_i^T \mathbf{U}_i = \mathbf{I}, \\ \mathcal{X} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3}} \left\langle \Gamma_1^{(k)}, \mathcal{Y} - \mathcal{X} - \mathcal{S}^{(k)} - \mathcal{N}^{(k)} \right\rangle + \left\langle \Gamma_2^{(k)}, \right.$$

$$\left. \mathcal{X} - \mathcal{Z}^{(k)} \right\rangle + \frac{\mu}{2} \left( \|\mathcal{Y} - \mathcal{X} - \mathcal{S}^{(k)} - \mathcal{N}^{(k)}\|_F^2 + \|\mathcal{X} - \mathcal{Z}^{(k)}\|_F^2 \right)$$

which can be easily transformed into the following equivalent problem:

$$\min_{\mathbf{U}_i^T \mathbf{U}_i = \mathbf{I}} \mu \left\| \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 - \frac{1}{2} \left( \mathcal{Y} - \mathcal{S}^{(k)} - \mathcal{N}^{(k)} \right. \right.$$

$$\left. \left. + \mathcal{Z}^{(k)} + \left( \Gamma_1^{(k)} - \Gamma_2^{(k)} \right) / \mu \right) \right\|_F^2.$$

By resorting to the classic HOOI algorithm [43], we can easily get $\mathcal{C}^{(k+1)}$ and $\mathbf{U}_i^{(k+1)}$ ($i = 1, 2, 3$). Then, $\mathcal{X}$ can be updated as follows:

$$\mathcal{X}^{(k+1)} = \mathcal{C}^{(k+1)} \times_1 \mathbf{U}_1^{(k+1)} \times_2 \mathbf{U}_2^{(k+1)} \times_3 \mathbf{U}_3^{(k+1)}. \quad (7)$$

2) *Update $\mathcal{Z}$:* Extracting all terms containing $\mathcal{Z}$ from the augmented Lagrangian function (6), we can deduce

$$\mathcal{Z}^{(k+1)} = \underset{\mathcal{Z}}{\operatorname{argmin}} \left\langle \Gamma_2^{(k)}, \mathcal{X}^{(k+1)} - \mathcal{Z} \right\rangle + \left\langle \Gamma_3^{(k)}, D_w(\mathcal{Z}) - \right.$$

$$\left. \mathcal{F}^{(k)} \right\rangle + \frac{\mu}{2} \left( \|\mathcal{X}^{(k+1)} - \mathcal{Z}\|_F^2 + \|D_w(\mathcal{Z}) - \mathcal{F}^{(k)}\|_F^2 \right).$$

Thus, optimizing this problem can be treated as solving the following linear system:

$$(\mu \mathbf{I} + \mu D_w^* D_w) \mathcal{Z} = \mu \mathcal{X}^{(k+1)} + \mu D_w^* \left( \mathcal{F}^{(k)} \right) + \Gamma_2^{(k)} - D_w^* \left( \Gamma_3^{(k)} \right)$$

where $D_w^*$ indicates the adjoint operator of $D_w$. Thanks to the block-circulant structure of the matrix corresponding to the operator $D_w^* D_w$, it can be diagonalized by the three-dimensional (3-D) FFT matrix. Therefore, we have

$$\begin{cases} H_z = \mu \mathcal{X}^{(k+1)} + \mu D_w^* \left( \mathcal{F}^{(k)} \right) + \Gamma_2^{(k)} - D_w^* \left( \Gamma_3^{(k)} \right), \\ T_z = w_1^2 |\text{fftn}(D_h)|^2 + w_2^2 |\text{fftn}(D_v)|^2 + w_3^2 |\text{fftn}(D_t)|^2, \\ \mathcal{Z}^{(k+1)} = \text{ifftn} \left( \frac{\text{fftn}(H_z)}{\mu \mathbf{1} + \mu T_z} \right) \end{cases} \quad (8)$$

where fftn and ifftn indicate the fast 3-D Fourier transform and its inverse transform, respectively, $|\cdot|^2$ is the element-wise square, and the division is also performed in an element-wise manner.

3) *Update $\mathcal{F}$:* Extracting all terms containing $\mathcal{F}$ from (6), we can get

$$\mathcal{F}^{(k+1)} = \underset{\mathcal{F}}{\operatorname{argmin}} \tau \|F\|_1 + \left\langle \Gamma_3^{(k)}, D_w \left( \mathcal{Z}^{(k+1)} \right) - \mathcal{F} \right\rangle$$

$$+ \frac{\mu}{2} \|D_w \left( \mathcal{Z}^{(k+1)} \right) - \mathcal{F}\|_F^2$$

$$= \underset{\mathcal{F}}{\operatorname{argmin}} \tau \|\mathcal{F}\|_1 + \frac{\mu}{2} \|\mathcal{F} - \left( D_w \left( \mathcal{Z}^{(k+1)} \right) + \frac{M_3^{(k)}}{\mu} \right) \|_F^2.$$

By introducing the so-called *soft-thresholding operator*

$$\mathcal{R}_\Delta(\mathbf{x}) = \begin{cases} x - \Delta, & \text{if } \mathbf{x} > \Delta \\ x + \Delta, & \text{if } \mathbf{x} < \Delta \\ 0, & \text{otherwise} \end{cases}$$

where $x \in \mathbb{R}$ and $\Delta > 0$, then we can update $\mathcal{F}^{k+1}$ as

$$\mathcal{F}^{k+1} = \mathcal{R}_{\frac{\tau}{\mu}} \left( D_w \left( \mathcal{Z}^{(k+1)} \right) + \frac{\Gamma_3^{(k)}}{\mu} \right). \quad (9)$$

4) *Update $\mathcal{S}$:* Similarly, we should consider

$$\mathcal{S}^{(k+1)} = \underset{\mathcal{S}}{\operatorname{argmin}} \lambda \|\mathcal{S}\|_1 + \left\langle \Gamma_1^{(k)}, \mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S} - \mathcal{N}^{(k)} \right\rangle$$

$$+ \frac{\mu}{2} \|\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S} - \mathcal{N}^{(k)}\|_F^2$$

$$= \underset{\mathcal{S}}{\operatorname{argmin}} \lambda \|\mathcal{S}\|_1 + \frac{\mu}{2} \|\mathcal{S} - \left( \mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{N}^{(k)} + \frac{\Gamma_1^{(k)}}{\mu} \right) \|_F^2.$$

By using the soft-thresholding operator introduced above, the solution of the above-mentioned problem can be obtained as

$$\mathcal{S}^{k+1} = \mathcal{R}_{\frac{\lambda}{\mu}} \left( \mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{N}^{(k)} + \frac{M_1^{(k)}}{\mu} \right). \quad (10)$$

---

**Algorithm 1:** LRTDTV solver.

---

**Input:** The noisy HSI $\mathcal{Y}$, desired rank $[r_1, r_2, r_3]$, stopping
  criteria $\epsilon$, and the regularization parameters $\tau$, $\lambda$ and
  $\beta$, and the weights $w_i$s.
**Output:** The restored HSI $\mathcal{X}$
  1: Initialize $\mathcal{X} = \mathcal{Z} = \mathcal{S} = \mathcal{N} = 0$, $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$,
     $\mu_{\max} = 10^6$, $\rho = 1.5$, and $k = 0$
  2: Repeat until convergence
     Update $\mathcal{X}, \mathcal{Z}, F, \mathcal{S}, \mathcal{N}, \Gamma_1, \Gamma_2, \Gamma_3$ via (6)–(11)
     Update the parameter $\mu := \min(\rho\mu, \mu_{\max})$
     Check the convergence condition
        $\frac{||\mathcal{X}^{(k)} - \mathcal{X}^{(k+1)}||_F^2}{||\mathcal{Y}||_F^2} \le \epsilon$

---

5) *Update $\mathcal{N}$:* Extracting all items related to $\mathcal{N}$ from (6), it can be easily deduced

$$\mathcal{N}^{(k+1)} = \underset{\mathcal{N}}{\arg\min} \; \beta\|\mathcal{N}\|_F^2 + \left\langle \Gamma_1^{(k)}, \mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S}^{(k+1)} - \mathcal{N} \right\rangle$$

$$+ \frac{\mu}{2}\|\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S}^{(k+1)} - \mathcal{N}\|_F^2$$

$$= \underset{\mathcal{N}}{\arg\min} \left(\beta + \frac{\mu}{2}\right) \left\| \mathcal{N} - \frac{\mu\left(\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S}^{(k+1)}\right) + \Gamma_1^{(k)}}{\mu + 2\beta} \right\|_F^2.$$

Thus, through a simple calculation, we can get its solution as follows:

$$\mathcal{N}^{k+1} = \frac{\mu\left(\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S}^{(k+1)}\right) + M_1^{(k)}}{\mu + 2\beta}. \qquad (11)$$

6) *Updating multipliers:* Based on the ALM algorithm, the multipliers are updated by the following equations:

$$\begin{cases} \Gamma_1^{(k+1)} = \Gamma_1^{(k)} + \mu\left(\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S}^{(k+1)} - \mathcal{N}^{(k+1)}\right) \\ \Gamma_2^{(k+1)} = \Gamma_2^{(k)} + \mu\left(\mathcal{X}^{(k+1)} - \mathcal{Z}^{(k+1)}\right) \\ \Gamma_3^{(k+1)} = \Gamma_3^{(k)} + \mu\left(D_w(\mathcal{Z}^{(k+1)}) - \mathcal{F}^{(k+1)}\right). \end{cases} \qquad (12)$$

Summarizing the aforementioned discussion, we now arrive at obtaining an ALM method to solve the proposed LRTDTV model (3), as presented in Algorithm 1.

