1. Introduction

Hyperspectral image (HSI) processing is one of the most powerful techniques in remote sensing. It has a wide range of applications in mineral exploration [1], environmental monitoring [2–4], military surveillance [5, 6], and other related fields. Despite the development of advanced techniques in hyperspectral sensors, noise is unavoidably introduced in the acquisition process, which limits HSI applications and has a negative impact on many subsequent tasks, such as unmixing, classification, and object detection. Therefore, HSI denoising is a classic problem that is still an active area of research in HSI preprocessing.

In the past decades, many algorithms have been proposed for noise reduction in HSI. A simple idea is to adopt a conventional denoising method for natural images band-by-band or pixel-by-pixel. However, these methods cannot obtain satisfactory results because they ignore the spectral correlation between different bands, which is of great importance for improving denoising quality. At the same time, the spectral correlation is destroyed by the above methods and produces extra distortion. Hence, the spatial and spectral dimensions should be considered simultaneously. Othman and Qian proposed a noise reduction scheme using a wavelet shrinkage based hybrid spatial-spectral derivative domain in which the noise level is elevated. The method utilized the dissimilarity of signal regularity in the corresponding spatial and spectral dimensions [7]. Assuming that the image noise exists in the low-energy principal component analysis (PCA) channels, Chen and Qian proposed a denoising algorithm based on PCA and wavelets [8]. In their algorithm, 2D bivariate wavelet thresholding is used to remove the noise in low-energy PCA channels and a 1D dual-tree complex wavelet transform (DCWT) based denoising method is utilized to remove the spectrum noise of the HSI data cube.

Inspired by the excellent performance of the total variation method for natural image denoising, in [9–11], a cubic total variation model was proposed for HSI denoising that combined the 2D total variation in the spatial domain with the 1D total variation in the spectral dimension and achieved competitive image quality [12]. Yuan et al. then proposed an adaptive spectral-spatial total variation model in which the differences of spectral noise and spatial information are both considered to better suppress the noise [13]. Motivated by this method, Liao et al. proposed a two-stage method that combines kernel principal component analysis (KPCA) and
Journal of Sensors

total variation, which obtained promising denoising results [14].

Recently, sparse representation (SR) based methods have been attracting more attention in image denoising as a powerful prior model [15–18]. In the SR framework, the signal is assumed to be represented by a linear combination of a few atoms in a learned or designed dictionary. Based on this viewpoint, the true signal can be recovered from its noisy observation with an ill-posed inverse problem by enforcing the sparsity prior. Rasti et al. proposed a new linear model that modeled the HSI with singular value decomposition and wavelets and then reconstructed the true image based on sparse regularization [19]. This approach achieves some visual improvement and can be applied to large data sets. In [20], a sparse low rank model for HSI was used to solve the denoising problem. The two benefits of the proposed model are dimensionality reduction with PCA and the exploitation of DCWT coefficients associated with the principle components separately. To better utilize the local and global redundancy in the spatial and spectral domains of HSI, Zhao and Yang [21] proposed a method using SR and the low rank constraint, which can be thought as utilizing two regularizers to constrain the local and global redundancy respectively and shows state-of-the-art performance.

In light of these methods, in this paper, we propose a novel algorithm for HSI denoising using composite regularization models. First, based on the patch SR, we present a specific way to extract the patches from the HSI by considering the information in both the spatial and spectral dimensions. On the one hand, this approach could enable the extracted patches to maintain their correlation in the local spectral dimension. On the other hand, with a nonlocal technique (NLT), it could also exploit the nonlocal self-similarity to preserve the detailed texture information in the spatial domain. In view of the excellent performance of the prior constraint for denoising, a framework with two composite regularization models (total variation and nonlocal group sparsity) is proposed for HSI denoising. Furthermore, to solve the nonconvex objective function in the framework, the alternating direction method of multipliers (ADMM) is utilized to obtain the approximate solution to the inverse problem.

The rest of the paper is organized as follows. In Section 2, basic theory for image denoising based on regularization will be introduced. Section 3 presents the proposed framework with composite regularization models for HSI denoising and its numerical solution method. The comparison experiments and results are presented in Section 4. Section 5 concludes the paper.

