Canadian Journal of Remote Sensing
Journal canadien de télédétection

ISSN: 0703-8992 (Print) 1712-7971 (Online) Journal homepage: http://www.tandfonline.com/loi/ujrs20

Hyperspectral Image Denoising with a Combined Spatial and Spectral Weighted Hyperspectral Total Variation Model

Cheng Jiang, Hongyan Zhang, Liangpei Zhang, Huanfeng Shen & Qiangqiang Yuan

To cite this article: Cheng Jiang, Hongyan Zhang, Liangpei Zhang, Huanfeng Shen & Qiangqiang Yuan (2016) Hyperspectral Image Denoising with a Combined Spatial and Spectral Weighted Hyperspectral Total Variation Model, Canadian Journal of Remote Sensing, 42:1, 53-72, DOI: 10.1080/07038992.2016.1158094

To link to this article: http://dx.doi.org/10.1080/07038992.2016.1158094

Canadian Journal of Remote Sensing, 42:53–72, 2016
Copyright © CASI
ISSN: 0703-8992 print / 1712-7971 online
DOI: 10.1080/07038992.2016.1158094

Hyperspectral Image Denoising with a Combined Spatial and Spectral Weighted Hyperspectral Total Variation Model

Cheng Jiang[1,4], Hongyan Zhang[2,4,*], Liangpei Zhang[2,4], Huanfeng Shen[3,4], and Qiangqiang Yuan[1,4]
[1]*School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China*
[2]*The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China*
[3]*School of Resources and Environmental Sciences, Wuhan University, 129 Luoyu Road, Wuhan 430079, China*
[4]*Collaborative Innovation Center of Geospatial Technology, Wuhan University, 129 Luoyu Road, Wuhan 430079, China*

Abstract. Hyperspectral image (HSI) denoising is a prerequisite for many subsequent applications. For an HSI, the level and type of noise often vary with different bands and spatial positions, which make it difficult to effectively remove noise while preserving textures and edges. To alleviate this problem, we propose a new total-variation model. The main contribution of the proposed approach lies in that the adaptive regularization terms in both the spatial and the spectral dimensions are designed separately and then combined into a unified framework. The 2 separate regularization terms allow a better description of the intrinsic nature of the original HSI data and can simultaneously penalize the noise from both the spatial and spectral perspectives. The designed weights for the regularization terms are positively correlated with the magnitude of the noise intensity and negatively correlated with the signal variation; thus, the original signal can be accurately retained and the noise can be effectively suppressed. To efficiently process the HSI, which appears as a huge data cube, a new optimization algorithm based on the alternating direction method of multipliers (ADMM) procedure is proposed to solve the new model. Experiments using HYDICE and AVIRIS images were conducted to validate the effectiveness of the proposed method.

Résumé. Hyperspectrale l'image (HSI) débruitage est une condition préalable pour de nombreuses applications ultérieures. Pour un HSI, le niveau et le type de bruit varie souvent avec différents groupes et positions spatiales, ce qui rend difficile d'éliminer efficacement le bruit tout en préservant les textures et les bords. Pour pallier ce problème, nous proposons un nouveau modèle de variation totale. Les principales contributions de l'approche proposée mensonge dans la conception des termes de régularisation adaptative dans les deux dimensions spatiales et spectrales, et en les combinant dans un cadre unifié. Les deux termes de régularisation séparés permettent une meilleure description de la nature intrinsèque des données HSI original et peuvent pénaliser simultanément le bruit à la fois des perspectives spatiales et spectrales. Les poids conçus pour les termes de régularisation sont en corrélation positive avec la grandeur de l'intensité du bruit et corrélation négative avec la variation de signal; ainsi, le signal d'origine peut être retenu avec précision et le bruit peut être efficacement supprimée. Pour traiter efficacement le HSI, qui apparaît comme un énorme cube de données, un nouvel algorithme d'optimisation basé sur la méthode de direction alternée de multiplicateurs «alternating direction method of multipliers» (ADMM) procédure est proposée pour résoudre le nouveau modèle. Des expériences utilisant des images AVIRIS et HYDICE et ont été menées afin de valider l'efficacité de la méthode proposée.