In the LRTDTV solver, the inputs are the noisy image $\mathcal{Y} \in \mathbb{R}^{M \times N \times p}$, desired rank $[r_1, r_2, r_3]$ for Tucker decomposition, the stopping criteria $\epsilon$, and the regularized parameters $\tau$, $\lambda$, and $\beta$. Considering the fact that these three parameters have certain proportional relationship, we simply set $\tau = 1$ and then tune $\lambda$ and $\beta$. More details and discussions will be presented in Section IV-C. For another important parameter $\mu$, we first initialize it as $\mu = 10^{-2}$ and then update it as $\mu = \min(\rho\mu, \mu_{\max})$ in each iteration. This strategy of determining the variable $\mu$ has been widely used in the ALM-based methods [45], [46], which can facilitate the convergence of the algorithm.

### D. LRTDTV Approximate Model

In the LRTDTV model (3), we separate the mixture noise into two parts, namely, the sparse noise and the Gaussian noise.

Because the TV regularization has certain effect on removing the Gaussian noise, when the variance of the Gaussian noise is not heavy, we can simplify (3) as follows:

$$\min_{\mathcal{X}, \mathcal{S}} \tau\|\mathcal{X}\|_{\text{SSTV}} + \lambda\|\mathcal{S}\|_1$$

$$\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{S}$$

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \mathbf{U}_i^T \mathbf{U}_i = \mathbf{I} \; (i = 1, 2, 3). \quad (13)$$

Similarly, we can derive an ALM-based optimization procedure. For simplicity, we directly give the closed-form solution of each subproblem in the following.

*Solving $\mathcal{X}$ subproblem:*
We need to consider the following optimization problem:

$$\min_{\mathbf{U}_i^T \mathbf{U}_i = \mathbf{I}} \mu \left\| \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 - \frac{1}{2}\left(\mathcal{Y} - \mathcal{S}^{(k)} + \mathcal{Z}^{(k)}\right.\right.$$

$$\left.\left. + \left(\Gamma_1^{(k)} - \Gamma_2^{(k)}\right)/\mu\right) \right\|_F^2.$$

By using the classical HOOI algorithm [43], we can get the estimation of the core tensor $\mathcal{C}^{(k)}$ and $\mathbf{U}_i^{(k)}$ $(i = 1, 2, 3)$. Then, $\mathcal{X}^{(k+1)}$ can be updated as follows:

$$\mathcal{X}^{(k+1)} = \mathcal{C}^{(k+1)} \times_1 U_1^{(k+1)} \times_2 U_2^{(k+1)} \times_3 U_3^{(k+1)}. \quad (14)$$

*Solving $\mathcal{Z}$ subproblem:*

$$\begin{cases} H_z = \mu\mathcal{X}^{(k+1)} + \mu D_w^*\left(\mathcal{F}^{(k)}\right) + \Gamma_2^{(k)} - D_w^*\left(\Gamma_3^{(k)}\right) \\ T_z = w_1^2|\text{fftn}(D_h)|^2 + w_2^2|\text{fftn}(D_v)|^2 + w_3^2|\text{fftn}(D_t)|^2 \\ \mathcal{Z}^{(k+1)} = \text{ifftn}\left(\frac{\text{fftn}(H_z)}{\mu\mathbf{1} + \mu T_z}\right). \end{cases} \quad (15)$$

*Solving $\mathcal{F}$ subproblem:*

$$\mathcal{F}^{k+1} = \mathcal{R}_{\frac{\tau}{\mu}}\left(D_w\left(\mathcal{Z}^{(k+1)}\right) + \frac{\Gamma_3^{(k)}}{\mu}\right). \quad (16)$$

*Solving $\mathcal{S}$ subproblem:*

$$\mathcal{S}^{k+1} = \mathcal{R}_{\frac{\lambda}{\mu}}\left(\mathcal{Y} - \mathcal{X}^{(k+1)} + \frac{\Gamma_1^{(k)}}{\mu}\right). \quad (17)$$

The multipliers are updated by the following equations:

$$\begin{cases} \Gamma_1^{(k+1)} = \Gamma_1^{(k)} + \mu\left(\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{S}^{(k+1)}\right) \\ \Gamma_2^{(k+1)} = \Gamma_2^{(k)} + \mu\left(\mathcal{X}^{(k+1)} - \mathcal{Z}^{(k+1)}\right) \\ \Gamma_3^{(k+1)} = \Gamma_3^{(k)} + \mu\left(D_w\left(\mathcal{Z}^{(k+1)}\right) - \mathcal{F}^{(k+1)}\right). \end{cases} \quad (18)$$

Similarly, we can get an efficient ALM method to solve the LRTDTV approximate model (13), that is, sequentially update (13)–(17) until certain convergence condition is satisfied. It should be noted that all the parameters used for updating (13)–(17) can be set as the same values used in Algorithm 1.

### E. Complexity Analysis

For an input noisy HSI with the size $h \times w \times B$, the main per-iteration cost of the LRTDTV solver lies in the subproblems

of updating of $\mathcal{X}$ and $\mathcal{Z}$, since other subproblems can be solved through a simple algebraic calculation. Updating $\mathcal{X}$ resorts to call the HOOI algorithm to get the estimation of $\mathbf{U}_i$ ($i = 1, 2, 3$), which requires performing several SVDs. This leads to $O(T(h^2 r_2 r_3 + (r_2 r_3)^3 + w^2 r_1 r_3 + (r_1 r_3)^3 + B^2 r_1 r_2 + (r_1 r_2)^3))$ cost where $T$ is the number of iterations used in HOOI, because of the fact that the computational complexity of SVD for an $M \times N$ matrix is $O(M^2 N + N^3)$. Then, updating $\mathcal{X}$ via (7) gives $O(r_2 r_3 \times r_1 \times h + r_3 h \times r_2 \times w + hw \times r_3 \times B)$ cost, due to the computational complexity of matrix multiplication. It is known that the computational complexity of FFT is $O(N \log N)$ where $N$ is the data size. As such, the main cost of the subproblem of updating $\mathcal{Z}$ is $O(hwB \log(hwB))$. In all, the complexity of the proposed LRTDTV solver is $O(T(h^2 r_2 r_3 + (r_2 r_3)^3 + w^2 r_1 r_3 + (r_1 r_3)^3 + B^2 r_1 r_2 + (r_1 r_2)^3) + r_2 r_3 r_1 h + r_3 h r_2 w + hw r_3 B + hwB \log(hwB))$. This complexity is comparable with the existing tensor-based HSI denoising methods, such as TDL [38] and BM4D [47], and thus reasonable in application.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

Both simulated and real image data experiments were carried out to demonstrate the effectiveness of the proposed LRTDTV method for HSI restoration. To thoroughly evaluate the performance of LRTDTV, we implemented seven different popular HSI restoration methods for comparison, i.e., the nuclear norm minimization (NNM)-based LRMA [28], the WNNM-based LRMA [31], the GoDec-based low-rank matrix recovery (LRMR) [13], the weighted Schatten norm minimization (WSNM)-based LRMA [33], the LRTV [17], the block matching 4-D filtering (BM4D) [47], the robust low-rank tensor recovery (RLRTR) [40], and the decomposable nonlocal tensor dictionary learning (TDL) [38]. These methods represent the state-of-the-art HSI restoration methods especially WSNM and LRTV, and their implementation codes can be directly obtained from the authors' websites.

In all the following experiments, the parameters in these compared methods were manually adjusted according to their default strategies. For our LRTDTV solver, we would like to present a detailed discussion of parameter setting in Section IV-C. In addition, to facilitate the numerical calculation and visualization, all the bands of the HSI are normalized into [0, 1] and they will be stretched to the original level after restoration.

### A. Simulated Data Experiments

The synthetic data were first considered by [13], which were generated using the ground truth of the Indian Pines dataset [48]. The size of the synthetic HSI was $145 \times 145 \times 224$, and the reflectance values of all the voxels in the HSI were linearly mapped to [0, 1].

To simulate noisy HSI data, we added several types of noise to the original synthetic HSI data under six different cases to test the performance of all the compared restoration methods, both in visual quality and quantitative perspective. We list the details in the following.

*Case 1:* For different bands, the noise intensity was equal. And the same distribution of zero-mean Gaussian noise was added to each band. The variance of the Gaussian noise was 0.1.

*Case 2:* In this case, the Gaussian noise was added to each band just as in case 1. In addition, some deadlines were added from bands 91 to 130, with the number of stripes randomly selected from 3 to 10, and the width of the stripes randomly generated from 1 to 3.

*Case 3:* In this case, the noise intensity was equal for different bands. And the same distribution of zero-mean Gaussian noise and the same percentage of impulse noise were added to each band. The variance of the Gaussian noise was 0.075, and the percentage of the impulse noise was 0.15.

*Case 4:* In this case, the Gaussian and impulse noise were added just like in case 3. Besides, the deadlines were added from bands 91 to 130, with the number of stripes randomly selected from 3 to 10, and the width of the deadlines randomly generated from 1 to 3.

*Case 5:* In this case, the noise intensity was different for different bands. That is, Gaussian noise of zero-mean and different variances was added to each band, with the variance value randomly selected from 0 to 0.2, and different percentages of impulse noise were added, which were randomly selected from 0 to 0.2. In addition, the deadline noise was added to some bands just as in case 4.

*Case 6:* In this case, the Gaussian noise, impulse noise, and deadlines were added just as in case 5. In addition, some stripes were added from bands 161 to 190, with the number of stripes randomly selected from 20 to 40.

*1) Visual Quality Comparison:* In terms of visual quality, two representative bands of restored HSIs in two typical cases obtained by different methods are shown in Figs. 5 and 6. More precisely, Fig. 5 shows the restoration results of band 36, which is corrupted by both the Gaussian noise and impulse noise in case 3. And Fig. 6 shows the restoration results of band 116, which is severely corrupted by a mixture of Gaussian noise, impulse noise, and deadline in case 5.