2. Regularization Model Based Image Denoising

Image denoising aims to recover the true image from its noisy observation, which can be formulated as

\[ f = u + \nu, \]

where \( f \in \mathbb{R}^N \) is the degradation observation, \( N \) is the whole pixel numbers in image, \( u \in \mathbb{R}^N \) is the true image, and \( \nu \sim \mathcal{N}(0, \sigma) \) is the zero-mean additive Gaussian noise. To get the estimation image \( \hat{u} \) as close as possible to the noise-free image \( u \), a solution with \( L_2 \)-norm constraint can be obtained as

\[
\hat{u} = \arg \min_u \| f - u \|_2^2,
\]

where \( \| \cdot \|_2 \) denotes the \( L_2 \)-norm operator and \( \| f - u \|_2 \) is called data fidelity term. Due to the nonuniqueness of the solution for the inverse problem in (2), the prior knowledge is utilized to supply a better solution space by regularizing the above ill-posed problem. So the formulation in (2) can be rewritten as follows:

\[
\hat{u} = \arg \min_u \| f - u \|_2^2 + \eta \varphi (u),
\]

where \( \varphi(u) \) is the prior constraint on the true image, which will help to obtain a better solution and improve the image quality. \( \eta \) is a scalar parameter to balance the above two terms.

It is widely known that prior knowledge plays an essential role in the regularization of a denoising algorithm. From the formulation in (3), we can also see that different priors produce different performance. Hence, designing a term to effectively reflect the prior is a critical issue in image denoising. Total variation is a classical regularization model that is widely used in image restoration, in which it is assumed that the image is locally smooth except at its discontinuous edges. Therefore, it could oversmooth the texture and achieve worse performance in fine structures, as it only accounts for the local statistical characteristics. Some additional studies on total variation have been proposed to further improve its denoising quality [22, 23].

Another popular model that has led to promising result for image denoising is patch-based sparsity prior knowledge. The sparse model assumes that each patch from an image can be represented by a few atoms of a dictionary. Thanks to a number of coding methods, the noise-free patches can also be accurately recovered using effective \( l_p \)-norm minimization or its relaxed version, \( l_1 \)-norm minimization. Typical methods to do this include the basis pursuit [24], matching pursuit [25], iterative thresholding [26], and split Bregman [27] methods.

In recent years, the NLT has attracted a large amount of attention for its powerful performance on sharper edges and ability to preserve fine-scale details. Because of the large amount of self-similarity exhibited in natural images from pixel to block level, a number of NLT-based regularization terms have been exploited for image inverse problem [28–30].

Although the regularization model successfully denoises images, there remain some shortcomings when it is used in HSI denoising. The first is that the aforementioned conventional regularization models can only be used band-by-band in HSI denoising. This ignores the significant correlation between spectral dimensions and produces additional spectral distortion. The second is that the single prior model used for denoising is not enough to obtain superior results. Hence, in the next section, we discuss how to solve these two shortcomings using our proposed novel framework.
3. Proposed Composite Regularization Model Framework

3.1. Patch-Based Extraction Method. For HSI denoising, to obtain better recovery quality, the patch extracting way should be designed carefully. To consider the spatial and spectral dimension simultaneously, a simplest way is to extract local cube patches from HSI directly. However, the problem of this way is that there are too many pixels in each cube with only a small size, such as $8 \times 8 \times 3$, which will cost huge computational complexity for solving the regularization optimization. Meanwhile, incorporating the NLT, some extra cost will be produced due to the cube matching. Furthermore, large error will be caused when the Euclidean distance between cubes is measured with increased dimension.

Hence, an effective means to extract the patch is key to HSI denoising. To avoid the above problem, we propose a specific way to extract the patch by considering the spatial-spectral information adaptively. To this end, we vectorize the images in each band from the HSI and then rearrange them as the columns of a new 2D matrix $U$, which can be expressed as

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_l \end{bmatrix},$$  \hspace{1cm} (4)

where $u_p \in \mathbb{R}^{MN}$ $(p = 1, \ldots, l)$ is the vectorized image in each band, $M$ and $N$ are the number of rows and columns of the band image, and $l$ is the number of spectral bands.

We take the image San Diego as an example. We reshape it into a 2D matrix, part of which is shown in Figure 1. As illustrated in Figure 1, the rows and columns of this 2D matrix present the spatial and spectral dimensions of the original band image separately. That is, we transform the spatial-spectral information included in the HSI cube into a more compact 2D form.