INTRODUCTION

With their high spectral resolution, hyperspectral images (HSIs) are commonly used in applications requiring fine identification of materials or precise estimation of physical parameters. As a result of the physical limitations of the sensors (Zhang et al. 2012), HSIs often contain different levels of noise, which not only affects the visual quality, but also can reduce the accuracy of the subsequent processing, e.g., unmixing (Iordache et al. 2012), classification (Harris et al. 2006), clustering (Zhang et al. 2016) and fusion (Jiang et al. 2014). Therefore, with the rapid development of applications using HSIs, the denoising task is becoming more and more important.

For gray images, there are a variety of denoising methods, e.g., the total variation (TV) model (Rudin et al. 1992), Gaussian scale mixtures (Portilla et al. 2003), nonlocal means (Buades et al. 2005), wavelets (Selesnick 2002), anisotropic diffusion (Perona and Malik 1990), and sparse representation (Elad and Aharon 2006). Intuitively, we can apply these meth-

---

Received 1 August 2015. Accepted 15 February 2016.
[*]Corresponding author. e-mail: zhanghongyan@whu.edu.cn.

ods to an HSI in a band-by-band manner. However, because the strong correlations between the HSI bands are overlooked, these methods often perform poorly. To deal with this problem, some researchers (Othman and Qian 2006; Chen and Qian 2011) have proposed to first decorrelate the HSI bands and then successively apply denoising methods in both the spatial and spectral dimensions. These methods work well when the HSI already has a high signal-to-noise ratio (SNR). The redundancy of the spectral bands can also be utilized by subspace-based methods, which first rearrange the HSI into a matrix whose columns contain the spectral signatures and then separate the signal from the noise by estimating the signal subspace (Kuybeda et al. 2007; Acito et al. 2010). However, the common approach is to extend the well-known 2-D denoising methods to higher dimensions. For example, Chen et al. (2011) extended Sendur and Selesnick's (2002) well-known bivariate wavelet thresholding method from gray image denoising to HSI denoising with a 3-D wavelet transform; Muti and Bourennane (2007) extended the classic 2-D Wiener filtering to multidimensional Wiener filtering based on a tensor model and multilinear algebra; Martín-Herrero (2007) extended the 2-D anisotropic diffusion for HSIs with 2 precisely defined diffusion processes in the spatial and spectral dimensions, respectively; and Qian et al. (2012) extended the traditional 2-D nonlocal means to a 3-D perspective. In recent years, methods based on tensor decomposition (Letexier and Bourennane 2008; Liu et al. 2012; Guo et al. 2013; Lin and Bourennane 2013b) and anisotropic diffusion (Méndez-Rial et al. 2010; Wang et al. 2010; Mendez-Rial and Martin-Herrero 2012) have been extensively studied. In addition, some sophisticated strategies have also been investigated with new, emerging technologies. For example, by stacking a local patch of an HSI into a 2-D matrix, Zhang et al. (2014) exploited low-rank matrix recovery theory to denoise the patch. By stacking cubes of voxels into a 4-D group, Maggioni (2013) applied a 4-D transform and collaborative filtering for volumetric data denoising and reconstruction. Based on these existing methods, some hybrid methods have also been investigated, such as combining nonlocal means with sparse representation (Qian and Ye 2013), wavelet packet transform with TUCKER3 decomposition (Lin and Bourennane 2013a), sparse representation with wavelets (Zelinski and Goyal 2006; Rasti et al. 2014), principal component analysis with block matching 3-D (Chen et al. 2012) or 4-D filter (Chen et al. 2014), sparse representation with low-rank constraint (Zhao and Yang 2015) and TV with low-rank constraint (He et al. 2016).