It is evident that, for both cases, the original clean HSIs have undergone great changes, as shown in Figs. 5(b) and 6(b). Aiming at better visual comparison, we marked the same subregion of each subfigure in Figs. 5 and 6 by a green box and then enlarged it in a red box. Several observations can be easily made from both Figs. 5 and 6. First, all the compared methods, to some extent, could remove such mixed noise. Second, the proposed LRTDTV method performs best among all the compared methods, effectively removing the noise while preserving the essential structures of the original HSIs. For instance, compared to the red dashed boxes in Fig. 6(c)–(i), the red dashed box in Fig. 6(j) produced by our LRTDTV method preserves clearer and sharper edges. Third, the two TV-regularized methods, i.e., LRTV and LRTDTV, perform much better than the other competing methods, and our proposed LRTDTV outperforms LRTV to a certain extent. This demonstrates the power of both the TV regularization and direct low-rank tensor modeling techniques for HSI restoration.

*2) Quantitative Comparison:* To better understand the overall performance of different methods, the mean peak
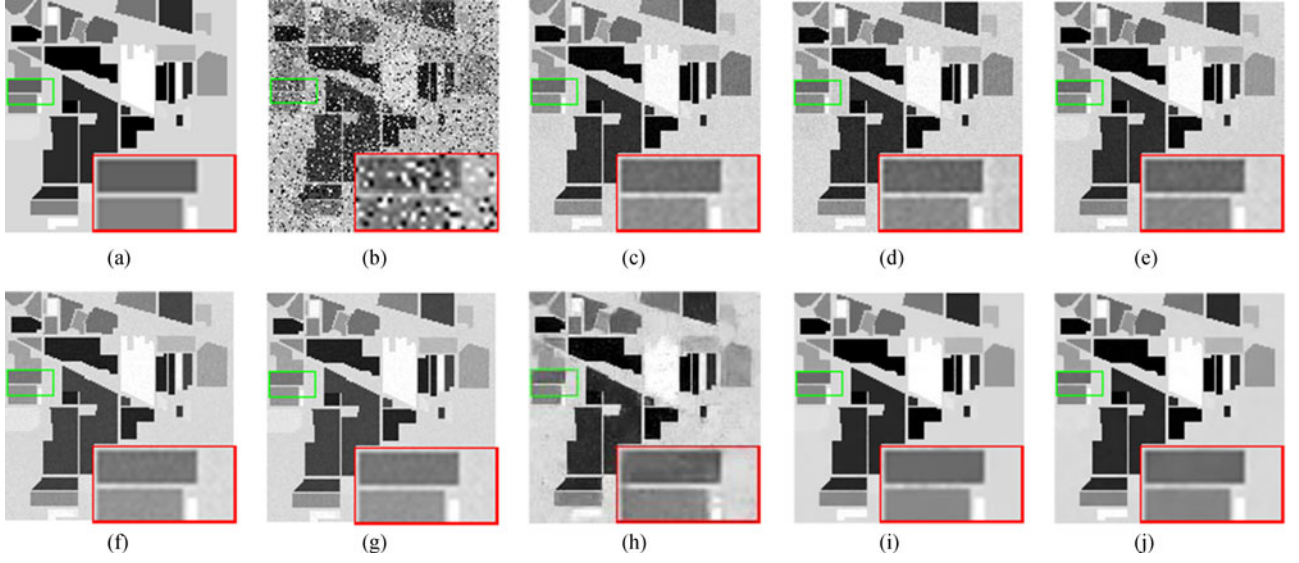
Fig. 5. Denoised results by all the compared methods: (a) the original band 36, (b) the simulated noise band of case 3, (c) WNNM, (d) LRMR, (e) RLRTR, (f) BM4D, (g) TDL, (h) WSNM, (i) LRTV, and (j) LRTDTV.
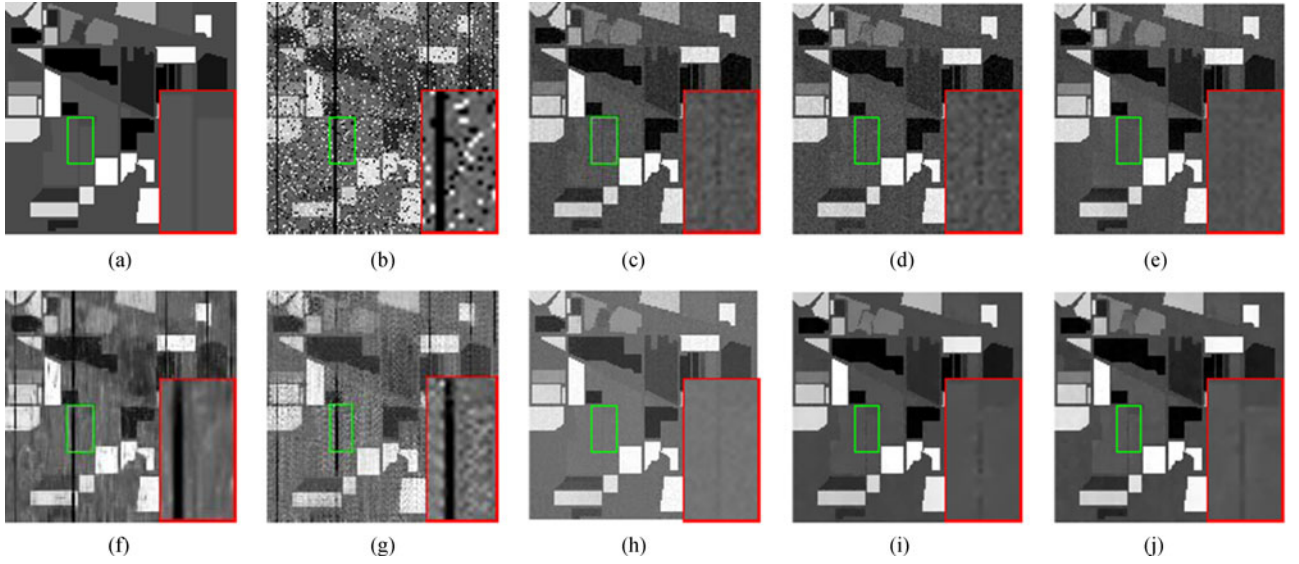


Fig. 6. Denoised results by all the compared methods: (a) the original band 116, (b) the simulated noise band of case 5, (c) WNNM, (d) LRMR, (e) RLRTR, (f) BM4D, (g) TDL, (h) WSNM, (i) LRTV, and (j) LRTDTV.

signal-to-noise ratio (MPSNR), mean structural similarity index (MSSIM), and Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) are adopted to objectively evaluate the restoration quality:

$$ \text{MPSNR} = \frac{1}{B} \sum_{i=1}^{B} \text{psnr}_i, $$

$$ \text{MSSIM} = \frac{1}{B} \sum_{i=1}^{B} \text{ssim}_i, $$

$$ \text{ERGAS} = \sqrt{\frac{1}{B} \sum_{i=1}^{B} \frac{\text{mse}(\text{ref}_i, \text{res}_i)}{\text{Mean}_2(\text{ref}_i)}} $$

where $\text{psnr}_i$ and $\text{ssim}_i$ denote the PSNR and SSIM values for the $i$th band, and $\text{ref}_i$ and $\text{res}_i$ denote the $i$th band of the reference image and the restoration image, respectively.

Table II presents the restoration results by all the compared methods in terms of the aforementioned three indices. Because the Gaussian noise is not the dominant noise in cases 2–6, we adopt the approximation model (15) in such cases except case 1 where the Gaussian noise dominates other noises. It is clear that the proposed LRTDTV enjoys a superior performance over the other popular approaches. We also show in Fig. 7 the PSNR and SSIM values of each band in all the simulated data cases. It is easy to see that the TV-regularized methods including LRTV and the proposed LRTDTV can get much higher SSIM and PSNR values than other ones for almost every band. In addition, except