The advantages of extracting patches from the reshaped 2D matrix are as follows. As shown in Figure 1, the patch with a white border includes the pixels located in the same spatial position within neighboring spectral bands. This is utilized as a unit in our framework, which employs the spatial-spectral information jointly. In Figure 1, the two similar patches in group 1 are spectrally correlated with each other because of their neighboring locations in the spectral dimension of the reshaped 2D matrix. Nevertheless, other similar patches in group 2 come from the same neighboring spectral bands with different spatial location. This indicates that the specific patch extraction method supplies a new way to exploit the spatial and spectral information simultaneously. That means we can construct a certain patch-based prior to constrain the nonlocal self-similarity across the spectral and spatial information by easily extracting patch from reshaped 2D images just as the general patch extracting method in [28], which will lead to a better performance on preserving fine-scale structures inevitability for HSI denoising problem.

3.2. Optimization Problem in the Proposed Framework. To the best of our knowledge, the single prior cannot be enough to obtain promising results. For example, to improve the conventional single total variation model, Liu et al. proposed integrating the nonlocal prior into total variation to take advantage of both of them, which achieved better reconstruction performance for CT image [31]. Similarly, Dong et al. exploited the nonlocal self-similarity in natural images and proposed a nonlocal central coding to impose the image to approximate to its clustering center under sparse representation [32]. Therefore, inspired by the achievement of total variation and NLT-based group sparsity, we propose a creative HSI denoising framework with them jointly, which can be expressed as

$$\min_u \left\{ \frac{1}{2} \|f - u\|_2^2 + \eta TV(u) + \tau \text{NGSR}(u) \right\},$$  \hspace{1cm} (5)

where $f$ and $u$ denote the vectoring reshaped observation 2D image and the true HSI 2D matrix. $\tau$ and $\eta$ are the positive scalar parameter to control the balance between the above three terms.

The second term $TV(u)$ presents the total variation prior model, which can be formulated as

$$TV(u) = \|Du\|_1,$$  \hspace{1cm} (6)

where $D = [D_v, D_h]$ is the synthesis gradient operator and $D_v$ and $D_h$ are the vertical and horizontal difference operator, which is well studied in many literatures [31, 32].

The third term $\text{NGSR}(u)$ presents the NLT-based group sparse representation prior model denoted by

$$\text{NGSR} = \sum_{i,j} \left\| W^T G(u_{i,j}) \right\|_{p,q},$$  \hspace{1cm} (7)

where $u_{i,j}$ denotes a patch centered on the element of reshaped 2D matrix located in $(i, j)$, $G(u_{i,j}) = [u_{i,j}, u_{i,j+1}, u_{i+1,j}, \ldots, u_{i,k}, \ldots]$ presents a matrix that takes the similar patches to $u_{i,j}$ as its columns, and $K$ is the index number. $W$ is the wavelet dictionary, which is proved to be an effective tool to represent the image. $\| \cdot \|_{p,q}$ is norm regularization to constrain the

---

**Figure 1:** The reshaped 2D matrix for San Diego.
group sparsity of matrix $A$, whose definition can be denoted by
\[ \|A\|_{p,q} = \sum_i \|\alpha_i\|_p^q \] (8)
where $\alpha_i$ denotes the $i$th row of matrix $A$.

In the optimization in (5), the second term constrains and guarantees the local smoothness, and the third term preserves finer structures by constraining the group sparsity with self-similarity in nonlocal patches. Moreover, because of the specific way patches are extracted, the framework not only utilizes the composite prior constraints, but also exploits the spatial-spectral information jointly. It is believed that better denoising results will be achieved by our proposed framework.