As a simple but effective model, the TV model has been widely used in imaging science, in applications such as image restoration, superresolution, segmentation, inpainting, and unmixing. Very recently, the TV model has also been applied to HSI denoising. For an HSI, there are 2 spatial dimensions (along-track and cross-track) and 1 spectral dimension (wavelength). Yuan et al. (2012) extended the traditional TV model to the spatial and spectral adaptive hyperspectral TV (SSAHTV) model. The SSAHTV model denoises an HSI with the TV regularization in the spatial dimensions and can adjust the degree of smoothing on different pixels and different bands with a single regularization parameter. This model has been proved to be much more effective than the procedure that directly applies the traditional TV model to each band of the HSI. Although the degree of smoothing can be adjusted in a 3-D manner, the model is still not a real 3-D TV model without considering the first-order difference in the spectral dimension. The later work by Yuan et al. (2014) first applied a 2-D adaptive TV model to an HSI twice, from the front and side views, respectively, and then adopted a Q-weighting strategy to fuse the 2 results. The improvement of this approach is a result of implicitly using the TV regularization in the spectral dimension. In addition, there are also methods that explicitly adopt the TV regularization in the spectral dimension, such as the methods proposed in Li et al. (2010) and Zhang (2012); however, the methods proposed in those studies cannot adjust the degree of smoothing, which can significantly affect the denoising performance for an HSI with a nonstationary noise variance. To sum, although the TV model has been used for HSI denoising, a 3-D TV model for HSI denoising, which is simple, effective, and can deal with the signal and noise variation, is still lacking.

In this work, we propose a new *combined spatial and spectral weighted hyperspectral total variation* (CSSWHTV) model for HSI denoising. As a 3-D data cube, the HSI is denoised not only in the spatial dimensions but also in the spectral dimension. The prior model is a combination of 2 separate, finely designed regularization terms that can penalize the variations of pixels in both the spatial and spectral dimensions. Unlike the existing 3-D TV models, the proposed model can automatically adjust the penalizing strength for the variations of the different pixels. To better attenuate the noise and protect the original information, the weight is designed to be positively correlated with the magnitude of the noise intensity and negatively correlated with the signal variation. To solve the new CSSWHTV model, we propose a fast algorithm based on the alternating direction method of multipliers (ADMM; Eckstein and Bertsekas 1992; Gabay and Mercier 1976), a method from the augmented Lagrangian family. From the experimental results with both simulated and real-image data, it is shown that the proposed CSSWHTV model works well, not only for random noise but also for striping noise (Acito et al. 2011b), which validates the effectiveness of the proposed CSSWHTV method.

The remainder of this article is organized as follows: In "Proposed Method," the proposed CSSWHTV model is introduced. In "Numerical Solution," we briefly review the ADMM procedure, and then the proposed algorithm to solve the CSSWHTV model is presented. The experimental results are presented in "Experiments and Discussion." "Conclusions" concludes the article.

PROPOSED METHOD

Many previous studies (Chen and Qian 2011; Othman and Qian 2006) have shown that, to acquire better results, the denoising process should operate not only on the spatial dimensions but also on the spectral dimension. Due to the difference in the signal nature of the spatial and the spectral dimensions (Othman and Qian 2006), we chose to first design the corresponding weighted TV regularization terms separately and then combine them effectively.

