A Novel 3D Anisotropic Total Variation Regularized Low Rank Method for Hyperspectral Image Mixed Denoising

Le Sun [1,2,3,†] ⓘ, **Tianming Zhan** [4], **Zebin Wu** [3] **and Byeungwoo Jeon** [5,*]

1   School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China; sunlecncom@nuist.edu.cn
2   Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China
3   School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; zebin.wu@gmail.com
4   School of Information and Technology, Nanjing Audit University, Nanjing 211815, China; ztm@ujs.edu.cn
5   Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 440746, Korea
*   Correspondence: bjeon@skku.edu; Tel.: +82-10-3709-7144
†   This paper is extended from our PIC2016 conference paper.

check for updates

Abstract: Known to be structured in several patterns at the same time, the prior image of interest is always modeled with the idea of enforcing multiple constraints on unknown signals. For instance, when dealing with a hyperspectral restoration problem, the combination of constraints with piece-wise smoothness and low rank has yielded promising reconstruction results. In this paper, we propose a novel mixed-noise removal method by employing 3D anisotropic total variation and low rank constraints simultaneously for the problem of hyperspectral image (HSI) restoration. The main idea of the proposed method is based on the assumption that the spectra in an HSI lies in the same low rank subspace and both spatial and spectral domains exhibit the property of piecewise smoothness. The low rankness of an HSI is approximately exploited by the nuclear norm, while the spectral-spatial smoothness is explored using 3D anisotropic total variation (3DATV), which is defined as a combination of 2D spatial TV and 1D spectral TV of the HSI cube. Finally, the proposed restoration model is effectively solved by the alternating direction method of multipliers (ADMM). Experimental results of both simulated and real HSI datasets validate the superior performance of the proposed method in terms of quantitative assessment and visual quality.

Keywords: hyperspectral restoration; multiple constraints; low rank; 3D anisotropic total variation; ADMM

1. Introduction

Hyperspectral images (HSIs), three-dimensional data cubes with hundreds of narrow spectral bands ranging from 0.4 to 2.5 μm, have been widely used in various fields [1], e.g., precise agriculture, mineral detection, environment monitoring, and urban planning. However, during acquisition of an image, real HSIs are always corrupted by several kinds of noises, e.g., Gaussian noise, impulse noise, stripes or dead lines, which severely limit the precision of the subsequent processing, e.g., unmixing [2,3], classification [4,5], and anomaly detection [6]. Therefore, it is essential to develop effective restoration techniques for HSIs.

Enormous existing single image denoising methods, e.g., k-means clustering dictionary learning algorithm via singular value decomposition (K-SVD) [7], non-local means [8,9], block-matching and

3D filtering (BM3D) [10] as well as their variants, can be directly applied to HSI by treating it as band-by-band images. However, without imposing the information between bands, these kinds of methods always produce ordinary results.

Several kinds of powerful HSI restoration methods are based on wavelet transform [11–13] which has shown great potential in multi-scale representation [14] and spatial-spectral correlation exploration. For instance, Othman and Qian [15] proposed a two-dimensional (2D) algorithm that benefits the spatial and spectral features of HSIs and operates the wavelet shrinkage in the derivative domain. Usually, a wavelet transform is always combined with principle component analysis (PCA) for HSI restoration [16,17]. For example, block-matching four-dimensional (BM4D) filtering was conducted in the remaining low-energy principal component analysis (PCA) output channel images for HSI denoising [18] and it achieved great progress in removing Gaussian noise. Generally, due to the exploration in spectral and spatial domains, 3D wavelet methods [19] produce much better results than 2D wavelet methods [20]. Moreover, the combination of sparse reduced-rank regression and wavelet transform is another way to produce appealing results [21]. Except for wavelet transform-based methods, there are still a number of techniques developed for hyperspectral restoration, e.g., tensor decomposition methods [22,23], sparse representation [24,25] or sparse dictionary learning methods [26,27], kernel-based methods [28], deep learning [29] or neural network [30,31] methods and Bayesian methods [32,33]. These methods have been proved to produce outstanding results. However, all the above mentioned methods only consider one or two kinds of noises, and fail to remove mixed noise completely.

Recently, low-rank representation (LRR), another powerful technique for image analysis [34,35], has drawn a lot of attention and achieved great success in the field of hyperspectral restoration. Lu et al. [36] proposed an LRR method for removing stripe noise in HSIs by exploiting the high spectral correlation among different bands, and graph regularization was constructed to preserve the intrinsic local structure. Motivated by the idea of robust principle component analysis (RPCA), Zhang et al. [37] exploited an HSI restoration method based on low-rank matrix recovery (LRMR) via so-called "Go Decomposition" algorithm. Even though LRMR was constructed on overlapped patches and achieved exciting performance for denoising mixed Gaussian and sparse noise simultaneously, it only considers the local similarity within patches. To overcome such drawback, by treating the different contribution of the overlapped patches, a spectral nonlocal weighted low-rank representation method was presented in [38]. Meanwhile, He et al. [39] then improved the patch-wise LRMR by a noise-adjusted iteration strategy, which makes LRMR adaptive to the noise intensity in different bands. To exploit both the local similarity within a patch and the nonlocal similarity across the patches in a group simultaneously, Wang et al. [40] proposed an approach of group low-rank representation (GLRR) for HSI denoising. GLRR enables the exploitation of both the local similarity within a patch and the nonlocal similarity across the patches in a group simultaneously and achieves better performance. Different from the above-mentioned low-rank based methods, Sun et al. [41] paved a new way to explore the low rank properties in the spectral difference images of HSI for mixed noise removal. Due to lack of spatial information, however, those low-rank methods only exploit the low-rank properties of HSIs from a spectral perspective and fail to remove the heavy Gaussian noise effectively. In addition, when sparse noise has its own structure, for example, the stripes or dead lines in an HSI are located at the same place in most of the bands, the low-rank-based method will treat the sparse noise as a low-rank component and fail to remove it [42,43]. More research related to low-rank representation for HSI denoising can be found in this literature [44,45] and references therein.

Total variation is an effective tool for imposing the spatial constraint and removing the Gaussian noise in HSI [46–48]. For instance, Yuan et al. [46] extended the vectorial total variation (VTV) [49] for HSI denoising, which is adaptive for noise intensity in each band and structures in the spatial domain. Chang et al. [48] proposed to employ an anisotropic total variation which is defined by maintaining the difference along the direction of strips for destriping of HSI. However, both methods can only remove a certain kind of noise for HSI and are not flexible for removing mixed noises,

i.e., Gaussian noise, stripes and deadlines. In order to deal with mixed noises for HSI, Aggarwal et al. [50] proposed a novel spatial-spectral TV to explore the spectral smoothness across the spatial total variation and achieved great success. In addition, by taking the advantage of combination of multiple constrains, He et al. [42] proposed a band-by-band total variation regularized low-rank matrix factorization (LRTV) method for the restoration of HSIs, which achieved great improvement for mixed noise removal by combining total variation regularization and low rank representation to enforce the local spatial smoothness and the low-rank property of the spectra in the whole scene simultaneously. Similarly, Wu et al. [51] proposed to combine the band-by-band TV regularization with weighted nuclear norm minimization (WNMM) for HSI mixed denoising. However, LRTV or WNMM only consider the band-by-band TV model, and ignore the consistency and smoothness in the spectral domain, which significantly limits the performance of HSI denoising especially in the case of heavy Gaussian noise and structured sparse noise. More recently, Wang et al. [52] combined the spatial-spectral TV with a low rank constraint for HSI mixed denoising and further improved the denoising performance. Another way to improve the performance of HSI mixed denoising is to involve three-dimensional total variation (3DTV) into the multiple combination framework. 3DTV which promotes both spatial and spectral smoothness is a natural extension of the conventional 2D TV for cubic data. It has been utilized for several applications of HSI, such as compressive sensing [53] and super-resolution [54]. In [53], 3DTV is defined as a 2D spatial TV and a 1D spectral TV on the abundance fractions of endmembers for HSI unmixing compressive sensing. Later, the 3DTV with the same definition as that in [53] was combined with low rank tensor prior for HSI super-resolution.

Based on our previous research [55], in this paper, we propose a novel hyperspectral restoration method based on 3D anisotropic total variation (ATV) regularization and low-rank representation, namely 3DATVLR, where the ATV is defined by the l_1 norm TV along two directions in spatial domain and the l_1 norm TV along one direction in spectral domain. This method could exploit both the spatial and spectral information of HSIs and is able to preserve the spectra and edges as well as remove the mixed noise simultaneously by treating an HSI as a 3D cube. Compared to [54], the proposed 3DATV intends to promote different smoothness strength from spatial and spectral TV while 3DTV defined in [53] enforces the spatial TV and spectral TV equally. This enables 3DATV to be more flexible for dealing with different mixed noises. In addition, our 3DATVLR method concerns the mixed noises removal problem while the method in [54] only considers the Gaussian noise for HSI super-resolution problem. More importantly, as analyzed in [56], the unfolding matrices of the HSI tensor along spectral mode and self-similar mode are of low rank. Enforcing low rankness on three modes of HSI tensor will reduce the performance of HSI denoising.

The contributions of this paper are summarized as follows:

1. By treating an HSI as a 3D cube, 3D anisotropic total variation (3DATV) is adopted to exploit the spatial smoothness and spectral consistency simultaneously along with the low-rank regularization.
2. The TV denoising process is updated iteratively in a spatial-spectral manner in the alternating direction method of multipliers (ADMM) algorithm instead of in a band-by-band manner, and it can be effectively solved by an n-D fast Fourier transform (nFFT).
3. When low-rank regularization fails to remove the structured sparse noise, i.e., structured stripes or dead lines, 3DATV could effectively get rid of them by simultaneously exploring the spatial and spectral consistency while LRTV even makes this case in a dilemma since the band-by-band TV preserves the stripes or deadlines along one of the directions.

The remainder of this paper is organized as follows: Section 2 formulates the restoration problem for HSIs. Section 3 describes the proposed 3DATVLR model and presents the details of the optimization algorithm. Section 4 illustrates the experimental results. Section 5 concludes the paper with some remarks.

2. Basic Formulation

2.1. Observation Model

Firstly, we define some notations for HSI restoration problem. Let $\mathbf{X} \in \mathbb{R}^{m \times n \times l}$ represent a clear HSI cube with $m \times n$ spatial pixels and l spectral bands. Then, the observed HSI cube $\mathbf{Y} \in \mathbb{R}^{m \times n \times l}$, corrupted by Gaussian noise $\mathbf{N} \in \mathbb{R}^{m \times n \times l}$ and sparse noise $\mathbf{S} \in \mathbb{R}^{m \times n \times l}$, can be formulated as

$$\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{N}. \tag{1}$$

This mixed noise formulation was first utilized for hyperspectral restoration in [37] using LRMR. It is used to remove the Gaussian noise caused by the imaging sensors and model system, as well as the sparse noise, e.g., dead lines or stripes, created by push-broom sensors.

2.2. Restoration Model

With Equation (1) in mind, the restoration model of removing mixed Gaussian and sparse noises for HSIs can be formulated as

$$\min_{\mathbf{X},\mathbf{S}} J_1(\mathbf{X}) + \lambda J_2(\mathbf{S})$$
$$s.t. \quad ||\mathbf{Y} - \mathbf{X} - \mathbf{S}||_F^2 \leq \varepsilon, \tag{2}$$

where λ is a parameter that maintains the trade-off between the two priori terms, and ε is a parameter related to the density of Gaussian noise. $J_2(\mathbf{S})$ is usually modeled as sparse regularization [37–39], i.e., $||\mathbf{S}||_1 = \sum_i |s_i|$, where s_i is the element of $\mathbf{S}$. $J_1(\mathbf{X})$ is the regularization associated with the clean HSI $\mathbf{X}$, which models the a priori properties of the clean HSI.

