remote sensing

MDPI

Article

Multispectral Image Denoising via Nonlocal Multitask Sparse Learning

Ya-Ru Fan [1], Ting-Zhu Huang [1,*], Xi-Le Zhao [1], Liang-Jian Deng [1] and Shanxiong Fan [2]

[1] School of Mathematical Sciences/Institute of Computational Science, University of Electronic Science and Technology of China, Chengdu 611731, China; 201411100101@std.uestc.edu.cn (Y.-R.F.); photographyer@uestc.edu.cn (X.-L.Z.); liangjian.deng@uestc.edu.cn (L.-J.D.)

[2] School of Mathematics and Statistics, Yunnan University, Kunming 650091, China; fanshanxiong@cdrcb.com

* Correspondence: tzhuang@uestc.edu.cn

Received: 7 November 2017; Accepted: 10 January 2018; Published: 16 January 2018

Abstract: The goal of multispectral imaging is to obtain the spectrum for each pixel in the image of a scene and deliver much reliable information. It has been widely applied to several fields including mineralogy, oceanography and astronomy. However, multispectral images (MSIs) are often corrupted by various noises. In this paper, we propose a MSI denoising model based on nonlocal multitask sparse learning. The nonlocal self-similarity across space and the high correlation of the MSI along the spectrum via multitask sparse learning are fully exploited in the proposed model. A nonnegative matrix factorization (NMF) based algorithm is developed to solve the proposed model. Experimental results on both simulated and real data demonstrate that the proposed method performs better than several existing state-of-the-art denoising methods.

Keywords: multispectral image denoising; nonlocal self-similarity; multitask sparse learning; nonnegative matrix factorization

1. Introduction

With plenty of available spectral information, multispectral image (MSI) has received considerable attention in wide applications such as image interpretation [1,2], superresolution [3–5], compression [6–8] and surveillance [9]. In practice, MSIs are inevitably corrupted by various noises due to the equipment limitations, the loss of radiance energy and narrow band width. Noise is always the major challenge that seriously degrades the quality of MSIs. Thus, it is necessary to remove noise in many MSI applications. Recently, there are several methods to efficiently solve MSI denoising problems. In particular, sparse coding has attracted much attention due to its state-of-the-art performance in MSI processing [10,11]. These methods assume that images can be linearly sparse represented by only a few basis atoms of a dictionary with the sparse coefficients. The noise can be greatly reduced, since noise is stochastic that cannot be sparsely approximated, but the noise-free MSI can be sparsely approximated by a few basis atoms.

The choice of dictionary is not tractable in sparse coding [12]. In early sparse coding based MSI applications, a fixed dictionary such as discrete cosine transform (DCT) dictionary and discrete wavelet transform (DWT) dictionary [13] is adopted, which are the orthogonal dictionaries and have the geometric invariant property [14]. In addition, the dictionary composed of two or more fixed basis transform such as DCT and DWT, is often employed for the MSI and hyperspectral image (HSI) processing [15], which obtains good performance due to combining the advantages of the different dictionaries.

Actually, the performance of the MSI denoising method using a data-driven dictionary is superior to the one with a fixed dictionary [16,17]. The data-driven dictionary can be learned according to the current result in the iterations. Therefore, dictionary learning for MSI denoising

problem is very popular. Based on principal component analysis (PCA) [5], K-means [18], K-singular value decomposition (K-SVD) [19] and nonnegative matrix factorization (NMF) [20] technologies, the dictionary learning approaches have achieved the outstanding performance in various MSI applications. In particular, the NMF based dictionary learning approaches have been widely used for solving the MSI unmixing, denoising and fusion problems since they can sufficiently utilize the sparsity property of MSIs [20,21].

In this paper, we propose a new MSI denoising model by fully employing the spatio-spectral sparsity of the MSI, and a NMF based dictionary learning algorithm to solve the proposed model, which alternately updates the basis atoms of the dictionary and the corresponding sparse representation. An emphasis behind the proposed model is the utilization of nonlocal sparsity by the patch grouping in each band. The other emphasis is that the high correlation of the MSI along spectrum is exploited via multitask sparse learning, as Figure 1 illustrated. Multitask sparse learning aims to enhance performance by accomplishing a task together with other related tasks, using a shared sparse representation across tasks, since what is learned for each task can help other tasks to be learned better [22]. We consider each band denoising problem as one task, and the denoising tasks on all bands share a common sparse representation according to the correlation of adjacent bands. In [23], a HSI denoising model based on the multitask learning was proposed, but it only adopted the correlation of adjacent patches and failed to use the nonlocal self-similarity of each band image. The proposed model simultaneously takes spectral and spatial sparsity of the MSI into account and generates a competitive denoising result.