TABLE II
QUANTITATIVE EVALUATION OF DIFFERENT METHODS ON THE SIMULATED DATA IN DIFFERENT CASES

| Noise case | Evaluation index | Noise | NNM | WNNM | LRMR | RLRTR | BM4D | TDL | WSNM | LRTV | LRTDTV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Case 1 | MPSNR (dB) | 19.99 | 30.97 | 32.58 | 36.20 | 33.32 | 38.44 | 38.05 | 37.32 | 38.68 | **40.76** |
| | MSSIM | 0.3672 | 0.8727 | 0.8420 | 0.9311 | 0.8632 | 0.9763 | 0.9674 | 0.9453 | **0.9853** | 0.9804 |
| | ERGAS | 233.99 | 69.02 | 57.65 | 36.85 | 53.38 | 29.04 | 30.72 | 32.37 | 28.47 | **23.02** |
| Case 2 | MPSNR (dB) | 19.34 | 30.81 | 32.46 | 35.67 | 33.03 | 35.10 | 33.22 | 36.12 | 38.04 | **40.54** |
| | MSSIM | 0.3592 | 0.8715 | 0.8415 | 0.9291 | 0.8570 | 0.9391 | 0.8743 | 0.9402 | 0.9818 | **0.9895** |
| | ERGAS | 257.88 | 70.06 | 58.39 | 39.84 | 54.70 | 110.40 | 112.35 | 44.89 | 49.18 | **23.44** |
| Case 3 | MPSNR (dB) | 13.07 | 31.61 | 32.36 | 36.40 | 32.92 | 28.59 | 27.50 | 38.14 | 39.54 | **41.08** |
| | MSSIM | 0.1778 | 0.8889 | 0.8786 | 0.9345 | 0.8931 | 0.8401 | 0.9076 | 0.9551 | 0.9866 | **0.9910** |
| | ERGAS | 520.53 | 64.36 | 59.71 | 36.04 | 55.88 | 90.29 | 102.08 | 29.52 | 35.22 | **21.98** |
| Case 4 | MPSNR (dB) | 12.92 | 31.45 | 32.29 | 35.76 | 32.66 | 27.02 | 26.54 | 36.63 | 38.75 | **40.72** |
| | MSSIM | 0.1748 | 0.8878 | 0.8777 | 0.9316 | 0.8885 | 0.8073 | 0.8365 | 0.9485 | 0.9826 | **0.9906** |
| | ERGAS | 529.82 | 65.45 | 60.10 | 39.99 | 57.33 | 128.11 | 120.54 | 46.32 | 55.44 | **22.90** |
| Case 5 | MPSNR (dB) | 13.80 | 29.62 | 31.33 | 33.72 | 31.82 | 27.95 | 24.34 | 35.02 | 36.54 | **38.83** |
| | MSSIM | 0.2038 | 0.8633 | 0.8445 | 0.8951 | 0.8596 | 0.8201 | 0.5931 | 0.9285 | 0.9742 | **0.9859** |
| | ERGAS | 500.68 | 81.16 | 66.39 | 50.16 | 62.53 | 121.95 | 158.75 | 51.33 | 72.52 | **28.66** |
| Case 6 | MPSNR (dB) | 13.73 | 29.54 | 29.97 | 33.42 | 30.45 | 27.53 | 23.34 | 33.88 | 36.35 | **38.63** |
| | MSSIM | 0.2022 | 0.8612 | 0.8431 | 0.8918 | 0.8542 | 0.8060 | 0.5583 | 0.9261 | 0.9736 | **0.9852** |
| | ERGAS | 504.37 | 82.18 | 82.48 | 52.62 | 80.87 | 127.28 | 174.28 | 53.29 | 72.05 | **29.82** |
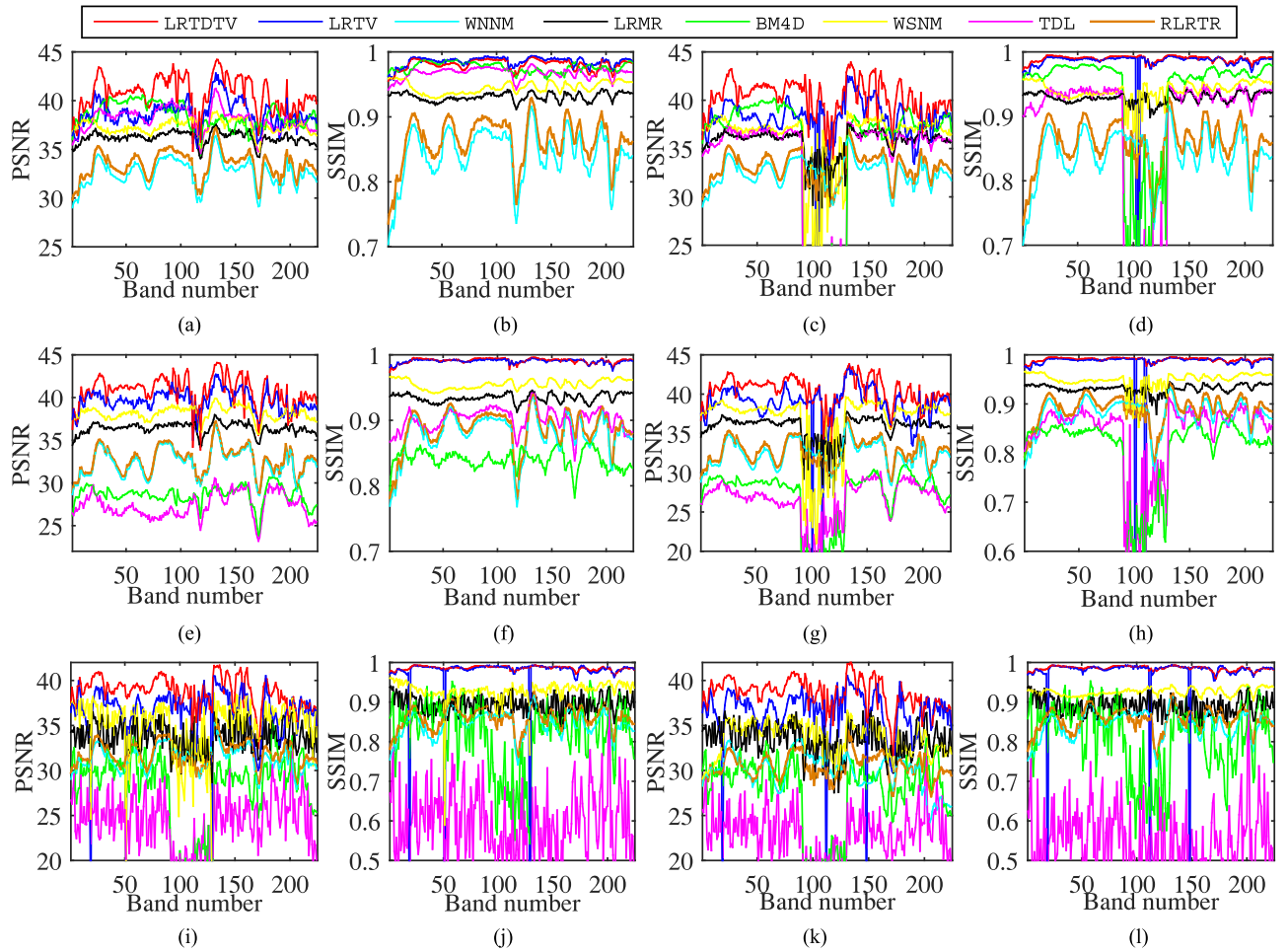


Fig. 7. Detailed quantitative evaluation of different methods for each band: (a) and (b) case 1, (c) and (d) case 2, (e) and (f) case 3, (g) and (h) case 4, (i) and (j) case 5, and (k) and (l) case 6.
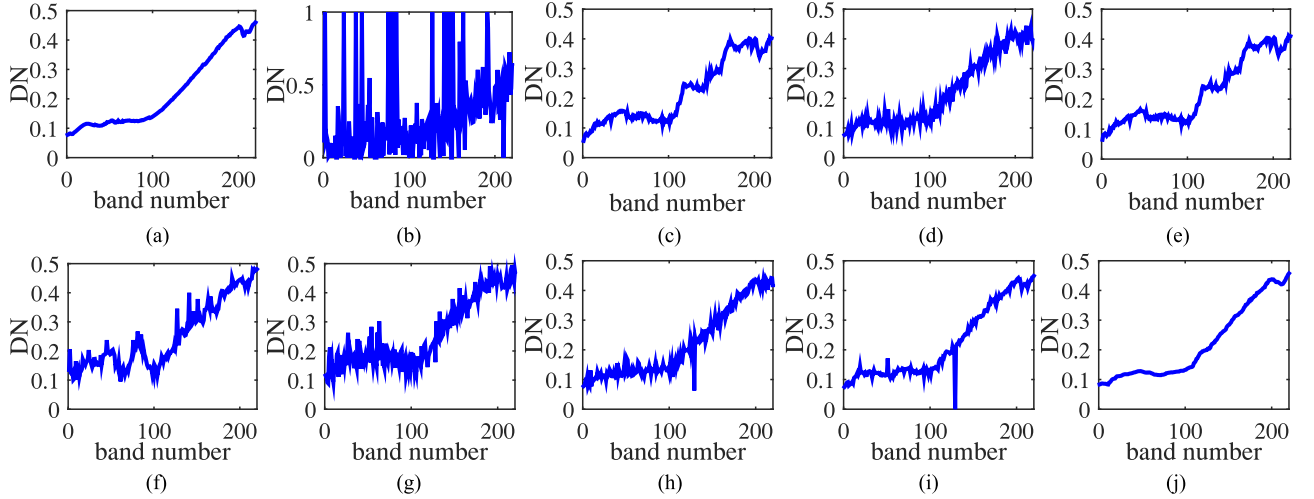
Fig. 8. Spectral signatures curve estimated by all the compared methods: (a) the original spectral signatures of pixel (20,20), (b) the noise spectral signatures, (c) WNNM, (d) LRMR, (e) RLRTR, (f) BM4D, (g) TDL, (h) WSNM, (i) LRTV, and (j) LRTDTV.

that the SSIM values of LRTV are slightly higher than those of LRTDTV in case 1, LRTDTV achieves the best performance among all the methods in terms of both SSIM and PSNR.

To further compare the performances of all the restoration methods, it would be necessary to show the spectral signatures before and after restoration. As such, Fig. 8 shows one simple example, namely, the spectral signatures of pixel (20, 20) in case 5. Combining with the ERGAS values in Table II, it can be clearly seen that the proposed LRTDTV method produces the best spectral signature among all the compared methods.

### B. Real Data Experiments

Two real-world HSI datasets were used in our experiments, i.e., the hyperspectral digital imagery collection experiment (HYDICE) urban dataset [49] and the airborne visible/infrared imaging spectrometer (AVIRIS) Indian Pines dataset [48]. Before conducting the restoration process, the gray values of each HSI band were normalized to [0, 1]. We have implemented the proposed two models (3) and (13), and indeed found that they performed similar in these real data experiments. Therefore, we only present the restoration results of the approximation model (13) for visual comparison.