2.3. Prior Regularizations

2.3.1. Low Rank Regularization

One choice for $J_1(\mathbf{X})$ is the nuclear norm, i.e., $||\mathbf{X}||_*$, which exploits the low-rank property by enforcing the sum of the singular values of $\mathbf{X}$, and has shown good performance in HSI restoration [37,42].

The objective function involving the low rank regularization and sparse regularization is formulated as:

$$\min_{\mathbf{X},\mathbf{S}} ||\mathbf{X}||_* + \lambda ||\mathbf{S}||_1$$
$$s.t. \quad ||\mathbf{Y} - \mathbf{X} - \mathbf{S}||_F^2 \leq \varepsilon. \tag{3}$$

Equation (3) is the well-known robust principle component analysis (RPCA) model which has been widely used in nature image processing [48]. Even working well for HSI mixed denoising, it still has two main drawbacks as follows:

- Due to the independent distribution of Gaussian noise in each pixel, Equation (3) can not completely remove the heavy Gaussian noise.
- Equation (3) can not remove the structured sparse noise, i.e., stripes or dead lines locate at the same place in each band because the low rank method will treat them as one of the low rank components.

2.3.2. Total Variation Regularization

Another supplementary choice for $J_1(\mathbf{X})$ is the total variation, which can effectively promote the spatial-spectral smoothness by suppressing the heavy Gaussian noise successfully.

Therefore, the band-by-band TV regularized low rank (LRTV) model is expressed as [42]:

$$\min_{\mathbf{X},\mathbf{S}} ||\mathbf{X}||_* + \tau ||\mathbf{X}||_{HTV} + \lambda ||\mathbf{S}||_1$$

$$s.t. \quad ||\mathbf{Y} - \mathbf{X} - \mathbf{S}||_F^2 \leq \varepsilon, \tag{4}$$

where $||\mathbf{X}||_{HTV} = \sum_{i=1}^l ||\mathbf{X}_i||_{tv}$ is the conventional band-by-band TV.

Benefitting from the band-by-band TV regularization, LRTV can produce more accurate results than LRMR, especially for the heavy Gaussian noise case. However, for the structured sparse noise, it still fails to remove it because band-by-band TV will inversely keep the stripes or dead lines in one of the directions, i.e., horizontal direction or vertical direction.

3. 3D Anisotropic TV Regularized Low-Rank Model for HSI Restoration

In this section, we systematically present the mathematical formulation of the proposed method and illustrate the framework of it in Figure 1.
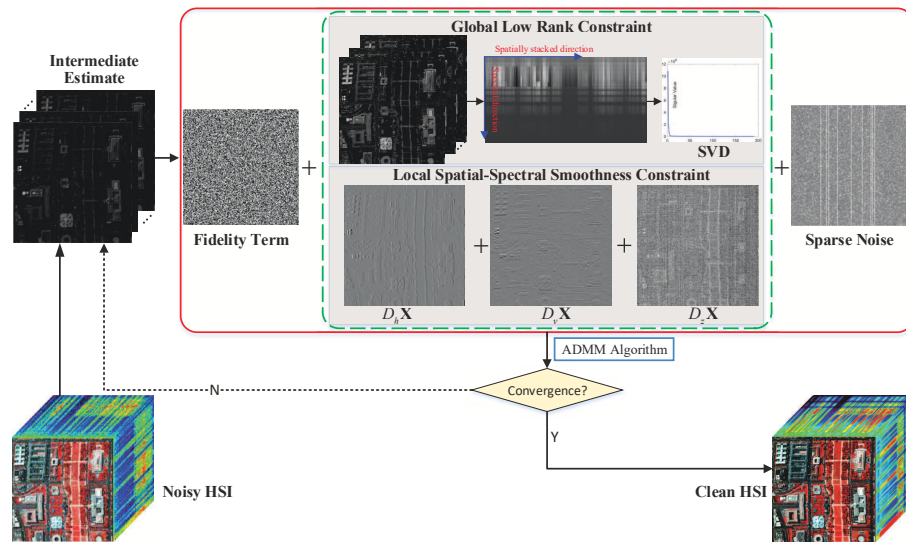

Figure 1. The flow chart of the proposed 3DATVLR method.

3.1. 3D Anisotropic Total Variation Regularization

A straightforward way to extend the conventional 2D TV, which has been widely used in HSI analysis [57,58], to a 3D version for HSI is to impose the spatial-spectral total variation into a unit system. Before doing this, we first analyze the statistical distribution of each gradient image, i.e., $D_h\mathbf{X}$, $D_v\mathbf{X}$, and $D_z\mathbf{X}$. Figure 2 exhibits the visual illustration and statistical table of each gradient image of Washington DC dataset. It is clear that the shape of each spatial difference (i.e., $D_h\mathbf{X}$, $D_v\mathbf{X}$) obeys Laplacian distribution while the shape of the spectral difference obeys Gaussian distribution. The phenomena follows the rule of optical imaging in spectral band. However, considering that the sparse noise will still exist in the spectral difference image, we finally employ the l_1 norm to model the distributions in three directions.
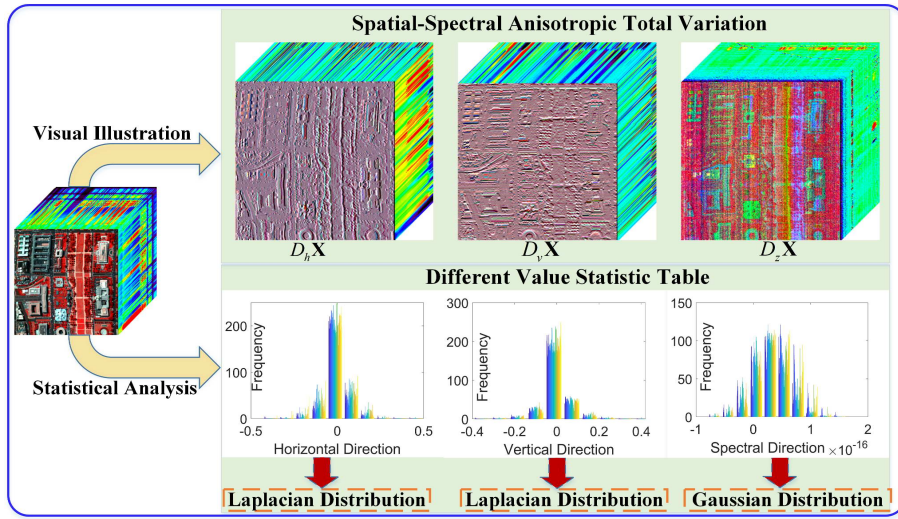
Figure 2. Visual illustration and statistical table of each gradient image of HSI (Washington DC).

Hence, the formulation of 3D anisotropic TV can be expressed as follows:

$$||\mathbf{X}||_{\text{3DATV}} = ||D_h\mathbf{X}||_1 + ||D_v\mathbf{X}||_1 + \rho||D_z\mathbf{X}||_1, \tag{5}$$

where D_h and D_v respectively denote the horizontal and vertical difference operators along the spatial dimension, while D_z is a linear operator corresponding to the spectral difference, and ρ controls the proportion of the spectral correlation. Now, the 3DATV in Equation (5) exploits both spatial and spectral smoothness for HSIs, and the definitions of its three operators are expressed as

$$
\begin{cases}
D_h\mathbf{X}(i,j,k) = & \mathbf{X}(i+1,j,k) - \mathbf{X}(i,j,k), \\
D_v\mathbf{X}(i,j,k) = & \mathbf{X}(i,j+1,k) - \mathbf{X}(i,j,k), \\
D_z\mathbf{X}(i,j,k) = & \mathbf{X}(i,j,k+1) - \mathbf{X}(i,j,k),
\end{cases} \tag{6}
$$

where $\mathbf{X}(i,j,k)$ denotes a pixel at spatial location (i,j) in the kth band. By enforcing the constraints on two spatial dimensions and one spectral dimension, the piecewise smoothness in both domains can be pursued.

3.2. Proposed 3DATVLR Model

By integrating the low rank constraint and 3DATV regularization into a unit framework, we propose a new model for HSI restoration:

$$\min_{\mathbf{X},\mathbf{S}\in\mathbb{R}^{m\times n\times l}} ||\mathbf{X}||_* + \lambda_{tv}||\mathbf{X}||_{\text{3DATV}} + \lambda_s||\mathbf{S}||_1,$$
$$s.t. \quad ||\mathbf{Y} - \mathbf{X} - \mathbf{S}||_F^2 \le \varepsilon, \quad \text{rank}(\mathbf{X}) \le r \tag{7}$$

where λ_{tv} and λ_s are two parameters that represent the trade-off of the three terms, ε represents the noise level of the Gaussian noise, and r is the desired rank of matrix $\mathbf{X}$.

This proposed 3D anisotropic total variation regularized low rank model, termed the 3DATVLR method, can explore the spectral similarity and spatial-spectral smoothness simultaneously. Not only can it remove the heavy Gaussian noise and structured stripes or deadlines by pursuing the spectral-spatial smoothness with 3DATV, but it can also remove the random sparse noise, thus preserving the spectra and keeping the fine spatial details in the original HSI.

3.3. Optimization Algorithm

To effectively solve the optimization Equation (7), we first convert it to the following equivalent constrained formulation:

$$\min_{\mathbf{L},\mathbf{X},\mathbf{S}} ||\mathbf{L}||_* + \lambda_{tv}(||\mathbf{V}_1||_1 + ||\mathbf{V}_2||_1 + \rho||\mathbf{V}_3||_1) + \lambda_s ||\mathbf{S}||_1,$$
$$s.t. \quad ||\mathbf{Y} - \mathbf{L} - \mathbf{S}||_F^2 \leq \varepsilon, \quad \mathbf{L} = \mathbf{X},$$
$$\mathbf{V}_1 = D_h\mathbf{X}, \quad \mathbf{V}_2 = D_v\mathbf{X},$$
$$\mathbf{V}_3 = D_z\mathbf{X}, \quad \text{rank}(\mathbf{L}) \leq r.$$

(8)

Up to this step, Equation (5) has been separated into five simper subproblems by the ADMM method [59], which then solves it by minimizing the following Lagrangian function of Equation (8):

$$\min \mathcal{L}(\mathbf{L}, \mathbf{X}, \mathbf{S}) = \min_{\mathbf{L},\mathbf{X},\mathbf{S}\in\mathbb{R}^{m\times n\times l}} ||\mathbf{L}||_* + \lambda_{tv}(||\mathbf{V}_1||_1 + ||\mathbf{V}_2||_1$$
$$+ \rho||\mathbf{V}_3||_1) + \lambda_s||\mathbf{S}||_1 + \langle \mathbf{\Lambda}_1, \mathbf{Y} - \mathbf{L} - \mathbf{S} \rangle$$
$$+ \langle \mathbf{\Lambda}_2, \mathbf{X} - \mathbf{L} \rangle + \langle \mathbf{\Lambda}_3, \mathbf{V}_1 - D_h\mathbf{X} \rangle$$
$$+ \langle \mathbf{\Lambda}_4, \mathbf{V}_2 - D_v\mathbf{X} \rangle + \langle \mathbf{\Lambda}_5, \mathbf{V}_3 - D_z\mathbf{X} \rangle$$
$$+ \frac{\mu}{2}(||\mathbf{Y} - \mathbf{L} - \mathbf{S}||_F^2 + ||\mathbf{X} - \mathbf{L}||_F^2$$
$$+ ||\mathbf{V}_1 - D_h\mathbf{X}||_F^2 + ||\mathbf{V}_2 - D_v\mathbf{X}||_F^2$$
$$+ ||\mathbf{V}_3 - D_z\mathbf{X}||_F^2)$$
$$s.t. \quad \text{rank}(\mathbf{L}) \leq r,$$

(9)

where μ is the penalty parameter, and $\mathbf{\Lambda}_i, i = 1, 2, ..., 5$ are the Lagrangian multipliers.