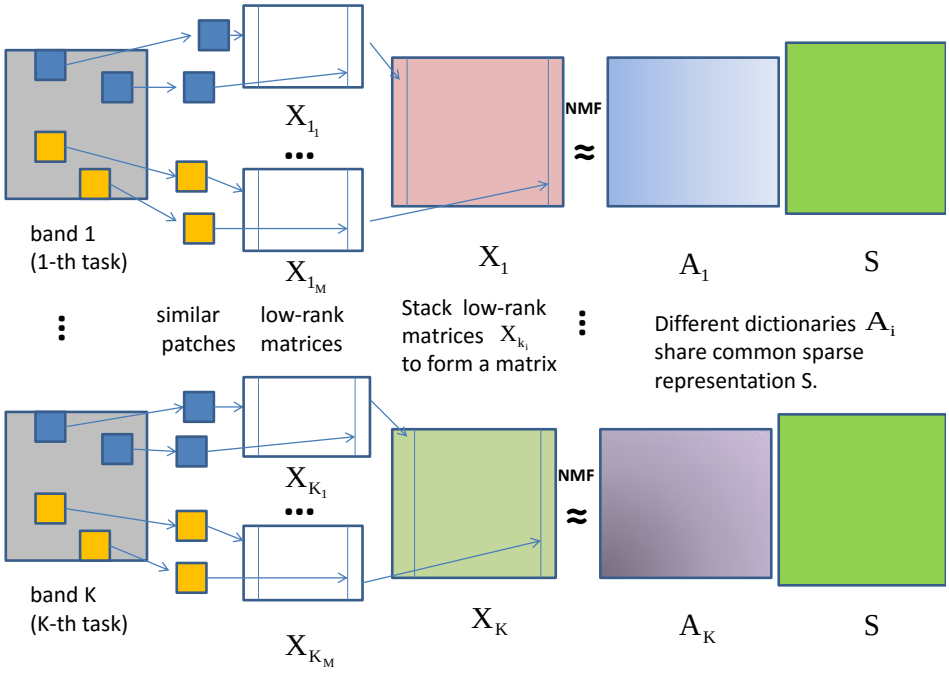


Figure 1. Framework of the proposed multispectral image denoising method. Each band denoising problem is considered as one task, and the denoising tasks on all bands share a common sparse representation S.

The outline of this paper is as follows: Section 2 introduces the MSI denoising problem based on sparse coding. In Section 3, we propose a new MSI denoising model, which simultaneously considers the spectral and spatial sparsity of the MSI. Section 4 shows the optimization algorithm for solving the proposed model. In Section 5, we provide numerical experiments on both simulated and real data to demonstrate the effectiveness of the proposed method. Finally, the conclusions and future works are given.

2. Problem Formulation

A MSI is made up of many band images and each band image is a two-dimensional image. For simplicity, the two-dimensional image is generally stacked as a column vector in lexicographic order $\mathbf{x} \in \mathbb{R}^n$ in image processing [24–26]. We focus on solving the MSI denoising problem by sparse coding. Mathematically, the sparse coding model for image denoising problem assumes that $\mathbf{x}$ can be represented by

$$\mathbf{x} \approx \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{n \times r}$ $(n < r)$ is a dictionary matrix and $\mathbf{s} \in \mathbb{R}^r$ is the coefficient vector. Due to the redundant property of the dictionary $\mathbf{A}$, most entries of the coefficient vector $\mathbf{s}$ are zero or close to zero, i.e., $\mathbf{s}$ is sparse.

If the dictionary $\mathbf{A}$ is known, the image denoising problem can be formulated as ℓ_1-norm regularized sparse optimization model, i.e.,

$$\min_{\mathbf{s}} \frac{1}{2}||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2 + \lambda||\mathbf{s}||_1, \tag{2}$$

where $||\cdot||_2$ is the ℓ_2-norm, $||\cdot||_1$ is the ℓ_1-norm and constant λ denotes the regularization parameter controlling the degree of sparsity. Obviously, problem (2) is convex. There are many efficient methods for solving ℓ_1-minimization problem (2), such as iterative thresholding algorithm [27] and Bregman split algorithm [28]. In MSI denoising problem, for the k-th band $\mathbf{x}_k$ of the MSI, we derive an unified model based on model (2) as follows:

$$\min_{\mathbf{s}_k} \frac{1}{2}||\mathbf{x}_k - \mathbf{A}_k\mathbf{s}_k||_2^2 + \lambda_k||\mathbf{s}_k||_1, \tag{3}$$

where λ_k is the regularization parameter of the k-th band denoising problem.

If the dictionary $\mathbf{A}_k$ is learned in problem (3), which becomes a non-convex optimization problem. A popular approach to simultaneously estimate $\mathbf{A}_k$ and $\mathbf{s}_k$ is the sparse NMF method, which integrates dictionary learning and sparse coding into one model, i.e.,

$$(\mathbf{s}_k, \mathbf{A}_k) = \arg\min_{\mathbf{s}_k, \mathbf{A}_k} \frac{1}{2}||\mathbf{x}_k - \mathbf{A}_k\mathbf{s}_k||_2^2 + \lambda_k||\mathbf{s}_k||_1, \tag{4}$$

$$\text{s.t.} \quad \mathbf{s}_k \geq 0, \mathbf{A}_k \geq 0.$$

The nonnegative constraint is consistent with the sparse constraint in the NMF method. Under fairly mild conditions, the sparsest decomposition of Formulation (1) by NMF is indeed unique [29]. In the following section, we will present the patch based MSI denoising method, which fully exploits the nonlocal similarity of the MSI across space and the high correlation of the MSI along the spectrum via multitask sparse learning.

3. The Proposed Method

We present a novel model for MSI denoising problem, which simultaneously uses the spectral and spatial sparsity priors of the MSI. The proposed model consists of two components: patch grouping for employing nonlocal self-similarity of MSI across space and multitask sparse learning for characterizing the high correlation of the MSI along the spectrum. Concretely, we denote the i-th image patch of size $\sqrt{n} \times \sqrt{n}$ at position i in the k-th band as $\mathbf{y}_{k_i} \in \mathbb{R}^n$, reordered into the column vector lexicographically. We divide the k-th band into a set of key patches $\mathbf{G}_k = \{\mathbf{y}_{k_i} \in \mathbb{R}^n\}_{i=1}^M$ (where M is the number of the key patches). The MSI contains much global and local redundancy and correlation in spatial and spectral dimensions. For each band, a small number of key patches could represent the overall band image, which means M is a small value. By performing block matching technoledge [30], a set of patches that are most similar to each key patch $\mathbf{y}_{k_i}$ can be found.

After image patch grouping, we align these similar patches to form a matrix $\mathbf{X}_{k_i} = [\mathbf{y}_{k_{i_1}}, \cdots, \mathbf{y}_{k_{i_T}}] \in \mathbb{R}^{n \times T}$, where T denotes the number of these nonlocal similar patches. The matrix $\mathbf{X}_{k_i}$ is low-rank since the similar patches have the similar structures. For the k-th band, we record these low-rank matrices as the column vector lexicographically to form a data matrix $\mathbf{X}_k \in \mathbb{R}^{N \times M}$, where $N = n \cdot T$, as Figure 1 illustrated. The data matrix $\mathbf{X}_k$ may be low-rank, and its size is not very large since n, T and M are small values such as $n = 49$, $T = 45$ in numerical experiments.