The Regularized Hyperspectral Image Denoising Model

A hyperspectral data cube **F** with M samples, N lines, and B bands can be viewed as a 3-D scalar image of size$M \times N \times B$, and it can be expressed as

$$\mathbf{F} = \mathbf{U} + \mathbf{N}, \qquad [1]$$

where **U** is the ideal noise-free data, and **N** is the noise. The denoising process is to seek the original data **U** from the noisy observation **F**. In general, noise can be classified as random noise or fixed-pattern noise. Photon noise (also called shot noise) and thermal noise are 2 examples of random noise, whereas striping noise is a typical example of fixed-pattern noise (Acito et al. 2011b). In push-broom sensors, striping noise, which is caused by the imperfect calibration of the detectors, might appear as a series of stripes in the along-track direction. For many existing sensors such as the Airborne Visible InfraRed Imaging Spectrometer (AVIRIS), the random photon noise, which is signal dependent, is negligible with respect to the random thermal noise, which is signal independent (Acito et al. 2011a). However, as the performance of the electronic components is improved in the new-generation hyperspectral sensors, the signal-dependent photon noise is becoming as important as the signal-independent thermal noise (Alparone et al. 2009). In this study, to focus on the denoising model, we simply assume **N** is random and independent of **U**. The noise variances of the different bands are assumed to be different because of the different wavelength responses of the sensors (Martín-Herrero 2007). For convenience, we define the (x, y, z) coordinates as shown in Figure 1; thus, $u_{i,j,k}$, $f_{i,j,k}$, and $n_{i,j,k}$ denote the pixels of **U**, **F**, and **N** with coordinates (i, j, k), respectively.

With these definitions, the HSI denoising model can be modeled as a regularization-based problem:

$$\hat{\mathbf{U}} = \arg\min \left\{ \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{B} \left(u_{i,j,k} - f_{i,j,k}\right)^2 + \lambda R(\mathbf{U}) \right\} \qquad [2]$$

The first term of Equation (2) is called the data fidelity term, and the second is named the regularization term, which describes the prior information of the original HSI. The regularization parameter λ is a scalar that controls the relative contribution of the data fidelity and the regularization terms.

Weighted Hyperspectral TV Regularization in the Spatial Dimensions

In this section, the classic TV regularization term for a gray image is first reviewed and then extended to HSIs based on vectorial TV theory. Finally, the weighted TV regularization term in the spatial dimensions for HSIs is proposed with a detailed analysis of the weight.

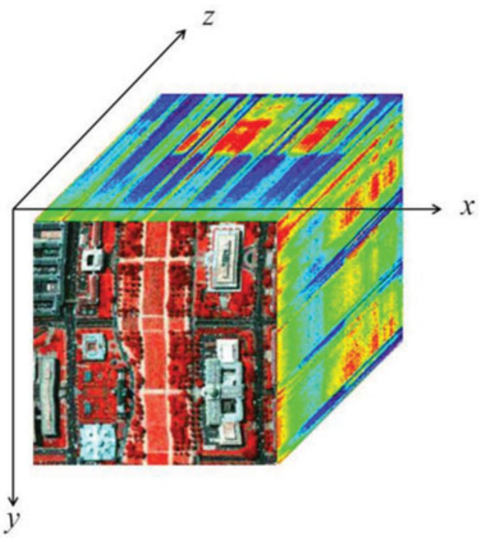


FIG. 1. A 3-D hyperspectral cube and the defined coordinates.

For a gray image $\tilde{\mathbf{u}}$ of size $M \times N$, the classic 2-D isotropic and anisotropic TV regularization terms can be expressed as Equations (3) and (4), respectively:

$$R(\tilde{\mathbf{u}}) = \mathrm{TV}_1(\tilde{\mathbf{u}}) = \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\left(\nabla_x \tilde{u}_{i,j}\right)^2 + \left(\nabla_y \tilde{u}_{i,j}\right)^2}, \qquad [3]$$

$$R(\tilde{\mathbf{u}}) = \mathrm{TV}_2(\tilde{\mathbf{u}}) = \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\left|\nabla_x \tilde{u}_{i,j}\right| + \left|\nabla_y \tilde{u}_{i,j}\right|\right), \qquad [4]$$

where ∇_x and ∇_y are the first-order difference operators in the x and y directions of the image plane, respectively, and $\tilde{u}_{i,j}$ is the pixel of $\tilde{\mathbf{u}}$ with coordinates (i, j). Usually, isotropic regularization is preferred over anisotropic regularization (Bioucas-Dias et al. 2006; Yang et al. 2009). For convenience, the isotropic TV regularization term is utilized in this study.