4. Numerical Strategy for Solving the Optimization

In this section, we present the details of the numerical strategy to solve the minimization problem in the proposed framework. The objective function in (5) can be solved for its inverse problems [33]. ADMM is a popular variable splitting technique for solving composite prior constraints via a variable splitting technique. In terms of (11)–(13), we impose auxiliary variable $p^k$ in (13) to be $p^k = \left[ q^k \right]$ and obtain their extension forms as follows:

\[ u^{k+1} = \arg \min_u \left\{ \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| Au + Bw - C + p^k \|_2^2 \right\} \] (14)

and
\[ z^{k+1} = \arg \min_z \left\{ \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \right\} \] (15)

So the objective function in (5) can be rewritten as
\[ \min_{u,z,v} \left\{ \frac{1}{2} \| f - u \|_2^2 + \gamma TV(z) + \tau NGSR(v) \right\} \] (16)

In terms of (11)–(13), we impose auxiliary variable $p^k$ in (13) to be $p^k = \left[ q^k \right]$ and obtain their extension forms as follows:

\[ u^{k+1} = \arg \min_u \left\{ \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| Au + Bw - C + p^k \|_2^2 \right\} \] (17)

\[ z^{k+1} = \arg \min_z \left\{ \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \right\} \] (18)

For more compact expression, the above subproblems are further separated as

\[ u^{k+1} = \arg \min_u \left\{ \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| Au + Bw - C + p^k \|_2^2 \right\} \] (19)

\[ z^{k+1} = \arg \min_z \left\{ \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \right\} \] (20)

\[ q^{k+1} = \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \] (21)

\[ r^{k+1} = \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \] (22)

\[ v^{k+1} = \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \] (23)

\[ w^{k+1} = \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \] (24)

\[ p^{k+1} = \frac{1}{2} \| f - u \|_2^2 + \frac{\rho}{2} \| z \|_2^2 + \gamma TV(z) + \tau NGSR(v) \] (25)
The extension auxiliary variables \( p^k \) can also be updated as
\[
q^{k+1} = q^k + u^{k+1} - z^{k+1},
\]
\[
t^{k+1} = t^k + u^{k+1} - v^{k+1}.
\]

Since the minimization in (18) is a quadratic function, a closed-form solution can be obtained as
\[
u^{k+1} = \frac{1}{1 + 2\rho} \left( f + \rho \left( z^k - p^k \right) + \rho \left( v^k - t^k \right) \right).
\]

The subproblem in (19) can be regarded as a standard total variation problem. Many equally good algorithms have been proposed to solve total variation such as fast iterative shrinkage threshold algorithm (FISTA) [34] and the split Bregman-based method [35]. In our paper, we adopted FISTA to guarantee fast convergence.

The third minimization in (20) is a joint sparsity coding problem, and the group coding can be computed by the method in [17]. Thus, \( \psi^{k+1} \) is obtained by averaging the products of the wavelet dictionary and group coding.

The specific steps of the overall method are described.

Detailed numerical strategy for proposed framework is as follows:

**Input.** The noisy observation HSI \( f \);

**Initialization.** \( u^0 = f, z^0 = \psi^0 = \emptyset, \rho^0 = t^0 = \emptyset, k = 0; \)

**Steps**

1. Update \( u^{k+1} \) with formulation (22);
2. Compute \( \gamma = \eta / \rho; \)
3. Update \( z^{k+1} = \text{argmin}_z \{(1/2)\|z - t^{k+1} - d^t\|_2^2 + \gamma \|Dz\|_1\} \) with FISTA;
4. Compute \( \lambda = \rho / \rho; \)
5. Update \( \psi^{k+1} = \text{argmin}_\psi \{(1/2)\|\psi - u^{k+1} - p^k\|_2^2 + \lambda \sum_{i,j} \|W^TG(v_{i,j})_{P,q}\| \) with the method in [17];
6. Updated the auxiliary variables with formulation (21);

\[
k \leftarrow k + 1;
\]

Until criterion satisfied;

**Experimental Results and Analysis.** To verify the performance of our proposed framework, we applied it to two different HSI data sets, San Diego and Indian Pines. To evaluate it with respect to other methods, we compared the proposed method with the cubic total variation model (CTVM) [12], two-stage denoising method (TwoS) for HSI [14], and PCA based HSI denoising using wavelet shrinkage (PCA+WS) [8]. The assessment indicators are the peak signal-to-noise ratio (PSNR) and recently popular structural similarity (SSIM), which is based on the human vision system [36]. To evaluate the synthesis denoising performance, we plot the mean values of PSNR and SSIM of all the test bands in the HSI data. Considering that some bands of the experimental data sets are heavily corrupted by noise and hence cannot be used as ground truth, we visually selected some cleaner continuous bands as the ground truth test data for the simulation. In the following experiments, the parameters in our proposed algorithm were set only according to the noise level, independently of the test data. The parameters for other comparison methods were set to the original values in their papers. The window size for similar patch search was set to \( 20 \times 20 \) and the patch size was \( 8 \times 8 \). In our experiment, we fixed the number of similar patches in every group, which enabled (20) to be solved more efficiently. The standard deviation \( \sigma \) of the noise was varied from 5 to 20 in increments of 5. Before the experiment, the extracted testing data were normalized to \([0, 1]\) on each channel. Specifically, the atmospheric and water absorption bands were removed from Indian Pines before the test data was extracted.