By treating the MSI denoising problem as a multitask learning problem, in which multiple tasks are solved at the same time with shared representation, we consider each band denoising problem as one task, and the denoising tasks on all bands share a sparse representation. Since adjacent bands of the MSI are related, we can well employ the high correlation of the MSI along the spectrum by solving all bands denoising problems in parallel using a shared representation. Given K data matrices $\mathbf{X}_1$, …, $\mathbf{X}_K$, we simultaneously learn K dictionaries $\mathbf{A}_1$, …, $\mathbf{A}_K$ with a common sparse representation $\mathbf{S}$. As [31,32] presented, learning multiple related tasks simultaneously could obtain better results than learning each task independently, since what is to learn for each task can help other tasks be learned better. Therefore, we conjecture that solving multiple related denoising tasks simultaneously could enhance denoising performance. To capture useful information from the corrupted MSI with as few basis atoms as possible, the shared representation is very sparse.

In [33], multitask learning based on the capped-ℓ_1, ℓ_1 regularization performed better than that with ℓ_1-norm regularization. Here, we use the capped-ℓ_1, ℓ_1 regularization in our model to enhance sparsity. We formulate the proposed model for MSI denoising as follows:

$$(\hat{\mathbf{A}}_1, \cdots, \hat{\mathbf{A}}_K, \hat{\mathbf{S}}) = \underset{\mathbf{A}_1, \cdots, \mathbf{A}_K, \mathbf{S}}{\arg\min} \sum_{k=1}^{K} (\frac{1}{2} ||\mathbf{X}_k - \mathbf{A}_k \mathbf{S}||_F^2 + \lambda_k \sum_{i=1}^{M} \min(||s^i||_1, \theta)), \tag{5}$$

where $\mathbf{A}_1, \cdots, \mathbf{A}_K \in \mathbb{R}_+^{N \times R}$ are the dictionaries on different bands, λ_k $(k = 1, \cdots, K)$ is the regularization parameter, $\mathbf{S} \in \mathbb{R}_+^{R \times M}$ is the common sparse representation and R is the size of the dictionary, s^i is the i-column of matrix $\mathbf{S}$ and $\theta > 0$ is a thresholding parameter. The first term in model (5) is the fidelity term and the second term is the capped-ℓ_1, ℓ_1 regularization. The proposed model simultaneously utilizes the nonlocal self-similarity by patch grouping across space and the global correlation along the spectrum via multitask sparse learning.

Obviously, model (5) is equivalent to the following model:

$$(\hat{\mathbf{A}}_1, \cdots, \hat{\mathbf{A}}_K, \hat{\mathbf{S}}) = \underset{\mathbf{A}_1, \cdots, \mathbf{A}_K, \mathbf{S}}{\arg\min} \sum_{k=1}^{K} \sum_{i=1}^{M} (\frac{1}{2} ||x_k^i - \mathbf{A}_k s^i||_F^2 + \lambda_k \min(||s^i||_1, \theta)), \tag{6}$$

where x_k^i denotes the i-column of data matrix $\mathbf{X}_k$.

By an indicator function, we rewrite the model (6) as follows:

$$(\hat{\mathbf{A}}_1, \cdots, \hat{\mathbf{A}}_K, \hat{\mathbf{S}}) = \underset{\mathbf{A}_1, \cdots, \mathbf{A}_K, \mathbf{S}}{\arg\min} \sum_{k=1}^{K} \sum_{i=1}^{M} (\frac{1}{2} ||x_k^i - \mathbf{A}_k s^i||_F^2 + \lambda_k \beta^i ||s^i||_1), \tag{7}$$

where $\beta^i = I(||s^i||_1 < \theta)$ and $I(\cdot)$ is $\{0, 1\}$ valued indicator function.

Each band has a corresponding dictionary since different bands deliver different information although they are related. For the MSI denoising problem, this model (5) exploits the high correlation of the MSI along the spectrum by the multitask sparse learning and nonlocal sparsity via patch grouping across space, which makes full use of the spectral and spatial sparsity priors of the MSI. By solving the above model (5), we can obtain the good estimations of dictionaries $\hat{\mathbf{A}}_1, \cdots, \hat{\mathbf{A}}_K$ and the shared sparse representation $\hat{\mathbf{S}}$. The matrix $\hat{\mathbf{X}}_k$ containing the noise-free patches is recovered by

$$\hat{\mathbf{X}}_k = \hat{\mathbf{A}}_k \hat{\mathbf{S}} \quad (k = 1, \cdots, K). \tag{8}$$

Finally, each band of the MSI is reconstructed simultaneously by averaging the corresponding pixels of the noise-free patches.

4. Optimization Algorithm

There are several algorithms to solve the proposed model such as alternating direction method of multipliers (ADMM) [34] and primal-dual method [35]. Here, to save the computational complexity, we adopt a multiplicative update rule based NMF algorithm [36] to solve the proposed model.

We first denote two matrix operations: $\otimes$ defines the elementwise product and $\oslash$ stands for the elementwise division. Let the objective function as follows:

$$F(\mathbf{A}_1, \cdots, \mathbf{A}_K, \mathbf{S}) = \sum_{k=1}^{K}\sum_{i=1}^{M}(\frac{1}{2}||x_k^i - \mathbf{A}_k s^i||_F^2 + \lambda_k \beta^i ||s^i||_1) \tag{9}$$

$$= \sum_{k=1}^{K}(\frac{1}{2}||\mathbf{X}_k - \mathbf{A}_k\mathbf{S}||_F^2 + \hat{\lambda}_k||\mathbf{S}||_1),$$

where $\hat{\lambda}_k = \lambda_k \beta^i$.