*1) HYDICE Urban Dataset:* The size of the original image is $307 \times 307 \times 210$. As shown in Fig. 9, the full urban image is polluted by stripes, deadlines, the atmosphere, water absorption, and other unknown noise. In this experiment, NNM, WNNM, BM4D, and TDL were implemented using the default values of the parameters. For other competing methods, we manually adjusted their parameters to give the best visual results. For instance, the two parameters of LRMR, i.e., $r$ and $k$, were set for 3 and 9500, respectively.

Figs. 10 and 11 show the restorations of bands 109 and 207, respectively. It is not hard to see that several low-rank matrix methods, including NNM, WNNM, LRMR, and WSNM, cannot effectively remove the stripes. In addition, some structures of the objects are not clearly differentiated. This is mainly
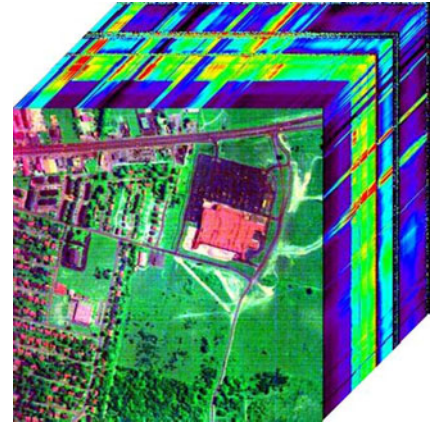


Fig. 9. HYDICE urban dataset used in first real data experiment.

because of the fact that stripes and deadlines exist in the same place from bands 104 to 110 and from bands 199 to 210. That is, in the low-rank and sparse decomposition, the stripes are more likely regarded as the low-rank part, which is assumed to be the clean image. As for BM4D, though it is a tensor filtering based method, it mainly assumes that the noise is Gaussian, so once the HSIs suffer from heavy sparse noise or some structure noise including stripes, this method cannot perform well. As for TDL, it loses its effectiveness when HSIs suffer from heavy mixture of corrupted noise since it was also designed for removing Gaussian noise. As for RLRTR, it does not model the local smoothness structure of the HSI cube, which makes it difficult to give a good denoising result. Similar to the aforementioned, since LRTV uses the spatial TV regularization to exploit the spatial information, it can remove this kind of stripes and preserve the edge information to a certain extent. By combining the spectral TV and spatial TV into a unified SSTV regularization and utilizing the low-rank Tucker decomposition to encode the global spectral-and-spatial correlation, our proposed LRTDTV method can better remove the complex mixed noise
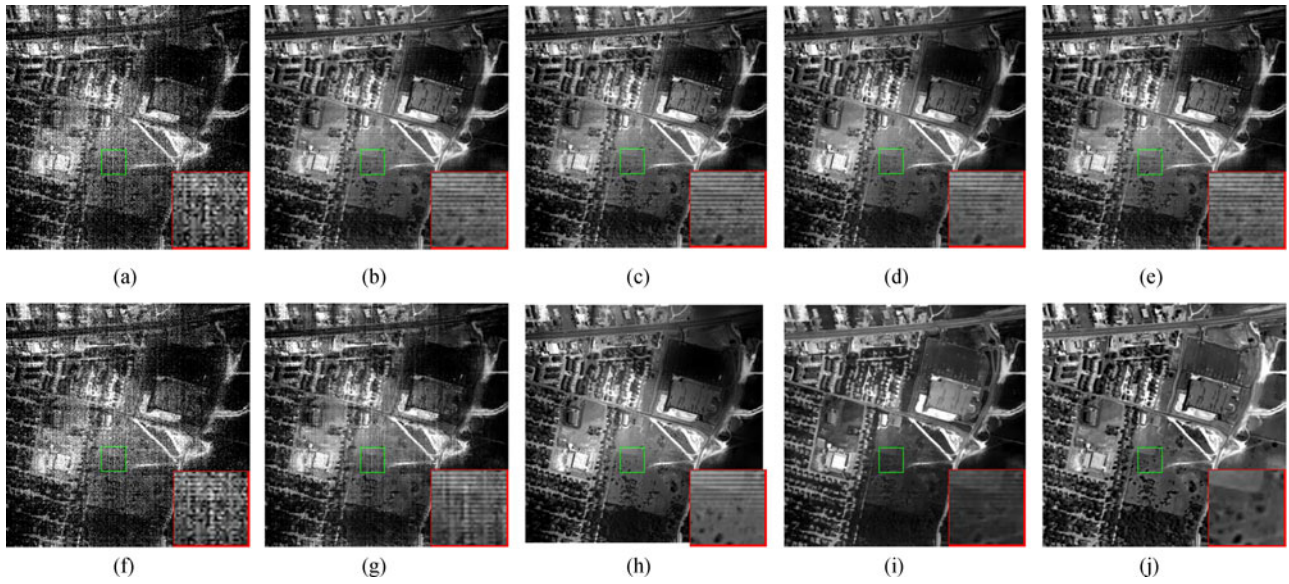
Fig. 10. Denoised results by different methods: (a) the original band 109, (b) NNM, (c) WNNM, (d) LRMR, (e) RLRTR, (f) TDL, (g) BM4D, (h) WSNM, (i) LRTV, and (j) LRTDTV.
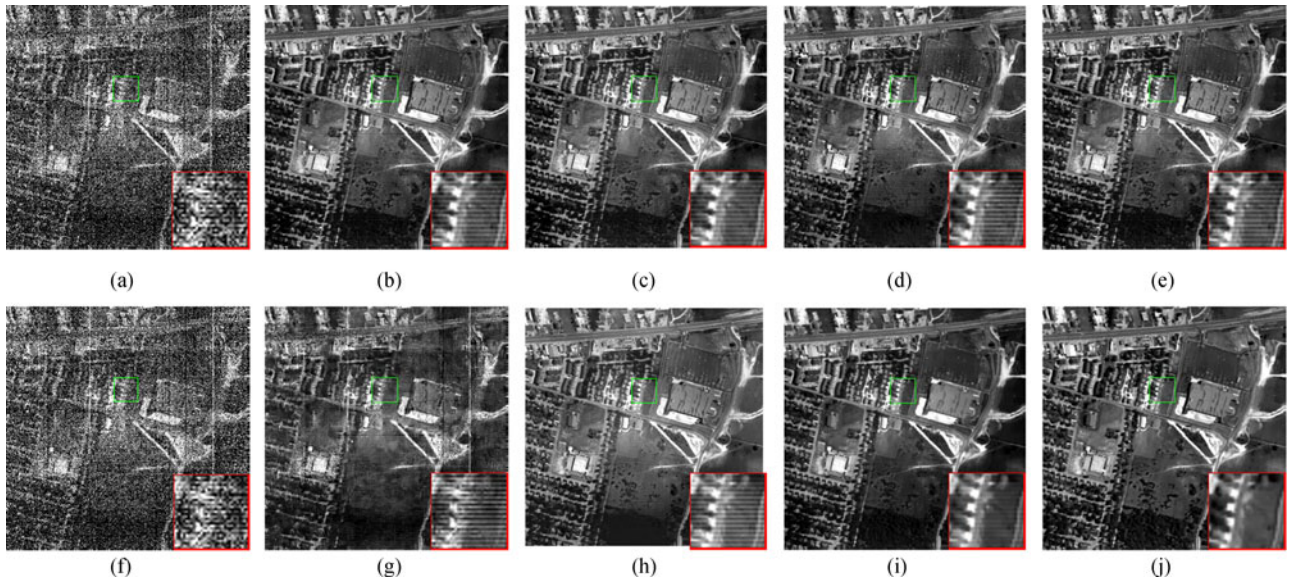


Fig. 11. Denoised results by different methods: (a) the original band 207, (b) NNM, (c) WNNM, (d) LRMR, (e) RLRTR, (f) TDL, (g) BM4D, (h) WSNM, (i) LRTV, and (j) LRTDTV.

and preserve the spatial texture information compared with the other methods.

Fig. 12 shows that the horizontal mean profiles of band 109 before and after restoration. As shown in Fig. 12(a), due to the existence of mixed noise, especially stripes and deadlines, there are rapid fluctuations in the curve before processing the restoration. After restoration, the fluctuations are more or less suppressed by all the methods. Here, one can see that the curve of the proposed LRTDTV method is most stable, which is in accordance with the visual results presented in Fig. 10.

*2) AVIRIS Indian Pines Dataset:* This dataset was acquired by the NASA AVIRIS instrument over the Indian Pines test site

in Northwestern Indiana in 1992, and it has $145 \times 145$ pixels and 220 bands. The full urban image shown in Fig. 13 is mainly corrupted by deadlines, atmosphere, water absorption, and others. Similar to HYDICE data experiment, we implemented WNNM, BM4D, and TDL using the default parameters and the rest using the fine-tuned parameters.