When viewed from the spatial perspective, an HSI of size $M \times N \times B$ can be seen as B stacked images of size $M \times N$, or a multivalued image. Because the noise level and signal variation vary significantly along the spectral dimension, the first thing to consider is defining a TV model that can automatically balance the penalizing strength for the different bands. This is achieved by adopting the theory of the vectorial TV norm (Bresson and Chan 2008). With the vectorial TV theory, the hyperspectral TV regularization term in the spatial dimensions can be extended from the classic isotropic TV regularization term in Equation (3) to a multivalued case, and defined as

$$\mathrm{HTV}_{\mathrm{Spatial}}(\mathbf{U}) = \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\sum_{k=1}^{B} \left[\left(\nabla_x u_{i,j,k}\right)^2 + \left(\nabla_y u_{i,j,k}\right)^2\right]}. \qquad [5]$$

This formula was first applied to HSI denoising in Yuan et al. (2012), and interested readers can refer to that study for more details.

It should be noted that by using Equation (5), the penalizing strength is balanced for the different bands, which means it is automatically adjusted to every HSI pixel just with the varying index k, rather than the indexes i and j. In other words, the penalizing strength is automatically balanced in the spectral dimension, rather than in the spatial dimensions. To automatically adjust the penalizing strength in the spatial dimensions, a weight $\mathbf{W}_{i,j}$ is added to Equation (5), and the weighted hyperspectral TV regularization term in the spatial dimensions is defined as:

$$\mathrm{WTV}_{\mathrm{Spatial}}(\mathbf{U}) = \sum_{i=1}^{M}\sum_{j=1}^{N}\mathbf{W}_{i,j}\sqrt{\sum_{k=1}^{B}\left[\left(\nabla_x u_{i,j,k}\right)^2+\left(\nabla_y u_{i,j,k}\right)^2\right]}. \quad [6]$$

As 3-D data, every pixel in an HSI shows different variations in both the spatial and spectral dimensions, e.g., the pixels near the edges often have larger variations than those in the smooth areas. For a noisy HSI, the noise intensity also varies in different pixels. Therefore, to better attenuate the noise and retain the variation in the pixels, the desired weight should be positively correlated with the magnitude of the noise intensity and negatively correlated with the variation of the signal. Based on this principle, the weight $W_{i,j}$ in the spatial dimensions is defined as

$$\tau_{i,j} = FV_{i,j} \times (1 - PV_{i,j}/FV_{i,j})^{\alpha}, \quad [7]$$

$$W_{i,j} = \frac{\tau_{i,j}}{\bar{\tau}},\ \bar{\tau} = \sum_{i=1}^{M}\sum_{j=1}^{N}\tau_{i,j}\Big/MN, \quad [8]$$

where α is a positive constant value and is empirically set as 2 in this work. The vectorial gradient magnitudes $FV_{i,j}$ and $PV_{i,j}$ are calculated on the 3-D data $\mathbf{F}$ and $\mathbf{P}$ as Equations (9) and (10), respectively.

$$FV_{i,j} = \sqrt{\sum_{k=1}^{B}\left[\left(\nabla_x f_{i,j,k}\right)^2+\left(\nabla_y f_{i,j,k}\right)^2\right]}, \quad [9]$$

$$PV_{i,j} = \sqrt{\sum_{k=1}^{B}\left[\left(\nabla_x p_{i,j,k}\right)^2+\left(\nabla_y p_{i,j,k}\right)^2\right]}, \quad [10]$$

where $p_{i,j,k}$ denotes the pixel of $\mathbf{P}$ with coordinates (i, j, k), and $\mathbf{P}$ is acquired by applying a simple 1-D mean filter [1/3 1/3 1/3] to $\mathbf{F}$ in the spectral dimension, as a coarse-noise attenuated result. We chose the mean filter in the spectral dimension to take advantage of its fast and effective noise attenuation property, and we bypassed its weakness of introducing blurring along the spectral dimension because only the gradients in the spatial dimensions are calculated, as in Equation (10).