Derive the Karush–Kuhn–Tucker (KKT) conditions of the model (5), i.e.,

$$\begin{cases} \mathbf{S} \geq 0, & \\ \mathbf{A}_k \geq 0, & k = 1, \ldots, K, \\ \nabla_{\mathbf{S}} F \geq 0, & \\ \nabla_{\mathbf{A}_k} F \geq 0, & k = 1, \ldots, K, \\ \mathbf{S} \otimes \nabla_{\mathbf{S}} F = 0, & \\ \mathbf{A}_k \otimes \nabla_{\mathbf{A}_k} F = 0, & k = 1, \ldots, K, \end{cases} \tag{10}$$

where

$$\nabla_{\mathbf{S}} F = \sum_{k=1}^{K}(\mathbf{A}_k^T\mathbf{A}_k\mathbf{S} - \mathbf{A}_k^T\mathbf{X}_k + \hat{\lambda}_k\mathbf{1}), \tag{11}$$

$$\nabla_{\mathbf{A}_k} F = \mathbf{A}_k\mathbf{S}\mathbf{S}^T - \mathbf{X}_k\mathbf{S}^T, \quad k = 1, \ldots, K. \tag{12}$$

Due to $\mathbf{S} \otimes \nabla_{\mathbf{S}} F = 0$ and $\mathbf{A}_k \otimes \nabla_{\mathbf{A}_k} F = 0$ in KKT conditions, we have

$$\mathbf{S} \otimes (\sum_{k=1}^{K}(\mathbf{A}_k^T\mathbf{A}_k\mathbf{S} + \hat{\lambda}_k\mathbf{1})) = \mathbf{S} \otimes (\sum_{k=1}^{K}(\mathbf{A}_k^T\mathbf{X}_k)), \tag{13}$$

$$\mathbf{A}_k \otimes (\mathbf{A}_k\mathbf{S}\mathbf{S}^T) = \mathbf{A}_k \otimes (\mathbf{X}_k\mathbf{S}^T), \quad k = 1, \ldots, K. \tag{14}$$

According to the multiplicative update rule based NMF algorithm [36], we obtain the following multiplicative update rules to solve the model (5):

$$\mathbf{S} \leftarrow \mathbf{S} \otimes (\sum_{k=1}^{K}(\mathbf{A}_k^T\mathbf{X}_k)) \oslash (\sum_{k=1}^{K}(\mathbf{A}_k^T\mathbf{A}_k\mathbf{S} + \hat{\lambda}_k\mathbf{1})), \tag{15}$$

$$\mathbf{A}_k \leftarrow \mathbf{A}_k \otimes (\mathbf{X}_k\mathbf{S}^T) \oslash (\mathbf{A}_k\mathbf{S}\mathbf{S}^T), \quad k = 1, \ldots, K. \tag{16}$$

The solution of model (5) can be achieved by alternatively employing the above update rules (15) and (16). The overall procedure is summarized in Algorithm 1. In addition, the computational cost is $KR(2N+1)(M+R) + MR(2R+3)$ for updating $\mathbf{S}$ in each iteration, and $2R(NM + MR + NR + N)$ for updating $\mathbf{A}_k$ in each iteration. Thus, the total complexity of the Algorithm 1 is the sum of the computational costs from computing $\mathbf{S}$ and $\mathbf{A}_k$, which is much less than the tensor based approaches for the MSI denoising problem.

Referring to [33,36], Algorithm 1 satisfies the block coordinate descent procedure. Thus, the value of objective function (9) is non-increasing.

Algorithm 1: MSI denoising via nonlocal multitask sparse learning

1. **Initialization**:
Stack the observed MSI into K matrices $\mathbf{X}_1, \ldots, \mathbf{X}_K \in \mathbb{R}_+^{N \times M}$;
Initialize $\mathbf{A}_1, \ldots, \mathbf{A}_K \in \mathbb{R}_+^{N \times R}$ and $\mathbf{S} \in R_+^{R \times M}$ with random elements in [0,1];
Set regularization parameters $\lambda_1, \ldots, \lambda_K$ and $\theta = \frac{1}{K} \sum_{i=1}^{K} \lambda_i$;
2. **Repeat**:
(a) compute $\mathbf{S}$ by the update rule (15);
(b) for $k = 1$ to K do
compute $\mathbf{A}_k$ by the update rule (16);
end for
(c) **Until** the maximum number of iterations is reached;
3. Output the dictionaries $\hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_K$ and the common coefficient $\hat{\mathbf{S}}$;
4. The final result: $\hat{\mathbf{X}}_k = \hat{\mathbf{A}}_k \hat{\mathbf{S}}, \quad k = 1, \cdots, K$.

5. Numerical Results

To demonstrate the effectiveness of the proposed method for MSI denoising, we present the comparisons on both simulated and real data. Five state-of-the-art MSI denoising methods are utilized for comparisons. They are band-wise KSVD [37], band-wise BM3D [30], 3D-cube KSVD [38], PARAFAC [39], and MTSNMF [23]. All parameters involved in the comparing algorithms are carefully tuned to obtain the best performance or automatically chosen as described in their papers. The experimental results are compared both quantitatively and visually.

5.1. Simulated Experiments for MSI Denoising

We select two MSIs from the Columbia MSI database for the simulated experiments. The MSI database contains 32 real-world scenes of a wide variety of real-world materials and objects, and each MSI has spatial resolution 512×512 and 31 bands, which includes full spectral resolution reflectance values collected from 400 nm to 700 nm with 10 nm steps. We scale the MSIs into the range of $[0, 1]$ in simulated experiments. In addition, we add the Gaussian noise into two test MSIs to generate noisy observations. In practice, different bands of the MSI contain different noise levels. Following [23], the standard deviations of Gaussian distribution for different bands are set as random number from the range of $[0.05, 0.2]$, and we store the random numbers in a descending order as $\sigma_1, \sigma_2, \cdots, \sigma_K$. Namely, the noise level decreases gradually from the first band to the last band. The dictionary size is suggested to be set to $R = 4n$ empirically in [38]. For the regularization parameters, we just simply set $\lambda_k = \sigma_k \sqrt{2 \log(R)}$, patch size 7×7 and threshold $\theta = \frac{1}{K} \sum_{i=1}^{K} \lambda_i$. The maximum number of iterations is 300 for the proposed method.