We selected two typical bands, namely, bands 108 and 220, to present the performance of all the compared methods in Figs. 14 and 15. Here, it is easy to observe that low-rank matrix methods including NNM, WNNM, LRMR, and WSNM can more or less remove some noise, but when the noise is heavy, such methods lose their utility and even cause some degradation
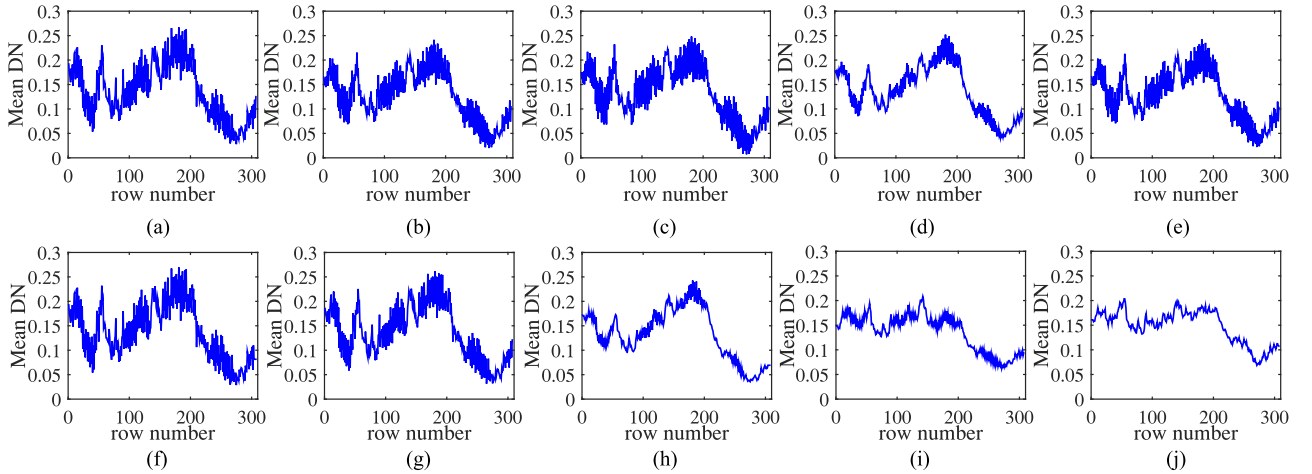
Fig. 12.    Spectral signatures curve estimated by different methods: (a) the original band 109, (b) NNM, (c) WNNM, (d) LRMR, (e) RLRTR, (f) TDL, (g) BM4D, (h) WSNM, (i) LRTV, and (j) LRTDTV.
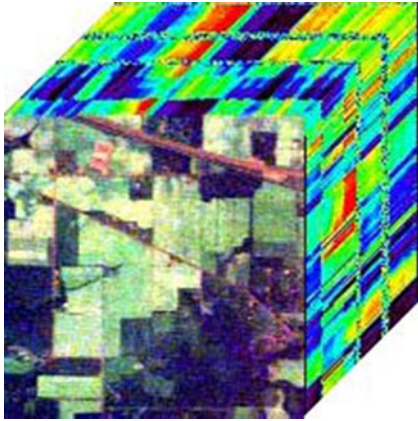


Fig. 13.    AVIRIS Indian Pines dataset used in second real data experiment.

of the gray value. See Fig. 14(h) for an example of WSNM. As for such tensor-based methods including BM4D, TDL, and RLRTR, these would fail to give satisfactory results because of the fact that some intrinsic information such as the local smoothness underlying the HSI cube was not exploited. See Fig. 15(e) for an example of RLRTR. Compared to the afore-mentioned non-TV-regularized methods, LRTV and the proposed LRTDTV can remove lots of noise, but LRTV cannot preserve edges and local detail information as well as LRTDTV. More detailed visual comparison results can be seen in such red boxes. In a word, for this dataset, the proposed LRTDTV can still get the best performance for removing such heavy mixed noise.

We also present the vertical mean profiles of band 108 before and after in Fig. 16. Similar to Fig. 12, it can again be clearly observed that LRTDTV gives the best curve among all the restored vertical mean profiles.

To sum up, extensive experiments on both simulated and real data demonstrate the clear superiority of LRTDTV over other popular methods, because more useful structures are delivered.

## C. Discussion

*1) Parameter Setting:* Basically, we introduced two LRT-DTV models, one is the general model (3), the other is the approximate model (13). When the noisy HSI was corrupted by very heavy Gaussian noise, the general model (3) performs better than the approximate model (13) because of the use of the additional Frobenius norm term for modeling the Gaussian noise. In fact, the noise of real HSI is usually a mixture of several types of noise, and compared to other noise terms, the extent of Gaussian noise was not severe. Thus, considering the fact that the TV regularization has the ability of removing Gaussian noise, we would like to use the approximate model (13) for simplicity.

In the LRTDTV model, there exist several parameters that need to be carefully identified. Specifically, for the Frobenius term regularization parameter $\beta$, we set it as the reciprocal of the variance of Gaussian noise in the general model, just as what WNNM did; for the TV regularization parameter $\tau$ in such two models, similar to many other works, it can be fixed as the constant 1. Then, we need to consider how to select the $\ell_1$ term regularization parameter $\lambda$, the rank $r_i$s among three dimensions of the HSI cube, and the weights of SSTV. In the following, we shall provide the reasons for choosing such parameters in our experiments. All the results were based on the simulated data experiment in case 5.

*Sensitivity analysis of the parameter $\lambda$:* It is easy to see that $\lambda$ is the parameter used to restrict the sparsity of the sparse noise. As stated in the RPCA model [28], the sparsity regularization parameter was set for $\lambda = 1/\sqrt{MN}$ for one matrix with size $M \times N$, which was good enough to guarantee the existence of an optimal solution. By bringing the new TV penalty, our model is different from RPCA, we, thus, set $\lambda = 100 \times C/\sqrt{hw}$, where $h$ and $w$ are the height and the width of one HSI band, respectively, and $C$ is a tuning parameter. Fig. 17 shows the restoration results as $C$ varied in the set {1, 2, 5, 8, 10, 15, 20, 25, 30, 40, 50}. It can be clearly observed from this figure that the results of the LRTV solver are relatively stable in terms of both MPSNR and
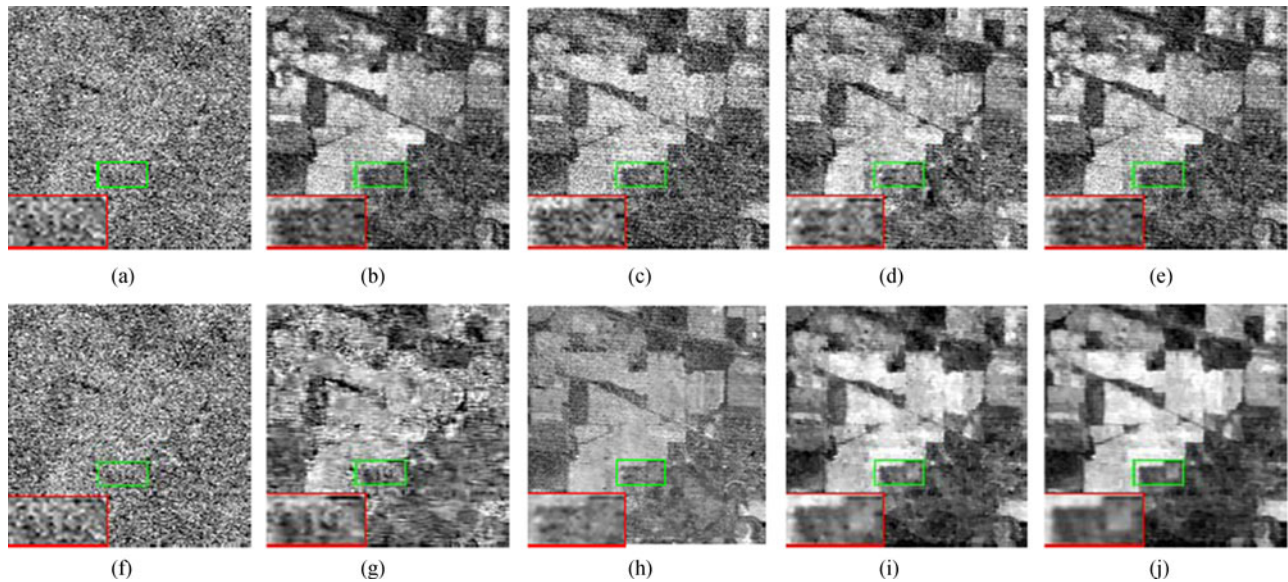
Fig. 14. Denoised results by different methods: (a) the original band 108, (b) NNM, (c) WNNM, (d) LRMR, (e) RLRTR, (f) TDL, (g) BM4D, (h) WSNM, (i) LRTV, and (j) LRTDTV.
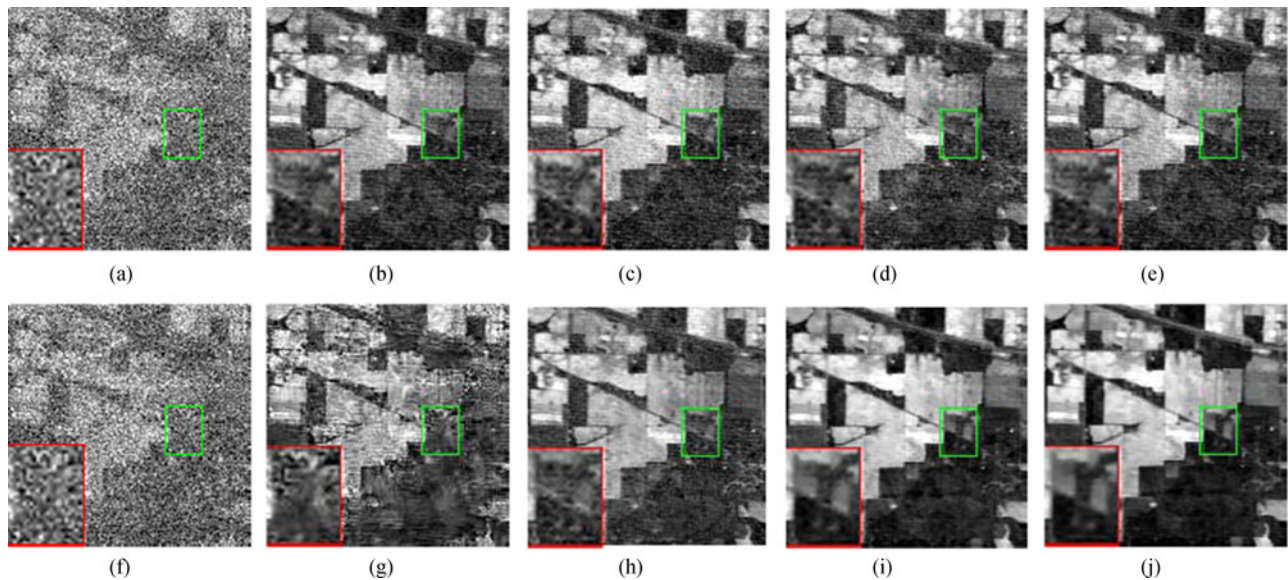


Fig. 15. Denoised results by different methods: (a) the original band 220, (b) NNM, (c) WNNM, (d) LRMR, (e) RLRTR, (f) TDL, (g) BM4D, (h) WSNM, (i) LRTV, and (j) LRTDTV.