Generally, we optimize the Lagrangian Equation (9) iteratively over one variable while fixing the other variables. The detailed pseudo-code for solving the proposed model via ADMM is summarized in Algorithm 1.

Line 4 in Algorithm 1 is related to variable $\mathbf{L}$, which is a nuclear norm subproblem and that can be effectively solved by the singular value shrinkage method [60], given by

$$\mathbf{L}^{(k+1)} = \underset{rank(\mathbf{L})\leq r}{\text{argmin}} \mathcal{L}(\mathbf{L}, \mathbf{X}^{(k)}, \mathbf{S}^{(k)})$$
$$= \underset{rank(\mathbf{L})\leq r}{\text{argmin}} ||\mathbf{L}||_* + \mu||\mathbf{L} - \frac{1}{2}\left(\mathbf{Y} + \mathbf{X}^{(k)} - \mathbf{S}^{(k)}\right.$$
$$\left. + (\mathbf{\Lambda}_1^k + \mathbf{\Lambda}_2^k)/\mu\right)||_F^2$$
$$= \mathcal{D}_{\frac{1}{2\mu}}\left(\frac{1}{2}(\mathbf{Y} + \mathbf{X}^{(k)} - \mathbf{S}^{(k)} + (\mathbf{\Lambda}_1^k + \mathbf{\Lambda}_2^k)/\mu)\right),$$

(10)

where

$$\mathcal{D}_\lambda(\mathbf{W}) = \mathbf{U}\mathcal{D}_\lambda(\Sigma_r)\mathbf{V}^T$$
$$= \mathbf{U}\text{diag}\{\max(\sigma_i - \lambda, 0)\}_{1\leq i\leq r}\mathbf{V}^T.$$

(11)

Here, $\mathbf{U}\Sigma\mathbf{V}^T$ is the singular value decomposition of the given matrix $\mathbf{W}$, and $\{\sigma_i\}_{1\leq i\leq r}$ are the first r singular values of $\mathbf{W}$ in descending order.

Algorithm 1 Extended ADMM for the 3DATVLR method.

1: **Input** HSI matrix $\mathbf{Y} \in \mathbb{R}^{m \times n \times l}$, desired rank r, regularization parameters λ_{tv} and λ_s, maximum

　iteration k_{max}

2: **Initialization** $\mathbf{L}^{(0)} = \mathbf{X}^{(0)} = \mathbf{S}^{(0)} = 0$, $\mathbf{\Lambda}_i^{(0)} = 0, i = 1, .., 5$, $\mathbf{V}_i^{(0)} = 0, i = 1, 2, 3$, $\mu = 0.05$,

　$\mu_{max} = 10^6, \gamma = 1.2, \rho > 0$ and $k = 0$

3: **While** $k < k_{max}$

4:　　$\mathbf{L}^{(k+1)} = \arg\min_{\text{rank}(\mathbf{L}) \leq r} \mathcal{L}(\mathbf{L}, \mathbf{X}^{(k)}, \mathbf{S}^{(k)})$

5:　　$\mathbf{X}^{(k+1)} = \arg\min_{\mathbf{X}} \mathcal{L}(\mathbf{L}^{(k+1)}, \mathbf{X}, \mathbf{S}^{(k)})$

6:　　$\mathbf{S}^{(k+1)} = \text{SoftTH}(\mathbf{Y} - \mathbf{L}^{(k+1)} + \mathbf{\Lambda}_1^{(k)}/^-, \check{\ }_s/^-)$

7:　　$\mathbf{V}_1^{(k+1)} = \text{SoftTH}(D_h\mathbf{X}^{(k+1)} - \mathbf{\Lambda}_3^{(k)}/\mu, \lambda_{tv}/\mu)$

8:　　$\mathbf{V}_2^{(k+1)} = \text{SoftTH}(D_v\mathbf{X}^{(k+1)} - \mathbf{\Lambda}_4^{(k)}/\mu, \lambda_{tv}/\mu)$

9:　　$\mathbf{V}_3^{(k+1)} = \text{SoftTH}(D_z\mathbf{X}^{(k+1)} - \mathbf{\Lambda}_5^{(k)}/\mu, \rho\lambda_{tv}/\mu)$

10:　　**Update Lagrangian multipliers**

11:　　　$\mathbf{\Lambda}_1^{(k+1)} = \mathbf{\Lambda}_1^{(k)} + \mu(\mathbf{Y} - \mathbf{L}^{(k+1)} - \mathbf{S}^{(k+1)})$

12:　　　$\mathbf{\Lambda}_2^{(k+1)} = \mathbf{\Lambda}_2^{(k)} + \mu(\mathbf{X}^{(k+1)} - \mathbf{L}^{(k+1)})$

13:　　　$\mathbf{\Lambda}_3^{(k+1)} = \mathbf{\Lambda}_3^{(k)} + \mu(\mathbf{V}_1^{(k+1)} - D_h\mathbf{X}^{(k+1)})$

14:　　　$\mathbf{\Lambda}_4^{(k+1)} = \mathbf{\Lambda}_4^{(k)} + \mu(\mathbf{V}_2^{(k+1)} - D_v\mathbf{X}^{(k+1)})$

15:　　　$\mathbf{\Lambda}_5^{(k+1)} = \mathbf{\Lambda}_5^{(k)} + \mu(\mathbf{V}_3^{(k+1)} - D_z\mathbf{X}^{(k+1)})$

16:　　**Update the parameter** $\mu = \min(\gamma\mu, \mu_{max})$

17:　　**Update iteration** $k = k + 1$

18: **End While**

Line 5 in Algorithm 1 is for solving a convex subproblem related to the variable $\mathbf{X}$, as follows:

$$
\begin{aligned}
\mathbf{X}^{(k+1)} &= \underset{\mathbf{X}}{\text{argmin}}\, \mathcal{L}(\mathbf{L}^{(k+1)}, \mathbf{X}, \mathbf{S}^{(k)}) \\
&= \underset{\mathbf{X}}{\text{argmin}} \left(||\mathbf{X} - \mathbf{L}^{(k+1)} + \frac{\mathbf{\Lambda}_2^{(k)}}{\mu}||_F^2 + ||\mathbf{V}_1^{(k)} - D_h\mathbf{X} \right. \\
&\quad + \frac{\mathbf{\Lambda}_3^{(k)}}{\mu}||_F^2 + ||\mathbf{V}_2^{(k)} - D_v\mathbf{X} + \frac{\mathbf{\Lambda}_4^{(k)}}{\mu}||_F^2 \\
&\quad \left. + ||\mathbf{V}_3^{(k)} - D_z\mathbf{X} + \frac{\mathbf{\Lambda}_5^{(k)}}{\mu}||_F^2 \right).
\end{aligned}
\tag{12}
$$

The solution to Equation (12) is equivalent to the following liner system of equations:

$$
(\mathbf{I} + \Delta)\mathbf{X}^{(k+1)} = \mathbf{L}^{(k+1)} - \frac{\mathbf{\Lambda}_2^{(k)}}{\mu} + \xi_1 + \xi_2 + \xi_3,
\tag{13}
$$

where

$$
\Delta = D_h^T D_h + D_v^T D_v^T + D_z^T D_z, \quad \xi_1 = D_h^T\left(\mathbf{V}_1^{(k)} + \frac{\mathbf{\Lambda}_3^{(k)}}{\mu}\right),
$$

$$\xi_2 = D_v^T \left(\mathbf{V}_2^{(k)} + \frac{\Lambda_4^{(k)}}{\mu} \right), \quad \xi_3 = D_z^T \left(\mathbf{V}_3^{(k)} + \frac{\Lambda_5^{(k)}}{\mu} \right),$$

and T is the matrix transpose. By treating $D_h\mathbf{X}$, $D_v\mathbf{X}$, and $D_z\mathbf{X}$ as convolutions along two spatial directions and one spectral direction, this problem has a closed form solution via the nFFT.

$$\mathbf{X}^{(k+1)} = \mathcal{F}^{-1} \left(\frac{\mathcal{F} \left(\mathbf{L}^{(k+1)} - \frac{\Lambda_2^{(k)}}{\mu} + \xi_1 + \xi_2 + \xi_3 \right)}{1 + \Sigma_{i \in \{h,v,z\}} \mathcal{F}(D_i)^H \mathcal{F}(D_i)} \right), \tag{14}$$

where $\mathcal{F}$ is the fast nFFT operator, $\mathcal{F}^{-1}$ is the inverse nFFT operator, and H is the complex conjugate.

For the rest of the variables, i.e., $\mathbf{S}$, $\mathbf{V}_i, i = 1, 2, 3$, the related problems can be solved easily by a soft threshold operator, i.e., $\text{SoftTH}(x, \lambda) = \text{sign}(x) \times \max\{0, |x| - \lambda/2\}$. More details can be found in [42,48].

The convergence of Algorithm 1 using ADMM is guaranteed theoretically in [59]. Moreover, the convergence rate of the ADMM algorithm depends on the parameter μ, which is first initialized as $\mu = 0.05$ and then updated automatically in the algorithm according to the formulate $\mu = min(\gamma\mu, \mu_{max}$ in each iteration, here $\gamma = 1.05$ and $\mu_{max} = 10^6$. This strategy has been proved to be an effective method for setting the value of μ for practical applications [42,61].

3.4. Parameters Determination

As presented in Algorithm 1, there are four parameters in total, λ_{tv}, λ_s, the desired rank r, and the spectral proportion ρ in the proposed method. λ_s is related to the density of sparse noise, which is set as $\lambda_s = 10/\sqrt{(m \times n)}$ in all the experiments. What we need to discuss is the desired rank r, the 3DATV regularized parameter λ_{tv}, and the spectral proportion ρ.

According to the linear mixture model of hyperspectral image, the desired rank r represents the number of endmembers in the scene, and it can be estimated by the HSI subspace identification method termed HySime [62]. However, due to the complexity of mixed noise in the scene, we still need to slightly tune it accordingly.

The regularized parameter λ_{tv} controls the trade-off between the nuclear norm and the 3DATV norm. As analyzed in [42], the role of the 3DATV regularization in (7) gets smaller as the iteration progresses. Thus, it is vital to determine the initial value of λ_{tv}.

The parameter ρ controls the proportion of spectral TV in the 3DATV term. When it is set to zero, the contribution of spectral smoothness is omitted and the proposed model degrades to LRTV model. Generally, ρ is set in the range of [0–1] for most of the noise cases, and a larger value of ρ is recommended for heavy noise case, and vice versa.

4. Experimental Results and Discussion

To validate the effectiveness of the proposed method, the simulated Indian Pines and real Urban datasets are employed for the experiments. It is worth mentioning that the gray values of each band of the HSI are normalized to the range of [0–1], and, after the restoration, they are stretched back to the original range.

4.1. Comparison Methods and Assessment Metrics

To thoroughly evaluate the performance of the proposed 3DATVLR method for HSI restoration, the block matching 4-d filtering (BM4D) [17], the principal component analysis based BM4D (PCABM4D) [18], parallel factor analysis (PARAFAC) [63], the decomposable nonlocal tensor dictionary learning (TDL) [64], the GoDec based low-rank matrix recovery (LRMR) [37], 3DTV in Equation (3), and the band-by-band total variation regularized low rank matrix factorization (LRTV) [42] are employed as the benchmark methods. BM4D and PCABM4D are two representative

tensor based methods by exploiting nonlocal similarities in wavelet domain. PARAFAC is a powerful rank-1 multilinear algebra model and TDL is a decomposable nonlocal tensor dictionary learning method. Both PARAFAC and TDL are proposed for Gaussian noise removal. LRMR is one of the outstanding methods with low rank property for HSI restoration, 3DTV is a 3D spatial-spectral TV based mixed noise removal method without low rank constraint, while LRTV is the most relevant method to the proposed one, which employs the low rank constraint with band-by-band TV regularization.