To evaluate the denoising performance of the proposed method, four quantitative picture quality indices (PQI) are exploited, including peak signal-to-noise ratio (PSNR), structure similarity [40] (SSIM), feature similarity [41] (FSIM), erreur relative globale adimensionnelle de synthèse [42] (ERGAS). In image processing and computer vision, PSNR and SSIM are two primary and conventional PQIs. They evaluate the similarity between the ground truth image and the restored image based on the mean squared error (MSE) and structural consistency, respectively. The FSIM focuses on the perceptual consistency with the restored image. The higher values of these three measures indicate that the restored image is more closer to the ground truth image. The ERGAS estimates fidelity of the restored image based on the weighted sum of MSE in each band. It is worth pointing out that the smaller ERGAS value means a better estimation of the ground truth image.

In Figures 2 and 3, we compare the denoising results obtained by five methods with the results achieved by the proposed approach on band 1 and band 30 of the MSI *peppers*. Obviously, the denoising

performances of band-wise KSVD and band-wise BM3D methods are not good since their restored band images are unclear. The MTSNMF and PARAFAC approaches create some artifacts in the denoising results maybe due to employing less self-similarity of image in their methods. The restored image obtained by the 3D-cube KSVD method shows blurring artifacts on the edges. However, the proposed method removes most noise and obtains clear image details especially in the region corresponding to the white rectangle in clean *peppers*.

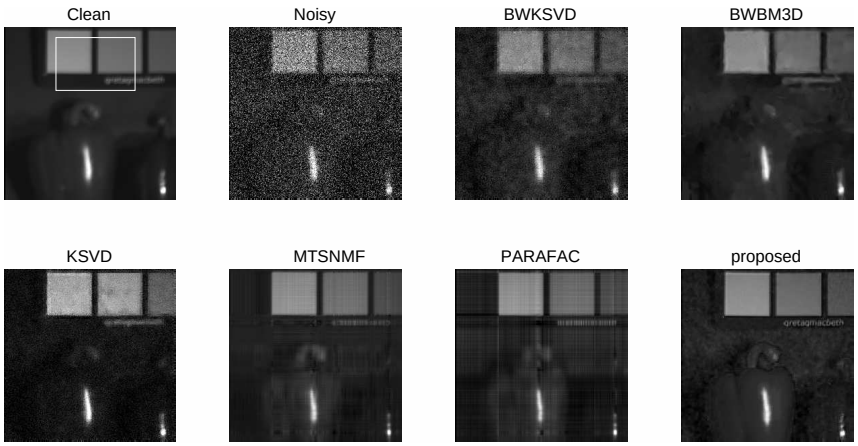


Figure 2. From left to right in the first row: band 1 of the clean *peppers*, the noisy *peppers*, the restored results obtained by band-wise KSVD, and band-wise BM3D; from left to right in the second row: the band 1 of the restored results obtained by 3D-cube KSVD, MTSNMF, PARAFAC and the proposed method. Focus on comparing the region of all test methods corresponding to the white rectangle in the clean band.

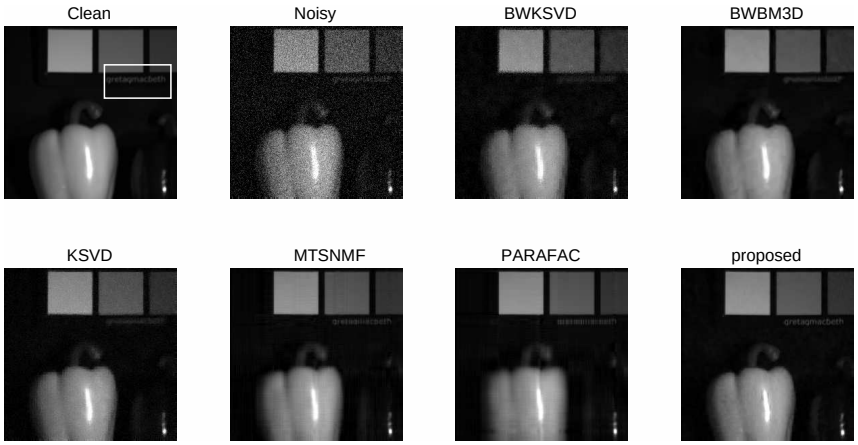


Figure 3. From left to right in the first row: band 30 of the clean *peppers*, the noisy *peppers*, the restored results obtained by band-wise KSVD, and band-wise BM3D; from left to right in the second row: band 30 of the restored results obtained by 3D-cube KSVD, MTSNMF, PARAFAC and the proposed method. Focus on comparing the region of all test methods corresponding to the white rectangle in the clean band.

Table 1 displays the comprehensive performance of six test methods on *peppers* with respect to four PQIs. From these quantitative comparisons, it is easy to observe that the proposed method outperforms other test methods according to all evaluation measures.

We also show the denoising results of six test methods on the MSI *pompoms*. The denoising results of band 6 and band 21 of *pompoms* are reported in Figures 4 and 5, respectively. We can see that the proposed method evidently performs better than others, both on the recovery of the edges and texture

structures, especially in the region corresponding to the white rectangle in the clean band. These restored images obtained by five test approaches are a little blurry and still contain some noise. Furthermore, Table 2 indicates the comprehensive performance of six test methods on *pompoms* with respect to four PQIs, which presents the advantage of the proposed method in quantitative comparisons.

Table 1. Performance of six comparing methods with respect to four picture quality indices on *peppers*.