MSSIM values, with the value of $C$ changing from 10 to 25. Therefore, we suggest the use of $C = 10$ in all the simulated data experiments.

*Effectiveness of the rank constraint:* In the LRTDTV optimization, we adopted the Tucker decomposition to encode the low-rank prior. So, we should give the estimated ranks along the three modes before running the algorithm. For two spatial modes, we empirically adopted the eighty percent of the image size to utilize the low-rank property. Fig. 18 presents the MP-SNR and MSSIM values of the LRTDTV solver with different rank-constrained values of the spectral mode. It can be easily observed that the MPSNR and MSSIM values first increase and

then decrease with the growth of the estimated rank value. This inspires us to use the HSI subspace estimation method (e.g., HySime [50]) to estimate the range of values for the rank, and then choose the best desired rank value among the candidates for the spectral mode. As such, in all our simulation experiments, the value of the spectral rank is set as 10.

*Effectiveness of the weight of the SSTV regularizer:* For an HSI, the spectral characteristics are a very important feature, which can be encoded by the spectral TV prior. This prior can help us remove the noise, especially for some structure noise whose distribution was different on adjacent bands as observed. Then, based on the statistical analysis of the TV value along
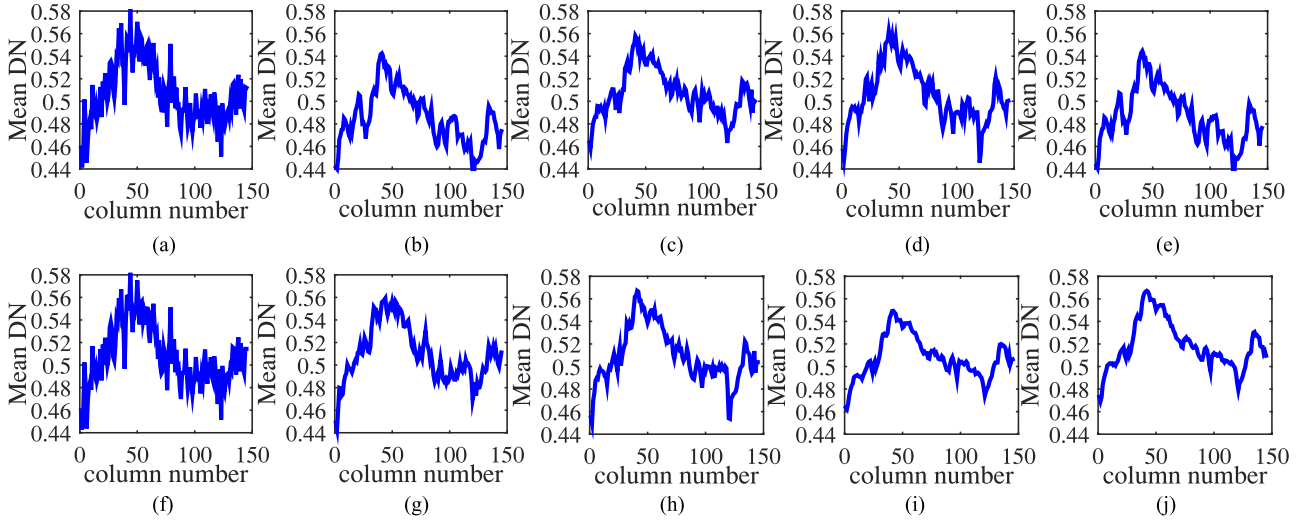
Fig. 16.   Spectral signatures curve estimated by all the compared methods: (a) the original band 108, (b) NNM, (c) WNNM, (d) LRMR, (e) RLRTR, (f) TDL, (g) BM4D, (h) WSNM, (i) LRTV, and (j) LRTDTV.
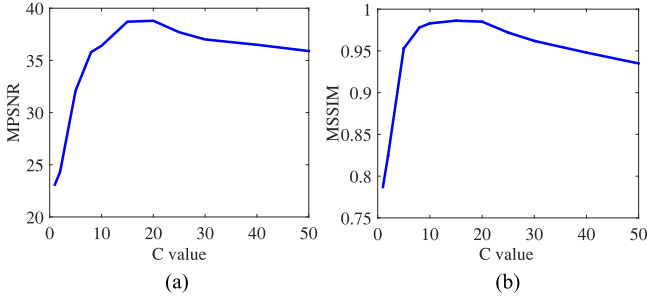


Fig. 17.   Sensitivity analysis of parameter ($C$ from 1 to 50 with $= \frac{100*C}{\sqrt{MN}}$). (a) Change in the MPSNR value and (b) change in the MSSIM value.
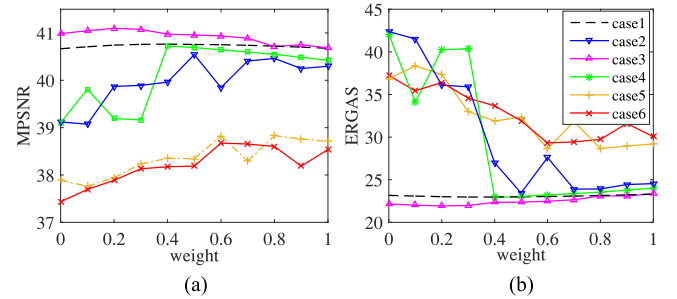


Fig. 19.   Performance with different weights of SSTV. (a) Change in the MPSNR value and (b) change in the ERGAS value.
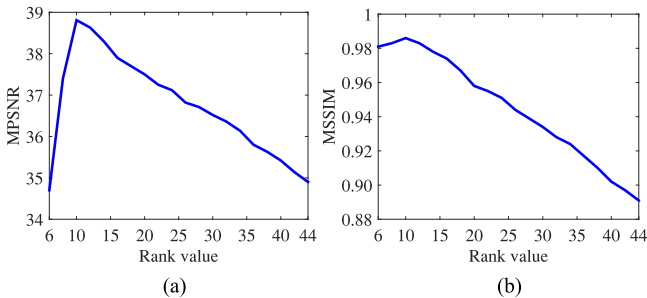


Fig. 18.   Sensitivity analysis of the spectral rank constraint (spectral mode rank values from 6 to 44). (a) Change in the MPSNR value and (b) change in the MSSIM value.



Fig. 20.   MPSNR and MSSIM values versus the iteration number of LRTDTV. (a) Change in the MPSNR value and (b) change in the MSSIM value.

three modes shown in Fig. 4, the weights of the SSTV along the spatial mode can be taken as the equivalent value 1, and we only change the weight of the SSTV along the spectral mode from 0 to 1. The influence of the weight of the spectral TV is presented in Fig. 19, which clearly shows the benefit of tuning the weight of the SSTV.

To sum up, we present Table III to express the parameter setting of the LRTDTV model based on the above-mentioned discussion.
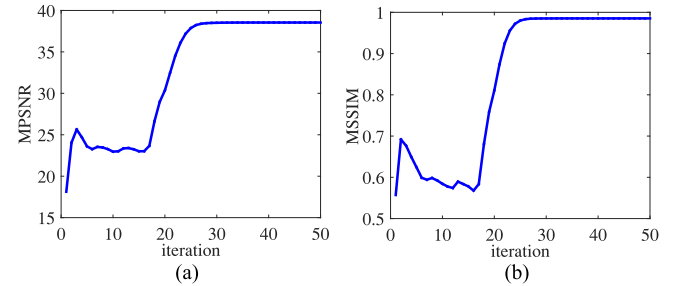
*2) Convergence of the LRTDTV Solver:* We now present an empirical analysis of convergence of our proposed LRTDTV solver. Fig. 20 depicts the MPSNR and MSSIM gains versus the iteration number of the LRTDTV solver on the simulated data experiment in case 5. Here, we can observe that as the number of iteration increases to a relatively large value, the relative changes of the MPSNR and MSSIM converge to zero. This clearly illustrates the empirical convergence behavior of the proposed method, which further supports its use for more practical situations.