As in [37,46], the mean peak signal-to-noise ratio (MPSNR) index, the mean structural similarity (MSSIM) index [65], the mean feature similarity (MFSIM) index, erreur relative globale adimensionelle de synthese (ERGAS) [66] and the mean spectral angle (MSA) are employed to give a quantitative assessment of the HSI denoising results. It is worth noting that the larger values of MPSNR, MSSIM, and MFSIM represent the better performance while the smaller values of ERGAS and MSA represent the better performance.

4.2. Experiments on Simulated Datasets

4.2.1. Datasets and Experimental Settings

Following the experiment in [42,67], the simulated data were generated using the ground truth of the Indian Pines data set (https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html) and the spectral signatures were extracted from the USGS digital library (http://speclab.cr.usgs.gov/spectral.lib). The pixels of the same class are replaced by a specific spectrum from the library. All of the unlabeled pixels were also replaced by the same signature. The final scene has 17 pure classes with 17 endmembers from the USGS library and the size of the scene is $145 \times 145 \times 224$. The simulated dataset, illustrated in Figure 3, is available online (https://sites.google.com/site/rshewei/home).
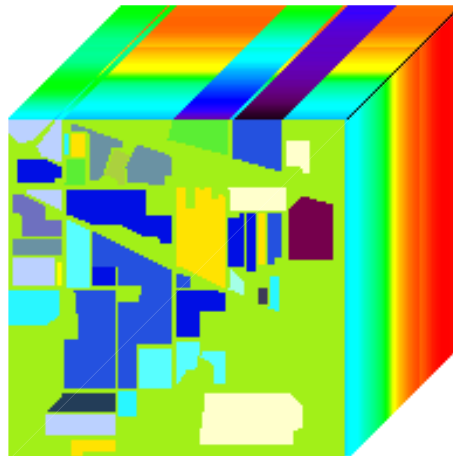

Figure 3. Simulated Indian Pines dataset used in the experiment (false color image composed of bands 6, 88, 220 for the red, green and blue wavelength, respectively).

After generating the clean dataset, five cases of noises are added to simulate the real HSI dataset according to the following rules:

(1) Case 1: Zero-mean Gaussian noise with the same variance of 0.01 was added to each band.

(2) Case 2: Zero-mean Gaussian noise with the same variance of 0.01, as well as the impulse noise with the same density of 15%, was added to each band.

(3) Case 3: In this case, the Gaussian noise and impulse noise were added just like that in Case 2. In addition, the deadlines were added from band 111 to band 150 with the number of deadlines being randomly selected from 3 to 10, and the width of each deadline was randomly generated from 1 to 3.

(4) Case 4: In this case, the noise intensity was different for each band. First, zero-mean Gaussian noise with different variances varying from 0 to 0.02 was added to each band randomly. Second, impulse noise with a density percentage varying from 0 to 20% was randomly added to each band. Finally, dead lines were simulated the same as that in Case 3.

(5) Case 5: In this case, the Gaussian noise, impulse noise and dead lines are simulated the same as that in Case 4. In addition, some periodical stripes are added from band 146 to 165; the number of stripes in each band is 30.

(6) Case 6: In this case, the Gaussian noise, impulse noise and stripes are simulated the same as that in Case 5. However, the dead lines are simulated to be located in the same place as the 40 bands randomly selected from all of the bands; the number of dead lines in each band is 15. This case is to simulate the structured sparse noise for the most severe noise situation.

For the simulated Indian Pines datasets with different noise levels, the parameters for the competing algorithms are set as follows: For the BM4D and PARAFAC methods, both of them are parameter free. For the PCABM4D method, the only parameter is the number of first principal components k; it is generally chosen from the set $\{0, 1, 2, 3\}$. For the TDL method, the only parameter is the deviation of the noise which is set to 0.1. For the 3DTV method, we set $\lambda_{tv} = 0.01$ and $\rho = 10$. For the LRMR method, we set window size $q = 20$, rank $r = 5$, and the cardinality of sparsity $k = 4000$. For the LRTV method, we set the parameter related to the band-by-band TV as $\tau = 0.005$, and rank $r = 5$ according to the corresponding references. For the proposed 3DATVLR method, parameter λ_{tv} is set to 0.005 while desired rank r is set to 6 to obtain optimal results. It is worth noting that the parameters for the competing methods are tuned slightly for different noise intensities to obtain optimal results, and the detailed description of the parameters are provided in Table 1.

Table 1. The parameters setting of the competing methods in the simulated experiments.

Noise	BM4D	PCABM4D	PARAFAC	3DTV	TDL	LRMR	LRTV	3DATVLR
Case 1	-	$k = 0$	-	$\lambda_{tv} = 0.01$ $\lambda_s = 0.05$ $\rho = 11$	$\sigma = 0.016$	rank(L) = 5 card(s) = 0.0	$\tau = 0.015$ $r = 10$	$\lambda_{tv} = 0.009$ $\rho = 0.3$ $r = 10$
Case 2	-	$k = 0$	-	$\lambda_{tv} = 0.01$ $\lambda_s = 0.05$ $\rho = 11$	$\sigma = 0.01$	rank(L) = 5 card(s) = 0.04	$\tau = 0.01$ $r = 9$	$\lambda_{tv} = 0.009$ $\rho = 0.3$ $r = 10$
Case 3	-	$k = 0$	-	$\lambda_{tv} = 0.01$ $\lambda_s = 0.05$ $\rho = 11$	$\sigma = 0.01$	rank(L) = 5 card(s) = 0.07	$\tau = 0.005$ $r = 10$	$\lambda_{tv} = 0.009$ $\rho = 0.5$ $r = 10$
Case 4	-	$k = 0$	-	$\lambda_{tv} = 0.01$ $\lambda_s = 0.05$ $\rho = 11$	$\sigma = 0.012$	rank(L) = 5 card(s) = 0.07	$\tau = 0.006$ $r = 10$	$\lambda_{tv} = 0.01$ $\rho = 0.5$ $r = 10$
Case 5	-	$k = 0$	-	$\lambda_{tv} = 0.01$ $\lambda_s = 0.05$ $\rho = 10$	$\sigma = 0.009$	rank(L) = 5 card(s) = 0.06	$\tau = 0.007$ $r = 8$	$\lambda_{tv} = 0.01$ $\rho = 0.5$ $r = 10$
Case 6	-	$k = 0$	-	$\lambda_{tv} = 0.01$ $\lambda_s = 0.05$ $\rho = 10$	$\sigma = 0.014$	rank(L) = 5 card(s) = 0.05	$\tau = 0.005$ $r = 10$	$\lambda_{tv} = 0.014$ $\rho = 5$ $r = 10$

4.2.2. Visual Performance Evaluation

To thoroughly show the performance of the competing methods, the results in Case 4, Case 5 and Case 6 are chosen for illustration. Figure 4 illustrates the false-color composite and corresponding zoomed-in portions of the denoising results with noise of Case 4. Figure 5 shows the mean spectra of the difference between the original spectra and the constructed ones in class 4 of the simulated Indian Pines dataset. From them, it is obvious that 3DATVLR produces the best visual quality among those competing methods, that is, 3DATVLR reconstructs the closest result to the original one in Figure 4j and plots the smoothest mean difference spectrum in Figure 5i. BM4D, PCABM4D, PARAFAC and TDL methods all remove the Gaussian noise well; however, four of them can not remove the sparse noise completely and produce some distortions around the sparse noise. LRMR performs the worst for preserving the spatial details among the four mixed denoising methods due to its poor ability to remove heavy Gaussian noise, as observed in the zoomed-in portion of Figure 4h. 3DTV can remove those mixed noises effectively, but the fine details of the images are somewhat oversmoothed, as observed in Figure 4f. The false color image shown in the zoomed-in portion of Figure 4i and the mean difference spectrum shown in Figure 5h both show that even LRTV can remove the Gaussian noise and impulse noise effectively, but it fails to preserve the spectra and spatial fine details; see the ellipse and rectangle zones in Figure 4i.
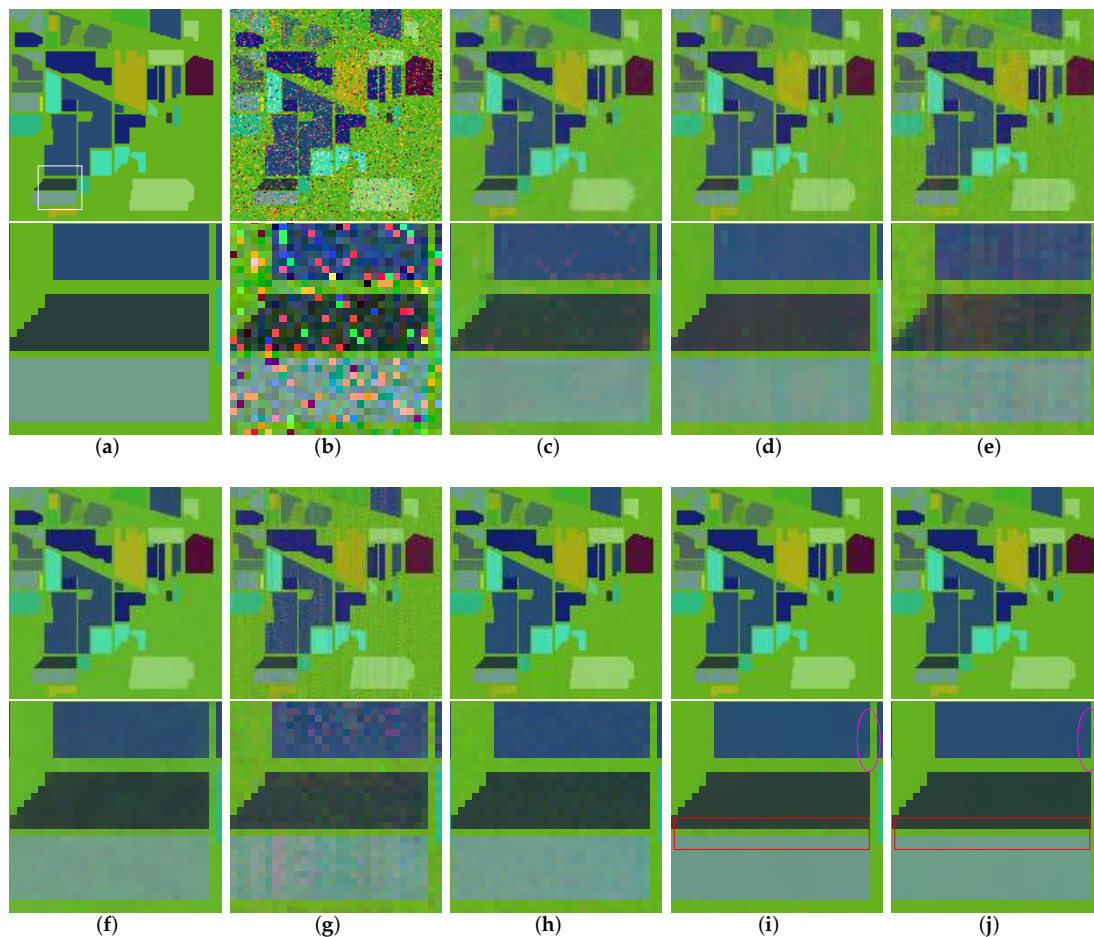