Because of space limitations, we only present the denoising results of one band (the 40th band) of San Diego using different methods and different standard deviations of the noise. These results are shown in Figures 2 and 3.

It is clear from the visual comparison in Figures 2 and 3 that PCA+WS obtains the worst visual performance among these methods because it accounts for the spatial and spectral dimensions separately, which eradicates many details found in the ground truth. Considering both the difference in the spectral noise and spatial information, CTVM retains more edges compared to PCA+WS because of its spatial-spectral technique. However, it strongly enforces local smoothness, which produces some mottled blocks in the denoised results. In contrast to PCA+WS, TwoS uses kernel PCA to first reduce the uncorrelated noise and then separate out the most significant information. Then, it removes the remaining noise from the low-energy channels with fast primal-dual total variation. Hence, the visual results of TwoS look better with sharper textures than those of PCA+WS. For example, TwoS shows more details of the parking apron at the bottom left, which is often used as a target in HSI target detection. Compared to the other algorithms, our method shows an even more promising performance. It not only suppresses most of the noise, but also preserves more details and sharp edges. Furthermore, from the plots of the mean PSNR and SSIM given in Figure 4, we conclude that our method outperforms other comparative algorithms with respect to these metrics.

To validate the further effectiveness of our proposed method, classification experiments for HSI are tested on Indian Pines. 10 percent of the numbers of testing samples are employed as the training samples. We take the support vector machine as the classifier and categorizes the image into 16 classes. The classification results are depicted in Figure 5. It is clear that the accuracy is improved greatly after denoising processing. Also, to more obviously, we list the overall accuracy (OA) in Table 1. The highest OA of proposed method

<table>
<thead>
<tr>
<th>Method</th>
<th>OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy</td>
<td>75.8%</td>
</tr>
<tr>
<td>PCA+WS</td>
<td>83.2%</td>
</tr>
<tr>
<td>CTVM</td>
<td>84.3%</td>
</tr>
<tr>
<td>TwoS</td>
<td>86.0%</td>
</tr>
<tr>
<td>Ours</td>
<td>87.5%</td>
</tr>
</tbody>
</table>
Figure 2: Denoising results with $\sigma = 20$.

Figure 3: Denoising results with $\sigma = 15$. 
Mean PSNR values

PCA+WS  
CTVM  
TwoS  
Our method  

Noise standard deviation

(a) Mean PSNR curves

Mean SSIM values

PCA+WS  
TwoS  
CTVM  
Our method  

Noise standard deviation

(b) Mean SSIM curves

Figure 4: The objective indicator for HSI denoising.

(a) Classification of noisy HSI  
(b) Classification result of PCA+WS  
(c) Classification of CTVM  
(d) Classification of TwoS  
(e) Classification of our method

Figure 5: Classification results with several methods.
illustrates that our method has the excellent spatial-spectral activity ability and produced less spectral distortion, which in belief is a significant factor to classification performance.

5. Conclusion

In this paper, a novel algorithm for HSI denoising with composite regularization models was proposed. The proposed method includes two main contributions. First, in order to utilize the spatial-spectral information simultaneously, a specific method is developed to extract a patch as a basic processing unit by reshaping the HSI cube into a 2D correlation matrix. Second, to overcome the shortcomings of a single prior, a framework with composite priors is used to characterize local smoothness and nonlocal group sparsity based on the extracted patch, which preserves more essential details. Compared to other state-of-the-art HSI denoising methods, our proposed method achieves competitive performance in both objective assessment and visual quality.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

The paper is supported in part by National Natural Science Foundation of China (Grant 61501147), Natural Science Foundation of Heilongjiang Province (Grant F2015040), and Postdoctoral Science Foundation of Heilongjiang Province (Grant LBH-Z15099).

References


Submit your manuscripts at http://www.hindawi.com