TABLE III
PARAMETER SETTING OF LRTDTV

| Parameter | Setting |
|---|---|
| The rank constraint $r = [r_1, r_2, r_3]$ | $r_1$ and $r_2$ are set as 80% of the HSI's height and width, and $r_3$ is selected from a candidate range that estimated by an HSI subspace estimation method, namely, HySime proposed by [50]. |
| The regularization parameter $\lambda$ | $\lambda = 100 \times C/\sqrt{hw}$, where $C$ is a parameter that can be tuned from a small candidate range. |
| The regularization parameter $\tau$ | $\tau$ is generally set as the constant 1. |
| The weight of the SSTV regularizer $w_i$ ($i = 1, 2, 3$) | $w_1$ and $w_2 = 1$ are set as the constant 1 and $w_3$ is tuned from the interval $[0, 1]$. |

TABLE IV
RUNNING TIME COMPARISON (IN SECONDS)

| Data | Size | NNM | LRMR | RLRTR | BM4D | TDL | WSNM | LRTV | LRTDTV |
|---|---|---|---|---|---|---|---|---|---|
| Simulated case 3 | $145 \times 145 \times 224$ | 24.97 | 156.32 | 72.78 | 516.36 | 64.55 | 5727.12 | 140.85 | 320.75 |
| Urban | $307 \times 307 \times 210$ | 94.10 | 592.14 | 330.64 | 2223.5 | 1957.12 | 16 089.24 | 578.56 | 983.16 |

TABLE V
EMPIRICAL ANALYSIS OF EACH REGULARIZER OF THE LRTDTV MODEL

| Indics | Model (3) | Model (13) | No $\ell_1$ | No TV | No Tucker |
|---|---|---|---|---|---|
| | | Simulated case 1 | | | |
| PSNR | 40.76 | 40.14 | 40.68 | 34.51 | 32.39 |
| SSIM | 0.9814 | 0.9807 | 0.9803 | 0.8731 | 0.9077 |
| ERGAS | 23.02 | 24.55 | 23.15 | 46.20 | 58.34 |
| hline | | Simulated case 2 | | | |
| PSNR | 40.66 | 40.58 | 35.98 | 32.63 | 31.32 |
| SSIM | 0.9893 | 0.9895 | 0.9566 | 0.8449 | 0.9001 |
| ERGAS | 23.18 | 23.44 | 55.26 | 64.03 | 65.82 |

*3) Computational Speeds:* We compare the running time of different models on one simulated data and one real data experiments. The details are listed in Table IV. It can be observed that the NNM method is the fastest method among all the compared methods. But, it cannot achieve well restoration results as shown in the previous experimental study. The WSMM method requires running an expensive iterative algorithm per-iteration and, as a result, gives much higher computation time than other competing methods. As for the proposed LRDTV, due to the use of the HOOI algorithm, which needs to perform a number of SVDs, its running time is relatively higher than other SVD-based mixed noise removal methods, namely, TLRTR, LRMR, and LRTV, but significantly lower than BM4D and TDL in dealing with the urban data with a larger size, which should perform the nonlocal search for similar 3-D patches with high computational cost. Actually, the computational cost of the proposed LRTDTV can be significantly reduced if one replaces the HOOI algorithm with the relatively simple higher order singular value decomposition algorithm [43] to update $\mathcal{X}$ in Algorithm 1, at the expense of degenerated denoising result, e.g., around 0.4 dB loss in PSNR.

*4) Effectiveness of Each Regularizer of LRTDTV:* In Table V, we show the role of each regularizer of the general LRT-DTV model (3) through an empirical study. Each regularizer's effect is displayed by removing it from the model (3). From this table, we can observe that each regularizer can bring some improvement in the HSI denoising task especially the Tucker decomposition and the TV term. It can also be observed that there has been no significant difference between the performance of the general model (3) and the approximate model (13) in case 2 with a mixture noise, while performing (3) can provide a certain improvement over (13) in case 1 with heavy Gaussian noise.

## V. CONCLUSION

In this paper, we have proposed a novel tensor-based approach for removing mixed noise in HSIs. Specifically, the low-rank tensor Tucker decomposition is utilized to describe the global spatial-and-spectral correlation among all the HSI bands, and an SSTV regularization is applied to characterize the piecewise smooth structure in both spatial and spectral domains of HSIs. Besides using the commonly used $\ell_1$ norm to detect the sparse noise, we have also considered an additional Frobenius norm term for modeling the heavy Gaussian noise that may exist in some practical situations. Though the resulting nonconvex tensor optimization problem seems to be difficult to solve, we have designed an efficient algorithm based on the ALM method. A series of simulated and real data experiments have been conducted to demonstrate the superior performance of the proposed method over some popular methods in terms of quantitative evaluation and visual comparison.

In the future, we are interested in conducting the following studies. First, we will incorporate the noise modeling idea of [51] into our SSTV-regularized low-rank tensor decomposition framework to further enhance its capability for removing more complex noise in some real-life scenarios. Second, we will extend the deep learning ideas of [52] and [53] to design a deep tensor architecture to learn the multilinear structure of the clean HSI and identify the noise structures of the observed HSI. Finally, we will design a new SSTV regularizer to exploit the gradient similarity among all the HSI bands to further characterize the smooth structure and, as a result, to improve the noise removal capability of the proposed procedure.

## REFERENCES

[1] A. F.-H. Goetz, "Three decades of hyperspectral remote sensing of the earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, 2009.

[2] R. Willett, M. Duarte, M. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014.

[3] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[4] 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.

[5] 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.

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2009, pp. 2272–2279.

[7] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[8] P. Zhong and R. Wang, "Multiple-spectral-band CRFs for denoising junk bands of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2260–2275, Apr. 2013.

[9] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[10] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[11] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.

[12] Y. Wang, R. Niu, and X. Yu, "Anisotropic diffusion for hyperspectral imagery enhancement," *IEEE Sens. J.*, vol. 10, no. 3, pp. 469–477, Mar. 2010.

[13] 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.

[14] 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, 2014.

[15] 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 Obs. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[16] M. Wang, J. Yu, J.-H. Xue, and W. Sun, "Denoising of hyperspectral images using group low-rank representation," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 9, no. 9, pp. 4420–4427, Sep. 2016.

[17] 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.

[18] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 208–220, Jan. 2013.

[19] M. Yuan and C.-H. Zhang, "On tensor completion via nuclear norm minimization," *Found. Comput. Math.*, vol. 16, no. 4, pp. 1031–1068, 2016.

[20] W. Cao *et al.* "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4075–4090, Sep. 2016.

[21] A. Anandkumar *et al.*, "Tensor vs. matrix methods: Robust tensor decomposition under block sparse perturbations," in *Proc. 19th Int. Conf. Artif. Intell. Statist.*, 2016, pp. 268–276.

[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. Vision Pattern Recog.*, 2016, pp. 5249–5257.

[23] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[24] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[25] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[26] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431–1445, Mar. 2016.

[27] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[28] E. Candès *et al.*, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–39, 2011.

[29] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 33–40.

[30] Z. Wu *et al.*, "Total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *J. Electron. Imag.*, vol. 25, no. 1, 2016, Art. no. 013037.

[31] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2014, pp. 2862–2869.

[32] Y. Peng, J. Suo, Q. Dai, and W. Xu, "Reweighted low-rank matrix recovery and its application in image restoration," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2418–2430, Dec. 2014.

[33] 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.

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

[35] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel non-negative tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.

[36] 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.

[37] X. Guo *et al.*, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogram. Remote Sens.*, vol. 83, pp. 50–63, 2013.

[38] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2014, pp. 2949–2956.

[39] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2016, pp. 1692–1700.

[40] C. Li *et al.*, "Hyperspectral image denoising using the robust low-rank tensor recovery." *J. Opt. Soc. Amer. A, Opt., Image Sci. Vision*, vol. 32, no. 9, pp. 1604–1612, 2015.

[41] Z. Wu *et al.*, "Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *Signal Process.*, vol. 131, pp. 202–219, 2017.

[42] S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser, "Structure tensor total variation," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 1090–1122, 2015.

[43] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[44] C. Jiang *et al.*, "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, 2016.

[45] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," arXiv:1009.5055.

[46] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 2014.

[47] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2012.

[48] [Online]. Available: https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html. Accessed on: Apr. 19, 2017.

[49] [Online]. Available: http://www.tec.army.mil/hypercube. Accessed on: Apr. 19, 2017.

[50] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identi-fication," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[51] X. Chen *et al.*, "A general model for robust tensor factorization with unknown noise," arXiv:1705.06755.

[52] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[53] P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Learning efficient sparse and low rank models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1821–1833, Sep. 2015.

**Yao Wang** received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2014.

He was a Visiting Student at the Georgia Institute of Technology from Oct. 2010 to Nov. 2011. He is currently an Assistant Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include statistical signal processing, high-dimensional data analysis, sparse recovery, and computational biology.



**Jiangjun Peng** is currently working toward the Master's degree at the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China.

He was a Research Assistant with the Chinese University of Hong Kong from Mar. 2017 to Oct. 2017. His current research interests include low-matrix/tensor analysis and deep learning.



**Qian Zhao** received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2015.

He was a Visiting Scholar at Carnegie Mellon University, Pittsburgh, PA, USA, from 2013 to 2014. He is currently an Assistant Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include low-matrix/tensor analysis, Bayesian modeling, and self-paced learning.



**Yee Leung** received the B.Sc. degree in geography from The Chinese University of Hong Kong, Shatin, Hong Kong, SAR, in 1972, and the M.Sc. and Ph.D. degrees in geography and the M.S. degree in engineering from the University of Colorado, Boulder, CO, USA, in 1974, 1977, and 1977, respectively.
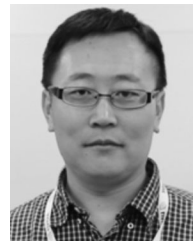
He is currently a Research Professor with the Department of Geography and Resource Management and the Director with the Institute of Future Cities, The Chinese University of Hong Kong. He has authored or co-authored more than 180 papers in reputed journals. His current research interests include the development and applications of intelligent spatial decision support systems, spatial data mining, fuzzy sets and logic, neural networks, and evolutionary computation.

Dr. Leung serves in several journal editorial boards.



**Xi-Le Zhao** received the M.S. and Ph.D. degrees both in applied mathematics from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively.

He is currently a Professor with the School of Mathematical Sciences, UESTC. His current research interests include image processing, computer vision, and machine learning.



**Deyu Meng** (M'09) received the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2008.

He is currently a Professor with the Institute for Information and System Sciences, Xi'an Jiaotong University. From 2012 to 2014, he took his two-year sabbatical leave from Carnegie Mellon University. His current research interests include self-paced learning, noise modeling, weekly supervised learning, and tensor sparsity.