Figure 4. False color image composed of bands 6, 88, 220 (top row) and its corresponding zoomed-in portion (bottom row) marked with white rectangle in the clean image of the simulated Indian Pines dataset with noise Case 4. (**a**) clean image; (**b**) noisy image (PSNR = 15.67 dB); (**c**) BM4D (PSNR = 30.01 dB); (**d**) PCABM4D (PSNR = 30.40 dB); (**e**) PARAFAC (PSNR = dB); (**f**) 3DTV (PSNR = 32.44 dB); (**g**) TDL (PSNR = 30.40 dB); (**h**) LRMR (PSNR = 35.63 dB); (**i**) LRTV (PSNR = 38.06 dB); (**j**) 3DATVLR (PSNR = 39.65 dB).
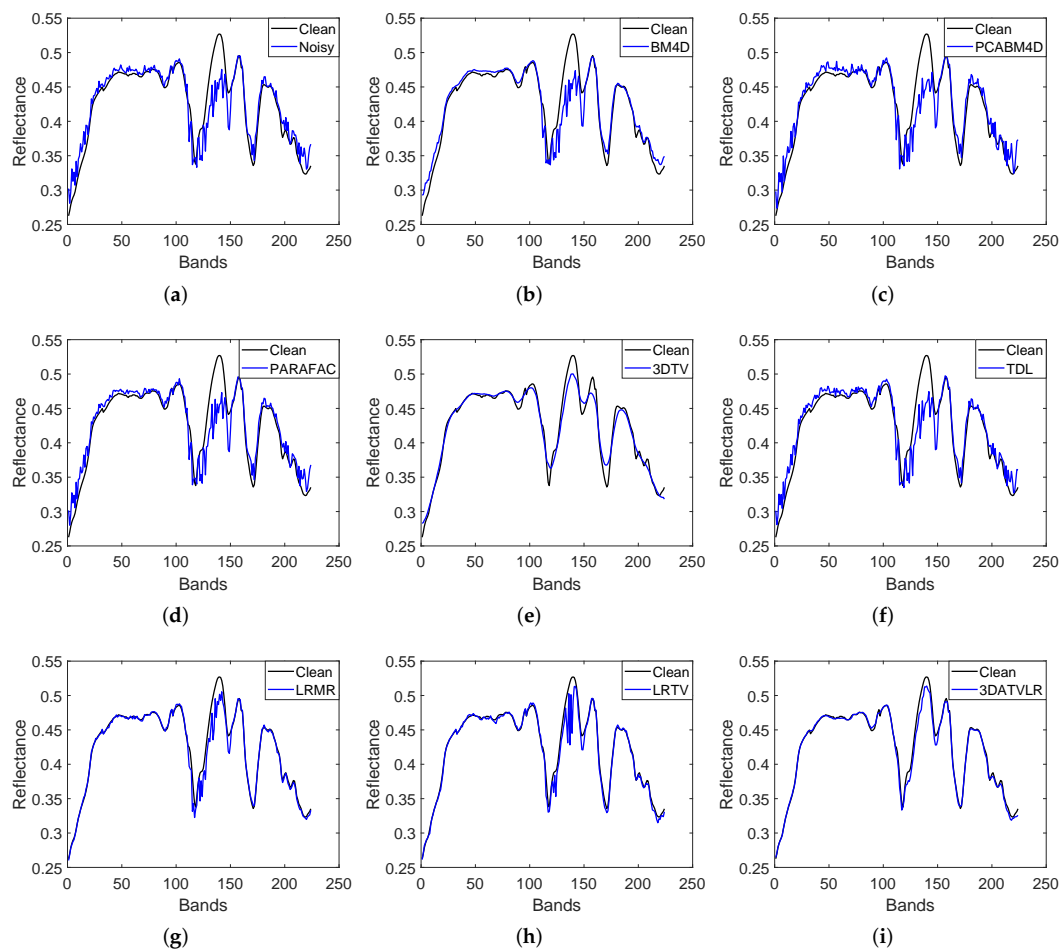
Figure 5. Mean spectra of the original spectra and the reconstructed spectra in class 2 of the simulated Indian Pines dataset with Case 4. (**a**) Noisy. (**b**) BM4D; (**c**) PCABM4D; (**d**) PARAFAC; (**e**) 3DTV; (**f**) TDL; (**g**) LRMR; (**h**) LRTV; (**i**) 3DATVLR.

Figures 6 and 7 illustrate the denoising results of band 123 in Case 5 and band 166 in Case 6 of the simulated dataset, respectively. Both of them are heavily corrupted by Gaussian noise, impulse noise, and dead lines. The difference is that the dead lines in Case 6 are located at the same place of the 40 randomly selected bands. From them, it easily leads us to conclude that the proposed 3DATVLR method produces the best performance in both cases. For BM4D, PCABM4D, PARAFAC and TDL methods, they all fail to remove the impulse noise and dead lines. For the LRMR method, if the dead lines are not structured, it can easily get rid of them completely; however, it is still not good at removing the heavy Gaussian noise and the structured dead lines. A similar performance happens for the LRTV method; it can only improve to remove the heavy Gaussian noise while failing to get rid of the structured dead lines.
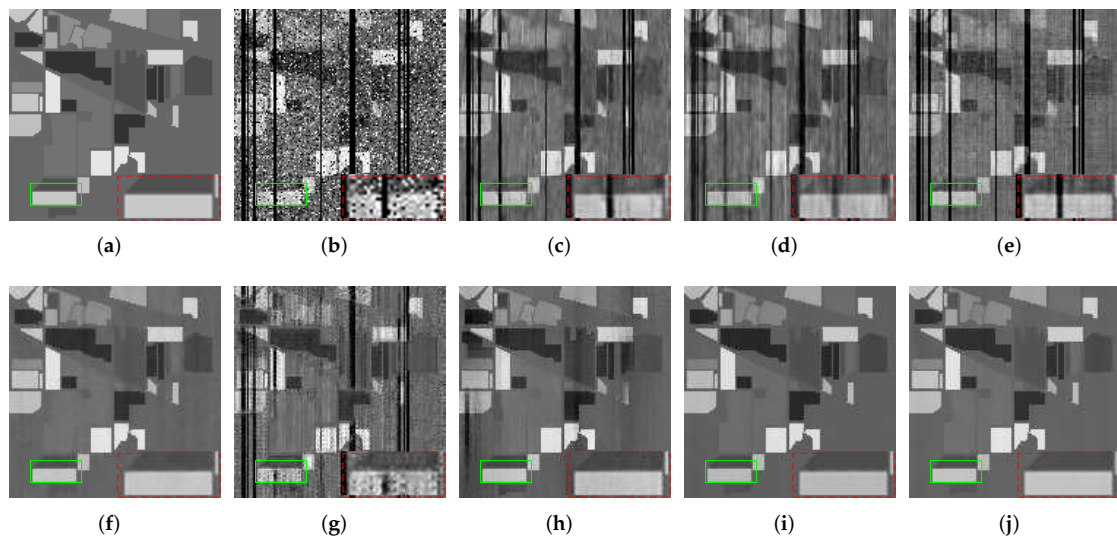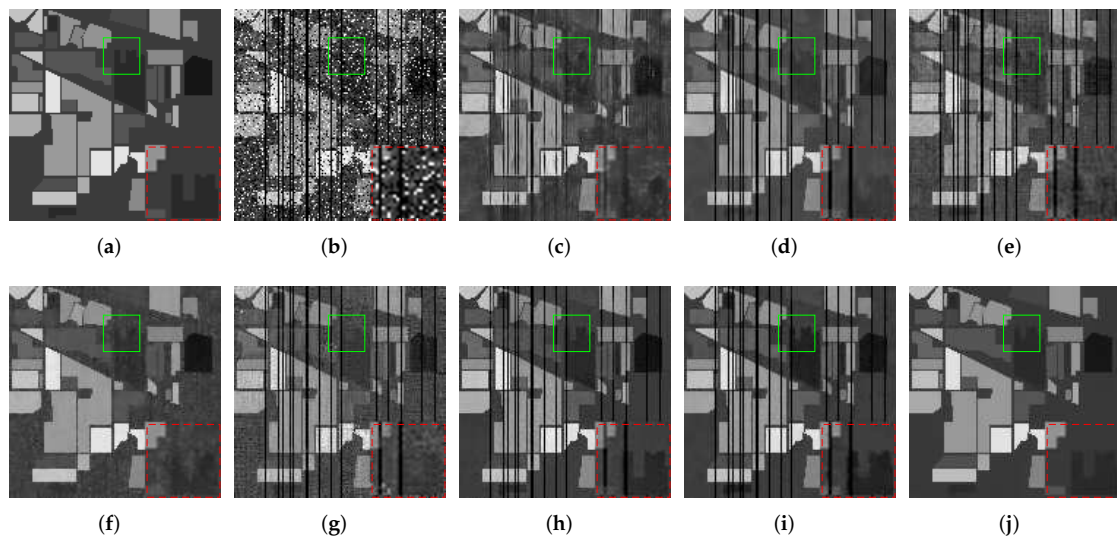
Figure 6. Restored results of band 123 in Simulated Indian Pines dataset with noise of Case 5. (**a**) original; (**b**) noisy; (**c**) BM4D; (**d**) PCABM4D; (**e**) PARAFAC; (**f**) 3DTV; (**g**) TDL; (**h**) LRMR; (**i**) LRTV; (**j**) 3DATVLR.

Figure 7. Restored results of band 166 in Simulated Indian Pines dataset with noise of Case 6. (**a**) original; (**b**) noisy; (**c**) BM4D; (**d**) PCABM4D; (**e**) PARAFAC; (**f**) 3DTV; (**g**) TDL; (**h**) LRMR; (**i**) LRTV; (**j**) 3DATVLR.

4.2.3. Quantitative Performance Evaluation

Table 2 exhibits the quantitative values in terms of MSPNR, MSSIM, MFSIM, ERGAS and MSA for the eight denoising methods above with different noise intensities. From Table 2, it can be observed that the proposed 3DATVLR method exhibits overwhelming superiority in terms of all assessment indexes among all the compared methods in each noise intensity. In addition, the MSPNR, MSSIM, MFSIM, ERGAS and MSA values have great consistency with the visual evaluations. From Table 2, we can also find a strange phenomenon that, when there are stripe noise or dead lines, the LRTV method works even worse than LRMR does in preserving the spectra; see the MSA values in Cases 4, 5 and 6.

Table 2. Metrics values of the denoising results in simulated Indian Pines dataset (The best results are in bold type).