Method	PSNR	SSIM	FSIM	ERGAS
band-wise KSVD	30.12	0.609	0.760	157.51
band-wise BM3D	38.14	0.934	0.949	62.433
3D-cube KSVD	33.08	0.775	0.844	116.67
PARAFAC	35.41	0.877	0.934	85.47
MTSNMF	38.41	0.816	0.772	60.75
the proposed method	**38.79**	**0.957**	**0.952**	**54.89**

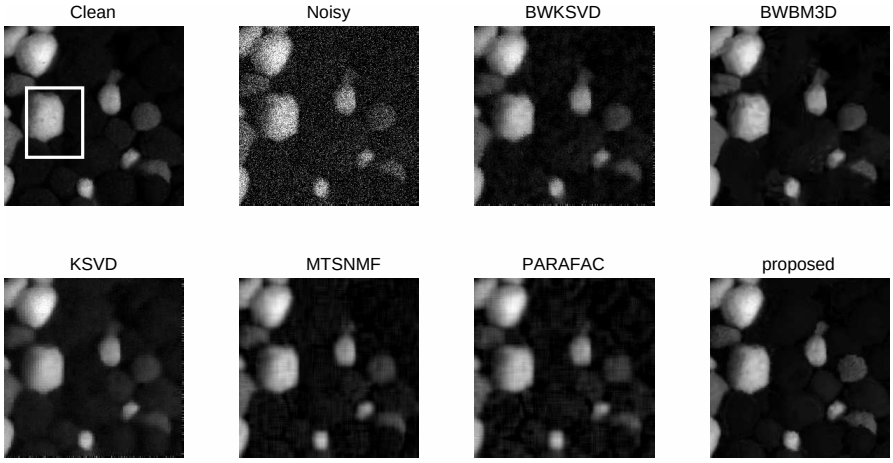


Figure 4. From left to right in first row: The band 6 of the clean *pompoms*, the noisy *pompoms*, the restored results obtained by band-wise KSVD, band-wise BM3D; From left to right in second row: The band 6 of the restored results obtained by 3D-cube KSVD, MTSNMF, PARAFAC and the proposed method. Focus on comparing the region of all test methods corresponding to the white rectangle in clean band.

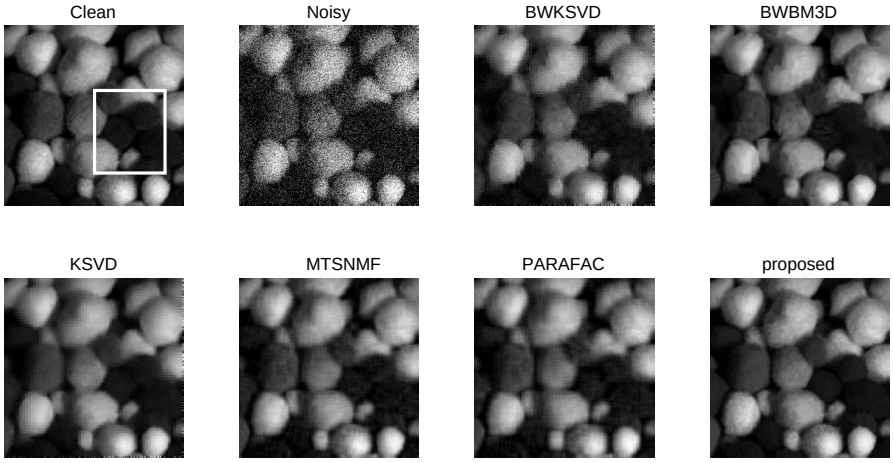


Figure 5. From left to right in first row: The band 21 of the clean *pompoms*, the noisy *pompoms*, the restored results obtained by band-wise KSVD, band-wise BM3D; From left to right in second row: The band 21 of the restored results obtained by 3D-cube KSVD, MTSNMF, PARAFAC and the proposed method. Focus on comparing the region of all test methods corresponding to the white rectangle in clean band.

Table 2. Performance of six comparing methods with respect to four picture quality indices on *pompoms*.

Method	PSNR	SSIM	FSIM	ERGAS
band-wise KSVD	27.64	0.624	0.838	253.74
band-wise BM3D	32.52	0.859	0.921	83.87
3D-cube KSVD	28.40	0.731	0.876	140.36
PARAFAC	30.09	0.782	0.907	111.15
MTSNMF	31.41	0.823	0.721	84.47
the proposed method	**33.16**	**0.871**	**0.923**	**75.04**

5.2. A Real Experiment

A real remote sensing data is utilized to evaluate the MSI denoising performance, i.e., urban area HYDICE MSI. The size of the original MSI is $304 \times 304 \times 210$. We remove some bands that are seriously polluted by the atmosphere. The remaining test data with the size of $304 \times 304 \times 157$. Like simulated experiments, we scale the test MSI into the interval [0, 1] and adopt the same parameter settings. In addition, since the noise level is unknown for real data, we show a qualitative evaluation based on the visual effect.

The denoising results of the band 50 obtained by five test methods are presented in Figure 6. The close-ups of the white rectangle in the noisy band 50 in Figure 6 are presented in Figure 7 corresponding to five test methods. We distinctly observe that the noisy band 50 is dark and contains some mixed noise. In Figure 6, the restored bands by test methods are similar visually. However, in Figure 7, the difference is evident between the proposed method and other test methods. There are some noise in the restored images obtained by the band-wise KSVD and 3D-cube KSVD approaches. The PARAFAC method returns a blurry image. The MTSNMF method removes some noise and achieves a clear restored band image, but its restored image loses some details. Nevertheless, the proposed method creates a clear image by finely recovering the details and edges, and removing a lot of noise.

All of the above numerical experiments attest that the proposed approach outperforms other test methods both quantitatively and visually.