Noise	Metrics	Noisy	BM4D	PCABM4D	PARAFAC	3DTV	TDL	LRMR	LRTV	3DATVLR
Case 1	MPSNR (dB)	20.12	37.91	**41.22**	31.92	33.32	37.24	37.91	38.24	40.32
	MSSIM	0.3710	0.9695	0.9876	0.8946	0.9395	0.9758	0.9695	0.9884	**0.9909**
	MFSIM	0.4724	0.9617	0.9860	0.8655	0.9344	0.9635	0.9617	0.9863	**0.9885**
	ERGAS	230.15	30.72	**21.00**	61.97	69.23	33.17	30.72	29.93	23.91
	MSA	0.1949	0.0234	**0.0158**	0.0391	0.0497	0.0201	0.0234	0.0174	0.0161
Case 2	MPSNR (dB)	12.78	28.50	28.36	25.70	32.14	26.88	35.12	38.13	**39.24**
	MSSIM	0.1682	0.8534	0.9287	0.6819	0.9183	0.9030	0.9168	0.9844	**0.9894**
	MFSIM	0.3397	0.8535	0.9372	0.7047	0.9140	0.8876	0.9156	0.9774	**0.9868**
	ERGAS	538.15	91.90	93.51	125.09	78.58	110.59	42.00	30.57	**26.79**
	MSA	0.4268	0.0673	0.0682	0.0937	0.0568	0.0805	0.0282	0.0198	**0.0179**
Case 3	MPSNR (dB)	12.63	26.93	26.83	24.48	32.03	26.18	34.62	37.30	**38.50**
	MSSIM	0.1656	0.8100	0.8437	0.6698	0.9259	0.8390	0.9112	0.9781	**0.9892**
	MFSIM	0.3372	0.8269	0.8653	0.7015	0.9236	0.8483	0.9099	0.9744	**0.9866**
	ERGAS	547.23	128.70	116.59	161.51	80.27	122.68	44.59	54.83	**29.29**
	MSA	0.4364	0.0997	0.0884	0.1289	0.0574	0.0893	0.0318	0.0430	**0.0197**
Case 4	MPSNR (dB)	14.26	27.62	27.27	25.62	33.16	24.46	34.06	36.09	**39.02**
	MSSIM	0.2236	0.7976	0.8169	0.7171	0.9373	0.6328	0.9249	0.9724	**0.9924**
	MFSIM	0.3920	0.8207	0.8253	0.7304	0.9354	0.6870	0.9211	0.9737	**0.9915**
	ERGAS	486.26	174.99	154.37	189.63	72.96	189.56	67.05	74.54	**29.46**
	MSA	0.3969	0.1422	0.1237	0.1597	0.0515	0.1518	0.0478	0.0578	**0.0202**
Case 5	MPSNR (dB)	13.94	26.92	26.86	24.87	33.19	25.24	34.76	36.40	**39.19**
	MSSIM	0.2123	0.7635	0.7946	0.6801	0.9339	0.6845	0.9323	0.9829	**0.9921**
	MFSIM	0.3802	0.8007	0.8098	0.7113	0.9334	0.7392	0.9285	0.9843	**0.9911**
	ERGAS	493.39	172.68	146.17	189.80	71.39	169.86	54.56	55.00	**28.21**
	MSA	0.4028	0.1410	0.1160	0.1594	0.0509	0.1316	0.0392	0.0398	**0.0192**
Case 6	MPSNR (dB)	14.28	26.78	27.59	25.14	33.32	25.18	32.10	33.86	**36.73**
	MSSIM	0.2197	0.7328	0.8338	0.6559	0.9333	0.6592	0.8526	0.8999	**0.9817**
	MFSIM	0.3912	0.7908	0.8793	0.7109	0.9327	0.7320	0.8902	0.9427	**0.9841**
	ERGAS	481.82	161.46	169.65	191.34	72.67	189.90	154.02	156.26	**53.69**
	MSA	0.3903	0.1069	0.0988	0.1281	0.0511	0.1245	0.0682	0.0751	**0.0370**
	runtime (s)	-	194.1	195.0	353.0	76.0	11.3	28.0	79.7	40.8

Figure 8 presents the PSNR and SSIM values in each band of the reconstructed results with different restoration algorithms in all cases. It leads us to conclude that the PSNR and SSIM values of almost all of the bands obtained by the proposed 3DATVLR are higher than those of the other seven competing algorithms, which once again validates the effectiveness of the proposed methods at removing the mixed noise of HSI. In addition, it is easy to see that the LRTV method always fails to remove one or two bands of the HSI when there are stripes or dead lines. This may be the reason why LRTV produces the worse results in preserving the spectra of the data when there are stripes noise or dead lines.

4.2.4. Parameters' Sensitive Analysis

Figure 9 presents the relationship between the MPSNR, MSSIM and MSA values with the regularization parameters, λ_{tv} and desired rank r. It shows that λ_{tv} and r really have an impact on the performance. The reason for this is that rank r has an inherent relationship with the number of endmembers in the scene, while λ_{tv} determines the strength of 3DATV regularization. Even so, the proposed method is relatively robust and stable to these two parameters and it can produce competitive results when λ_{tv} and r lie in the range of [0.003–0.01] and [4–8], respectively.
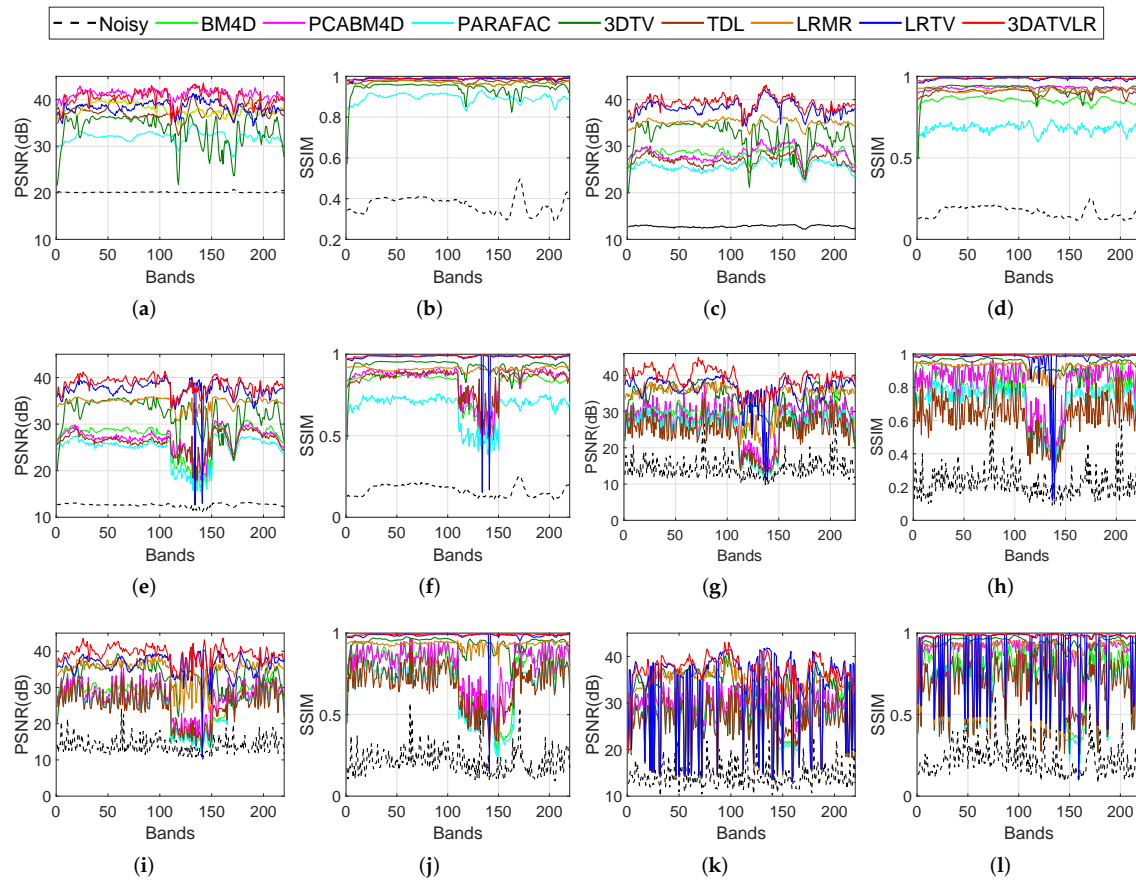
Figure 8. PSNR and SSIM values in each band of the reconstructed simulated Indian Pines dataset with different algorithms. (**a**,**b**) Case 1; (**c**,**d**) Case 2; (**e**,**f**) Case 3; (**g**,**h**) Case 4; (**i**,**j**) Case 5; (**k**,**l**) Case 6.
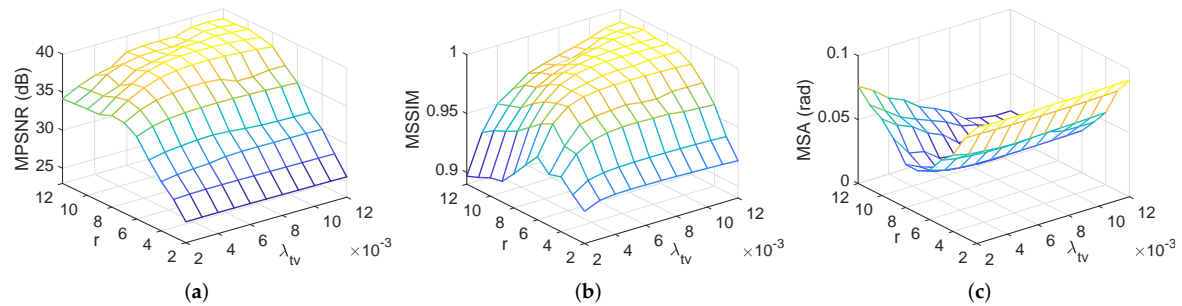
Figure 9. MPSNR, MSSIM and MSA values as a function of parameter λ_{tv} and desired rank r. (**a**) MPSNR versus λ_{tv} and r; (**b**) MSSIM versus λ_{tv} and r; (**c**) MSA versus λ_{tv} and r.

Figure 10 plots the MPSNR values as a function of the parameter ρ. From it, we can see that ρ really has an impact on the reconstruction results. This also means that the spectral difference in the 3DATV regularization contributes to improving the performance of the proposed method. In addition, we can see that if there is only Gaussian noise and impulse noise in the simulated dataset, the value of ρ lies in the range of [0–1] to obtain optimal results. If there are stripe noise or dead lines in the dataset, it needs a greater value of ρ, e.g., $\rho = 5$, to significantly improve the performance.
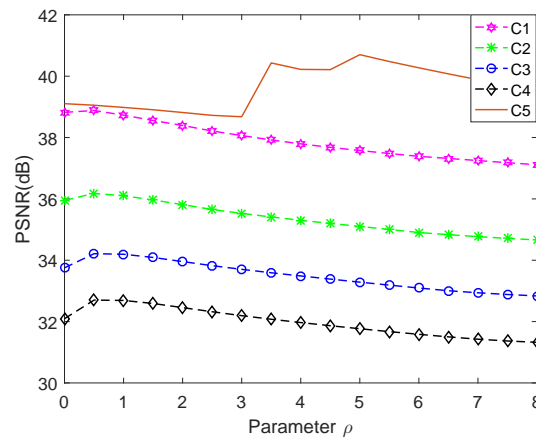
Figure 10. MPSNR values as a function of parameter ρ while fixing the other variables in different noise levels for a simulated Indian Pines dataset.

Figure 11 plots the convergence curves (PSNR as a function of the iteration and SSIM as a function of the iteration) of the proposed method. It is quite clear that the algorithm will converge sharply as the incensement of iteration. It experimentally validates that the proposed 3DATVLR method has a fast convergence rate.
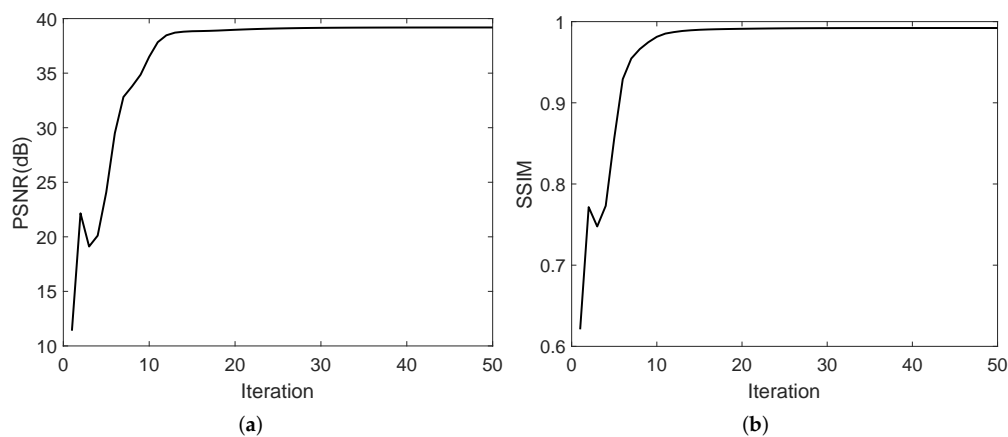

(**a**)

(**b**)

Figure 11. MPSNR, MSSIM values as a function of the iterations for the proposed method in Simulated Indian Pines. (**a**) MPSNR versus the iterations; (**b**) MSSIM versus the iterations.