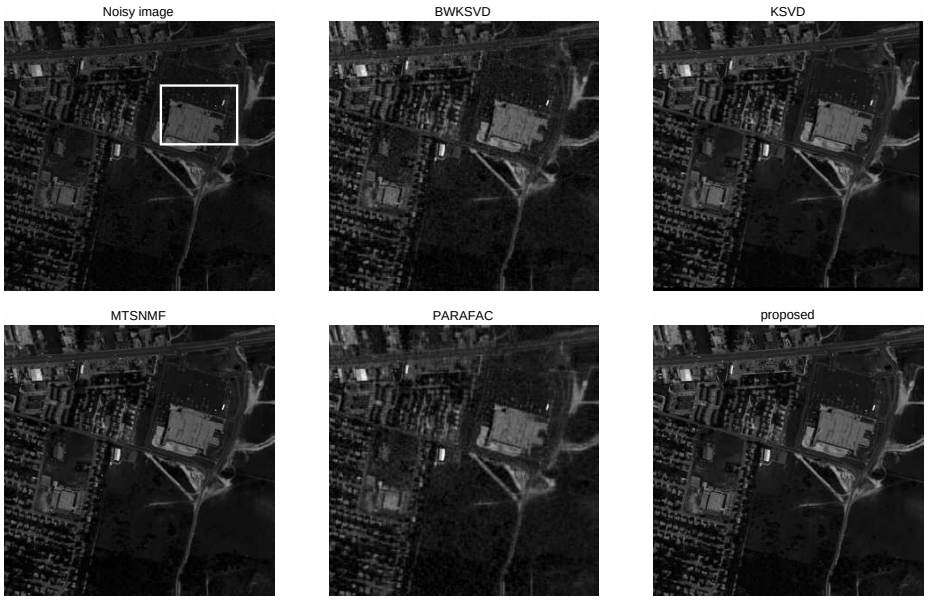


Figure 6. From left to right in the first row: band 50 of urban area HYDICE, the restored results obtained by band-wise KSVD, 3D-cube KSVD; from left to right in the second row: band 50 of the restored results obtained by MTSNMF, PARAFAC and the proposed method.

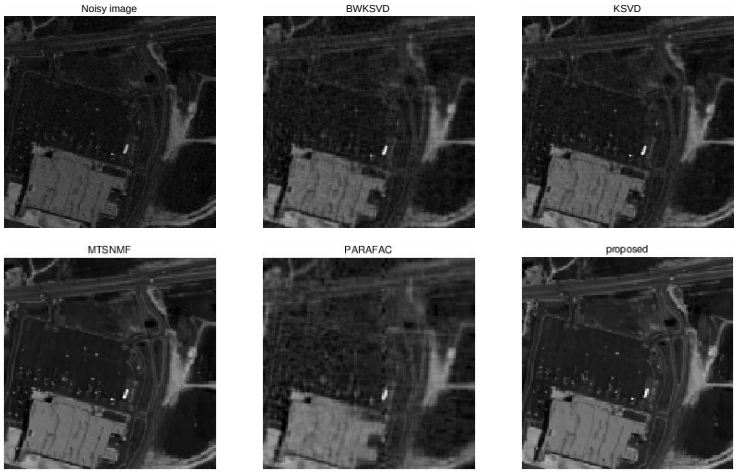
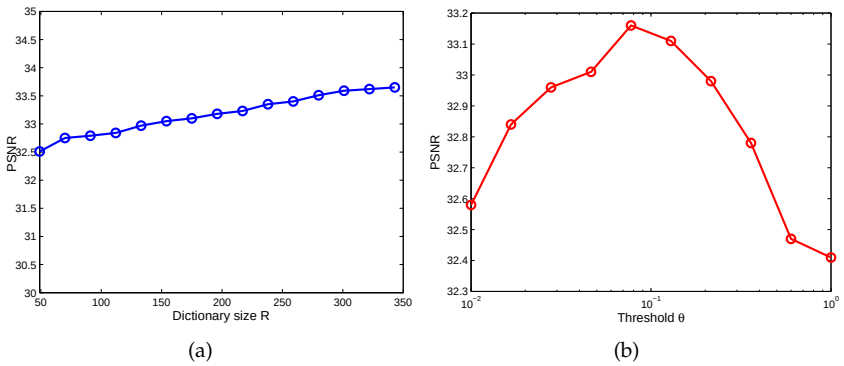


Figure 7. Close-ups of the white rectangle in the noisy band 50 of urban area HYDICE.

6. Discussion

The proposed method uses nonlocal self-similarity across space and the high correlation of the MSI along the spectrum; thus, it achieves the excellent performance, as reported in Tables 1 and 2. The band-wise KSVD method and the band-wise BM3D method are dictionary learning based denoising methods for two-dimensional images. They don't consider the correlation of the MSI along the spectrum. Thus, they obtain the lower PSNR and SSIM values than the proposed method. The PARAFAC method exploits decomposition uniqueness and single rank character to obtain the better performance than some two-dimensional filter based denoising methods. However, the PARAFAC method only considers the correlation along the spectrum. As shown in Figures 2 and 3, the results of the PARAFAC method are not good. The MTSNMF method takes the spatio-spectral sparsity priors into account, but it doesn't exploit the nonlocal self-similarity, which is very efficient for image denoising. Therefore, the results of the proposed method are much better than results of the MTSNMF method as shown in Tables 1 and 2.

Figure 8 shows the results of the proposed method with different parameters on *pompoms*. In Figure 8a, we can see a larger size of dictionary R may lead to higher PSNR value but more complexity. With the increasing R, the PSNR value of the proposed method becomes almost stable, which means overcompleteness of sparse representation in the dictionary learning offers some geometric invariant properties [43] and advantages the true signal recovery. In Figure 8b, we discover that an appropriate threshold θ is helpful for our algorithm to achieve a good performance.



(a) (b)

Figure 8. Results of the proposed method with different parameters. (**a**) dictionary size R vs. PSNR (**b**) threshold θ vs. PSNR.

7. Conclusions

In this paper, we proposed a MSI denoising model, which fully exploited the nonlocal self-similarity of the MSI on the spatial domain and the global correlation on the spectral domain via multitask sparse learning. An efficient NMF based algorithm was developed to solve the proposed model. Experimental results both on simulated and real data demonstrated that the proposed method performed better than several existing state-of-the-art MSI denoising methods. Usually, the MSIs are huge, and implementation of parallel-computing for the MSI denoising problem will be considered as our future work.