In summary, the experimental results of the simulated Indian Pines dataset all indicate that the proposed 3DATVLR outperforms the other three state-of-the-art restoration methods in terms of quantitative assessment and visual effect. In addition, for the structured sparse noise, only the proposed 3DATVLR method can get rid of it completely.

4.3. Experiments on a Real Dataset

The Hyperspectral Digital Imagery Collection Experiment (HYDICE) Urban dataset (http://www.tec.army.mil/Hypercube/) is adopted for real data experiments. The size of the original scene is $307 \times 307 \times 210$. Even though bands 104–108, 139–151 and 207–210 are heavily polluted by the atmosphere and water absorption, we still maintain them to thoroughly validate the performance of the proposed method in removing the severe mixed noise. Therefore, the real scene employed in the experiment is of the size $307 \times 307 \times 210$. An image of band 123 is illustrated in Figure 12a. From it, we can find that there are heavy Gaussian, strip noise and dead lines as well as the uneven illumination in the scene. In this experiment, a new state-of-the-art denoising method, i.e., noise-adjusted iterative low-rank matrix approximation (NAILRMA) [39] is added for comparison.

Figure 12 illustrates the image of band 139 and its corresponding zoomed-in portion of the denoising results of the real Urban dataset. It clearly shows that our proposed method produces the best results in visual quality. From the zoomed-in portion of Figure 12e, it is obvious that 3DTV removes the stripe and Gaussian noise effectively, but there is some over-smoothness in the restored image. From Figure 12g–i, we can see that LRMR, NAILRMA, and LRTV all fail to remove the stripe noise. This is because the stripes in Urban dataset are all located at the same place from band 108 to band 140. Moreover, a small amount of Gaussian noise still exists in the image restored by LRMR and NAILRMA, while the proposed 3DATVLR method promotes proper smoothness and preserves the fine details better.
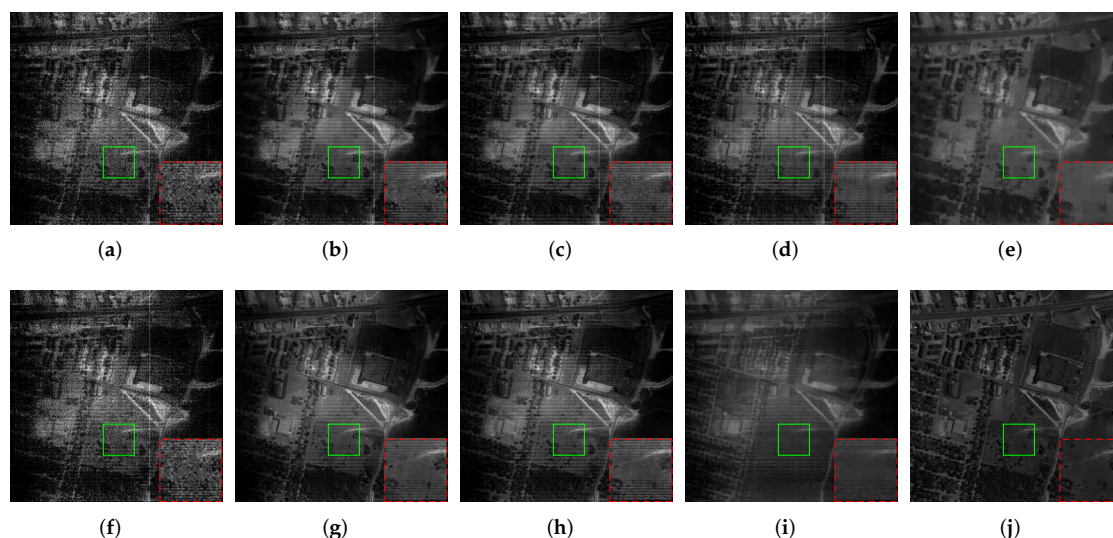
Figure 12. Restored results of band 139 in Urban dataset. (**a**) Observed; (**b**) BM4D; (**c**) PCABM4D; (**d**) PARAFAC; (**e**) 3DTV; (**f**) TDL; (**g**) LRMR; (**h**) NAIRLMA; (**i**) LRTV; (**j**) 3DATVLR.

Figure 13 presents the image of band 151 and its corresponding zoomed-in protion of the denoising results of the real Urban dataset. From it, we can draw a similar conclusion as that from Figure 12. That is, BM4D, PCABM4D, PARAFAC and DTL methods can remove the heavy Gaussian noise to an extent, but all of them fail to remove the stripes and dead lines. All of those LR based methods, i.e., LRMR, NAILRMA and LRTV, fail to remove the structured stripes in the scene. Benefitting from the 3DATV regularization, the proposed 3DATVLR is successful in getting rid of the mixed noise completely and preserving the spatial fine details well.

Figure 14 plots the horizontal mean profile of band 151 before and after restoration. Due to the mixed Gaussian and stripe noise, there are rapid fluctuations in the original curve. All eight competing methods can essentially suppress the mixed noise more or less, and our method appears to produce the best performance since it shows the smoothest and closest regression to the original curve.

To assess the spatial information restoration, Q-metric [68], which is a blind image content measurement index, is computed by averaging the Q values (aQ) in all bands to evaluate the results. From the above experiments, due to the poor performance of the BM4D, PCABM4D, PARAFA and TDL methods, we do not calculate the aQ values of them for comparison. The results in Table 3 demonstrate that our method outperforms other state-of-the-art methods in restoring the spatial details. The computation times (on a laptop with a 2.5 GHz CPU and 8 GB of RAM, running MATLAB 2016a) of the competing methods on the Urban dataset are also presented in Table 3, from which we can conclude that our method is only slightly slower than LRTV.
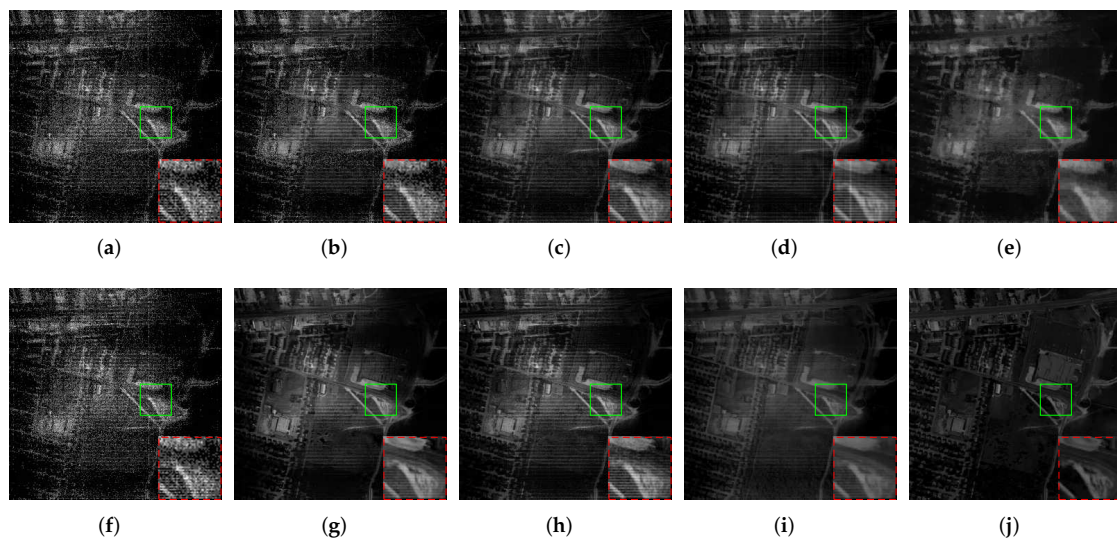
Figure 13. Restored results of band 151 in Urban dataset. (**a**) observed; (**b**) BM4D; (**c**) PCABM4D; (**d**) PARAFAC; (**e**) 3DTV; (**f**) TDL; (**g**) LRMR; (**h**) NAIRLMA; (**i**) LRTV; (**j**) 3DATVLR.
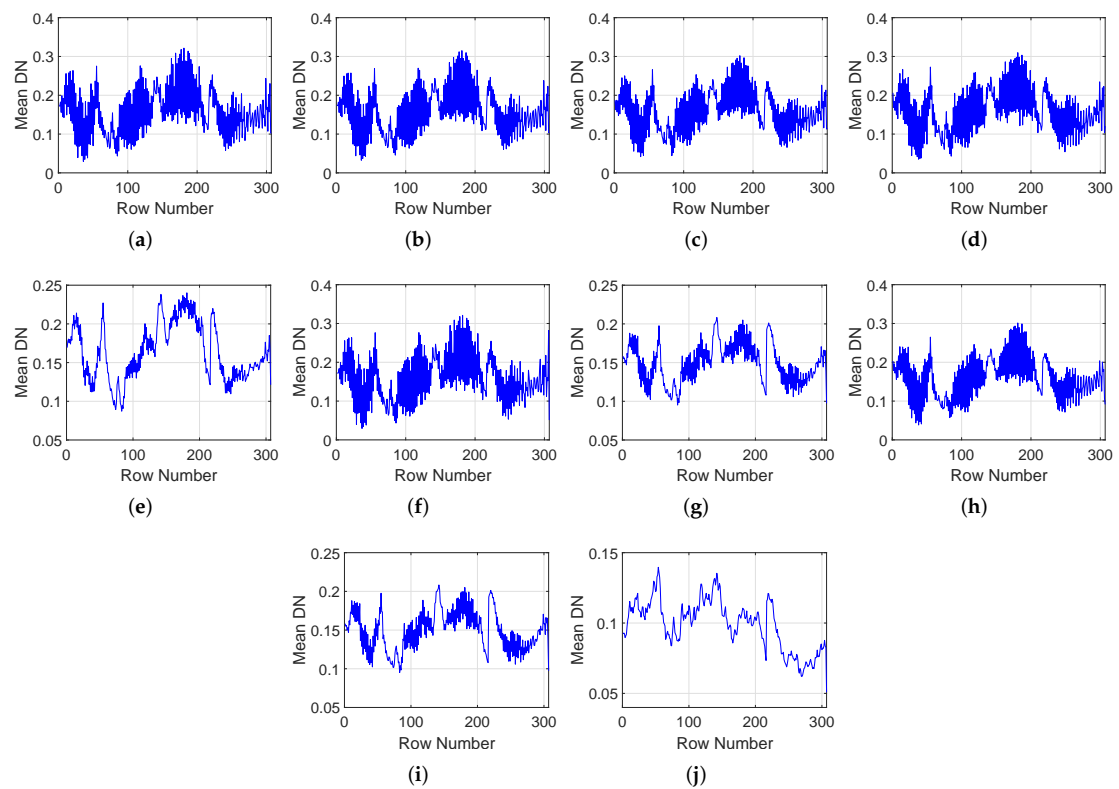
Figure 14. Horizontal mean profile of band 151 in Urban dataset. (**a**) Observed; (**b**) BM4D; (**c**) PCABM4D; (**d**) PARAFAC; (**e**) 3DTV; (**f**) TDL; (**g**) LRMR; (**h**) NAIRLMA; (**i**) LRTV; (**j**) 3DATVLR.

Table 3. Blind quantitative evaluation for real Urban dataset using the Q-metric with boldface indicating the best results.

Methods	Original	3DTV	LRMR	LRTV	3DATVLR
aQ	60.78	61.44	62.38	62.84	**65.13**
Times (s)	-	3590	2811	549	648

5. Conclusions

Hyperspectral image mixed denoising is one of the important and critical pre-processing steps for subsequent applications. In this paper, we present a novel restoration method for HSIs mixed denoising that uses the low-rank constraint and 3DATV regularization. The proposed 3DATVLR method can effectively exploit the spectral similarity by low-rank representation and it can also explore the spectral-spatial smoothness by 3D anisotropic TV regularization from the HSI cube. The problem is finally formulated as a combination of all convex regularizations (i.e., low rank regularization and 3DATV regularization) that could be easily solved via ADMM by alternatively separating it into several easier subproblems.