Acknowledgments: This work was supported by the National Nature Science Foundation of China (61772003).

Author Contributions: All authors contributed to the design of the methodology and the validation of experiments. Ya-Ru Fan wrote the paper. Ting-Zhu Huang, Xi-Le Zhao and Liang-Jian Deng reviewed and revised the paper.

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

References

1. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*; John Wiley & Sons, New York, NY, USA, 1994.
2. Zhao, X.L.; Wang, F.; Huang, T.Z.; Ng, M.K.; Plemmons, R.J. Deblurring and sparse unmixing for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4045–4058.
3. Amro, I.; Mateos, J.; Vega, M.; Molina, R.; Katsaggelos, A.K. A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP J. Adv. Sig. Proc.* **2011**, *2011*, 79.
4. Zhang, H.; Zhang, L.; Shen, H. A super-resolution reconstruction algorithm for hyperspectral images. *Signal Process.* **2012**, *92*, 2082–2096.
5. Jiang, C.; Zhang, H.; Shen, H.; Zhang, L. A practical compressed sensing-based pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 629–633.
6. Markas, T.; Reif, J. Multispectral image compression algorithms. In Proceedings of the IEEE Data Compression Conference, Snowbird, UT, USA, 30 March–2 April 1993; pp. 391–400.
7. König, F.; Praefcke, W. Multispectral image encoding. In Proceedings of the IEEE International Conference on Image Processing, Kobe, Japan, 24–28 October 1999; Volume 3, pp. 45–49.
8. Parmar, M.; Lansel, S.; Wandell, B.A. Spatio-spectral reconstruction of the multispectral datacube using sparse recovery. In Proceedings of the 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 473–476.
9. Yuen, P.W.; Richardson, M. An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. *Imaging Sci. J.* **2010**, *58*, 241–253.
10. Mairal, J.; Elad, M.; Sapiro, G. Sparse representation for color image restoration. *IEEE Trans. Image Process.* **2008**, *17*, 53–69.
11. Wei, Q.; Bioucas-Dias, J.; Dobigeon, N.; Tourneret, J.Y. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3658–3668.
12. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Hyperspectral image denoising using 3D wavelets. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 1349–1352.
13. Duijster, A.; Scheunders, P.; De Backer, S. Wavelet-based em algorithm for multispectral-image restoration. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3892–3898.
14. Simoncelli, E.P.; Freeman, W.T.; Adelson, E.H.; Heeger, D.J. Shiftable multiscale transforms. *IEEE Trans. Inf. Theory* **1992**, *38*, 587–607.
15. 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.

16. Tosic, I.; Frossard, P. Dictionary learning. *IEEE Signal Process. Mag.* **2011**, *28*, 27–38.
17. Mairal, J.; Bach, F.; Ponce, J. Task-driven dictionary learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 791–804.
18. Theiler, J.P.; Gisler, G. Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation. In Proceedings of the International Society for Optics and Photonics, Optical Science, Engineering and Instrumentation, San Diego, CA, USA, 24 October 1997; pp. 108–118.
19. Wang, L.; Lu, K.; Liu, P.; Ranjan, R.; Chen, L. IK-SVD: Dictionary learning for spatial big data via incremental atom update. *Comput. Sci. Eng.* **2014**, *16*, 41–52.
20. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 528–537.
21. Lu, X.; Wu, H.; Yuan, Y.; Yan, P.; Li, X. Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2815–2826.
22. Evgeniou, T.; Pontil, M. Regularized multi—Task learning. In Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; ACM: New York, NY, USA, 2004; pp. 109–117.
23. Ye, M.; Qian, Y.; Zhou, J. Multitask sparse nonnegative matrix factorization for joint spectral–spatial hyperspectral imagery denoising. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2621–2639.
24. Ma, T.H.; Huang, T.Z.; Zhao, X.L. Group-based image decomposition using 3-D cartoon and texture priors. *Inf. Sci.* **2016**, *328*, 510–527.
25. Deng, L.J.; Guo, W.; Huang, T.Z. Single-image super-resolution via an iterative reproducing kernel Hilbert space method. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 2001–2014.
26. Fan, Y.R.; Huang, T.Z.; Liu, J.; Zhao, X.L. Compressive sensing via nonlocal smoothed rank function. *PLoS ONE* **2016**, *11*, e0162041.
27. Daubechies, I.; Defrise, M.; De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **2004**, *57*, 1413–1457.
28. Zhang, X.; Burger, M.; Bresson, X.; Osher, S. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM J. Imaging Sci.* **2010**, *3*, 253–276.
29. Kawakami, R.; Matsushita, Y.; Wright, J.; Ben-Ezra, M.; Tai, Y.W.; Ikeuchi, K. High-resolution hyperspectral imaging via matrix factorization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 2329–2336.
30. 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.
31. Ando, R.K.; Zhang, T. A framework for learning predictive structures from multiple tasks and unlabeled data. *J. Mach. Learn. Res.* **2005**, *6*, 1817–1853.
32. Argyriou, A.A.; Evgeniou, T.; Pontil, M. Convex multi-task feature learning. *Mach. Learn.* **2008**, *73*, 243–272.
33. Gong, P.; Ye, J.; Zhang, C. Multi-stage multi-task feature learning. *J. Mach. Learn. Res.* **2013**, *14*, 2979–3010.
34. He, B.; Tao, M.; Yuan, X. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Optim.* **2012**, *22*, 313–340.
35. Chambolle, A.; Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **2011**, *40*, 120–145.
36. Lee, D.D.; Seung, H.S. Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* **2001**, *13*, 556–562.
37. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322.
38. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745.
39. 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.
40. 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.
41. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386.

42. Wald, L. *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*; Presses des MINES: Paris, French, 2002.
43. Rubinstein, R.; Bruckstein, A.M.; Elad, M. Dictionaries for sparse representation modeling. *Proc. IEEE* **2010**, *98*, 1045–1057.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).