This study mainly concerns on HSI denoising with mixed noises, e.g., Gaussian noise, impulse noise and stripes or dead lines. Two hyperspectral data sets, i.e., the simulated Indian Pines and real Urban, were used as simulated and real data, respectively. Both simulated and real data experiments validate that the 3DATV regularization can effectively alleviate the impact of the structured sparse noise and finally help to improve the accuracy of the denoising results. In addition, both experiments provide evidence that the proposed 3DATVLR method overwhelmingly outperforms the state-of-the-art methods at removing the mixed noises and preserving the spectra of the original HSI. For an LRTV method, it often fails to remove the sparse noise when severe strips or dead lines exist in the scene, and the reason may be that the band-by-band TV has one of the same directions with the strips or dead lines, whereas 3DATVLR can effectively deal with such noise case by taking advantage of the spectral TV.

In the near future, we will implement the proposed 3DATV method in the high-performance computing platform, such as NVIDIA CUDA and cloud computing, to accelerate the speed of it for more applications.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
2. Sun, L.; Wu, Z.; Xiao, L.; Liu, J.; Wei, Z.; Dang, F. A novel $l_{1/2}$ sparse regression method for hyperspectral unmixing. *Int. J. Remote Sens.* **2013**, *34*, 6983–7001. [CrossRef]
3. Sun, L.; Ge, W.; Chen, Y.; Zhang, J.; Jeon, B. Hyperspectral unmixing employing l_1-l_2 sparsity and total variation regularization. *Int. J. Remote Sens.* **2018**, *34*, 1–24. [CrossRef]
4. Wu, Z.; Shi, L.; Li, J.; Wang, Q.; Sun, L.; Wei, Z.; Plaza, J.; Plaza, A. GPU parallel implementation of spatially adaptive hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1131–1143. [CrossRef]
5. Gu, B.; Sheng, V.S. A robust regularization path algorithm for ν-support vector classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 1241–1248. [CrossRef] [PubMed]
6. Zhang, L.; Zhao, C. A spectral-spatial method based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection. *Int. J. Remote Sens.* **2017**, *38*, 4047–4068. [CrossRef]

7. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [CrossRef] [PubMed]

8. Buades, A.; Coll, B.; Morel, J. A non-local algorithm for image denoising. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 60–65. [CrossRef]

9. Kumar, B.S. Image denoising based on non-local means filter and its method noise thresholding. *Signal Image Video Process.* **2013**, *7*, 1211–1227. [CrossRef]

10. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef] [PubMed]

11. Wang, J.; Li, T.; Shi, Y.Q.; Lian, S.; Ye, J. Forensics feature analysis in quaternion wavelet domain for distinguishing photographic images and computer graphics. *Multimed. Tools Appl.* **2017**, *76*, 23721–23737. [CrossRef]

12. Xiong, L.; Xu, Z.; Shi, Y.Q. An integer wavelet transform based scheme for reversible data hiding in encrypted images. *Multidimens. Syst. Signal Process.* **2017**, *29*, 1–12. [CrossRef]

13. Wang, J.; Lian, S.; Shi, Y.Q. Hybrid multiplicative multi-watermarking in DWT domain. *Multidimens. Syst. Signal Process.* **2017**, *28*, 617–636. [CrossRef]

14. Yuan, C.; Sun, X.; Lv, R. Fingerprint liveness detection based on multi-scale LPQ and PCA. *China Commun.* **2016**, *13*, 60–65. [CrossRef]

15. Othman, H.; Qian, S. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 397–408. [CrossRef]

16. Chen, G.; Qian, S. Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis. *Can. J. Remote Sens.* **2008**, *34*, 447–454. [CrossRef]

17. Chen, G.; Qian, S.E. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 973–980. [CrossRef]

18. Chen, G.; Bui, T.D.; Quach, K.G.; Qian, S.E. Denoising hyperspectral imagery using principal component analysis and block-matching 4D filtering. *Can. J. Remote Sens.* **2014**, *40*, 60–66. [CrossRef]

19. Brook, A. Three-dimensional wavelets-based denoising of hyperspectral imagery. *J. Electron. Imaging* **2015**, *24*, 013034. [CrossRef]

20. Rasti, B.; Sveinsson, J.; Ulfarsson, M.; Benediktsson, J. Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2458–2467. [CrossRef]

21. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O. Wavelet-based sparse reduced-rank regression for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6688–6698. [CrossRef]

22. Renard, N.; Bourennane, S.; Blanc-Talon, J. Denoising and dimensionality reduction using multilinear tools for hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 138–142. [CrossRef]

23. Guo, X.; Huang, X.; Zhang, L.; Zhang, L. Hyperspectral image noise reduction based on rank-1 tensor decomposition. *ISPRS J. Photogramm. Remote Sens.* **2013**, *83*, 50–63. [CrossRef]

24. Lu, T.; Li, S.; Fang, L.; Ma, Y.; Benediktsson, J. Spectral–spatial adaptive sparse representation for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 373–385. [CrossRef]

25. Li, J.; Yuan, Q.; Shen, H.; Zhang, L. Noise removal from hyperspectral image with joint spectral–spatial distributed sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5425–5439. [CrossRef]

26. Fu, Y.; Lam, A.; Sato, I.; Sato, Y. Adaptive spatial-spectral dictionary learning for hyperspectral image restoration. *Int. J. Comput. Vis.* **2017**, *122*, 228–245. [CrossRef]

27. Tian, Q.; Chen, S. Cross-heterogeneous-database age estimation through correlation representation learning. *Neurocomputing* **2017**, *238*, 286–295. [CrossRef]

28. Yuan, Y.; Zheng, X.; Lu, X. Spectral–spatial kernel regularized for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3815–3832. [CrossRef]

29. Xie, W.; Li, Y. Hyperspectral imagery denoising by deep learning with trainable nonlinearity function. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1963–1967. [CrossRef]

30. Wang, B.; Gu, X.; Ma, L.; Yan, S. Temperature error correction based on BP neural network in meteorological wireless sensor network. *Int. J. Sens. Netw.* **2017**, *23*, 265–278. [CrossRef]

31. Qu, Z.; Keeney, J.; Robitzsch, S.; Zaman, F.; Wang, X. Multilevel pattern mining architecture for automatic network monitoring in heterogeneous wireless communication networks. *China Commun.* **2016**, *13*, 108–116. [CrossRef]

32. Xu, L.; Li, F.; Sato, I.; Wong, A.; Clausi, D. Hyperspectral image denoising using a spatial–spectral Monte Carlo sampling approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3025–3038. [CrossRef]

33. Zheng, Y.; Jeon, B.; Sun, L.; Zhang, J.; Zhang, H. Student's t-hidden Markov model for unsupervised learning using localized feature selection. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, 1–14. [CrossRef]

34. Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; Wei, Z. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1990–2000. [CrossRef]

35. Sun, L.; Wang, S.; Wang, J.; Zheng, Y.; Jeon, B. Hyperspectral classification employing spatial–spectral low rank representation in hidden fields. *J. Ambient Intell. Hum. Comput.* **2017**, 1–12. [CrossRef]

36. Lu, X.; Wang, Y.; Yuan, Y. Graph-regularized low-rank representation for destriping of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4009–4018. [CrossRef]

37. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [CrossRef]

38. Zhu, R.; Dong, M.; Xue, J. Spectral nonlocal restoration of hyperspectral images with low-rank property. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3062–3067. [CrossRef]

39. He, W.; Zhang, H.; Zhang, L.; Shen, H. Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3050–3061. [CrossRef]

40. Wang, M.; Yu, J.; Xue, J.; Sun, W. Denoising of hyperspectral images using group low-rank representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4420–4427. [CrossRef]

41. Sun, L.; Jeon, B.; Zheng, Y.; Wu, Z. Hyperspectral image restoration using low-rank representation on spectral difference image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1150–1155. [CrossRef]

42. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188. [CrossRef]

43. Sun, L.; Jeon, B.; Soomro, B.N.; Zheng, Y.; Wu, Z.; Xiao, L. Fast superpixel based subspace low rank learning method for hyperspectral denoising. *IEEE Access* **2018**, *6*, 12031–12043. [CrossRef]

44. Cao, W.; Wang, K.; Han, G.; Yao, J.; Cichocki, A. A Robust PCA Approach With Noise Structure Learning and Spatial–Spectral Low-Rank Modeling for Hyperspectral Image Restoration. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *1*, 1–17. [CrossRef]

45. Chang, X.; Zhong, Y.; Wang, Y.; Lin, S. Unified Low-Rank Matrix Estimate via Penalized Matrix Least Squares Approximation. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *1*, 1–12. [CrossRef] [PubMed]

46. Yuan, Q.; Zhang, L.; Shen, H. Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3660–3677. [CrossRef]

47. Zhang, H. Hyperspectral image denoising with cubic total variation model. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *7*, 95–98. [CrossRef]

48. Chang, Y.; Yan, L.; Fang, H.; Luo, C. Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping. *IEEE Trans. Image Process.* **2015**, *24*, 1852–1866. [CrossRef] [PubMed]

49. Goldluecke, B.; Cremers, D. An approach to vectorial total variation based on geometric measure theory. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 327–333.

50. Aggarwal, H.K.; Majumdar, A. Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 442–446. [CrossRef]

51. Wu, Z.; Wang, Q.; Wu, Z.; Shen, Y. Total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising. *J. Electron. Imaging* **2016**, *25*, 013037. [CrossRef]

52. Wang, Q.; Wu, Z.; Jin, J.; Wang, T.; Shen, Y. Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising. *Signal Process.* **2018**, *142*, 11–26. [CrossRef]

53. Zhang, L.; Zhang, Y.; Wei, W.; Li, F. 3D total variation hyperspectral compressive sensing using unmixing. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 2961–2964.

54. He, S.; Zhou, H.; Wang, Y.; Cao, W.; Han, Z. Super-resolution reconstruction of hyperspectral images via low rank tensor modeling and total variation regularization. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 6962–6965.

55. Sun, L.; Zheng, Y.; Jeon, B. Hyperspectral restoration employing low rank and 3d total variation regularization. In Proceedings of the International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 23–25 December 2016; pp. 326–329. [CrossRef]

56. Chang, Y.; Yan, L.; Zhong, S. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4260–4268.

57. Sun, L.; Wu, Z.; Liu, J.; Wei, Z. Supervised hyperspectral image classification using sparse logistic regression and spatial-tv regularization. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, Australia, 21–26 July 2013; pp. 1019–1022. [CrossRef]

58. Sun, L.; Wu, Z.; Liu, J.; Xiao, L.; Wei, Z. Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1490–1503. [CrossRef]

59. Eckstein, J.; Yao, W. Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.* **2015**, *11*, 619–644.

60. Cai, J.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [CrossRef]

61. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–184. [CrossRef] [PubMed]

62. Bioucas-Dias, J.M.; Nascimento, J.M. Hyperspectral subspace identification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2435–2445. [CrossRef]

63. Liu, X.; Bourennane, S.; Fossati, C. Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3717–3724. [CrossRef]

64. Peng, Y.; Meng, D.; Xu, Z.; Gao, C.; Yang, Y.; Zhang, B. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2949–2956.

65. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]

66. Sun, L.; Jeon, B.; Zheng, Y.; Wu, Z. A novel weighted cross total variation method for hyperspectral image mixed denoising. *IEEE Access* **2017**, *5*, 27172–27188. [CrossRef]

67. Qian, Y.; Ye, M. Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 499–515. [CrossRef]

68. Zhu, X.; Milanfar, P. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE Trans. Image Process.* **2010**, *19*, 3116–3132. [CrossRef] [PubMed]