For the choice of the weight, a brief qualitative analyzation can be given as follows. Because $FV_{i,j}$ is the vectorial gradient, it will not only be large for the pixels that have large noise intensities but also for those on the edges. Recall that the desired weight should be positively correlated with the magnitude of the noise intensity and negatively correlated with the variation of the signal; we have to distinguish the edges from the smooth areas under the influence of noise. This is achieved by introducing$PV_{i,j}/FV_{i,j}$, because after the smoothing process the vectorial gradient values will decrease much more significantly for the pixels in plain areas than for those on the edges. A simulation experiment will be given in "Experiments and Discussion" in order for the meaning of these terms to be visualized.

Weighted Hyperspectral TV Regularization in the Spectral Dimension

As we know, the noise in an HSI can be viewed from both the spatial dimensions and the spectral dimension, and the spatial denoising process is not sufficient to effectively suppress the noise in the spectral dimension. Therefore, a spectral denoising process is needed to suppress the noise in the spectral dimension and to avoid the artifacts that might have been introduced by the spatial denoising (Othman and Qian 2006; Martín-Herrero 2007). As with the discussion of the weighted hyperspectral TV regularization in the spatial dimensions, the weighted hyperspectral TV regularization in the spectral dimension for HSIs is discussed in this section, and the complementary nature of the spatial and spectral denoising is further discussed in the next section.

When viewed from the spectral perspective, an HSI of size $M \times N \times B$ can be viewed as a set of spectral signatures, or a multivalued 1-D signal. By defining the first-order difference operator in the spectral direction as ∇_z, the traditional TV regularization term in the spectral dimension can be written as:

$$\mathrm{TV}_{\mathrm{Spectral}}(\mathbf{U}) = \sum_{k=1}^{B}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|\nabla_z u_{i,j,k}\right| \quad [11]$$

In an HSI, the noise level and signal variation vary with different spectral signatures (Acito et al. 2011a), especially for the spectral signatures with striping noise in certain bands. Therefore, the denoising processes on every spectral signature should not be separate, and Equation (11) does not meet this requirement. If we view the HSI as a multivalued 1-D signal and apply the vectorial TV norm (Bresson and Chan 2008), the hyperspectral TV regularization term in the spectral dimension is proposed as

$$\mathrm{HTV}_{\mathrm{Spectral}}(\mathbf{U}) = \sum_{k=1}^{B}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\nabla_z u_{i,j,k}\right)^2}. \quad [12]$$

Similar to that in the spatial dimensions, Equation (12) introduces a coupling between the spectral signatures. In fact, Equation (12) can allow the adjustment of the denoising strength for different spectral signatures, due to the property of the vectorial TV norm.

Note that the regularization term (12) can adjust only the penalizing strength for different spectral signatures, rather than the signal in different bands. To overcome this problem, a weight W'_k, which is similar to that in the spatial dimensions, is added to Equation (12), and the weighted hyperspectral TV regularization term in the spectral dimension is defined as

$$\mathrm{WTV}_{\mathrm{Spectral}}(\mathbf{U}) = \sum_{k=1}^{B} W'_k \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\nabla_z u_{i,j,k}\right)^2}. \qquad [13]$$

Similar to the case in the spatial dimensions, the weight W_k is defined as

$$\tau'_k = FV_k \times (1 - QV_k/FV_k)^{\alpha}, \qquad [14]$$

$$W'_k = \frac{\tau'_k}{\bar{\tau}'}, \quad \bar{\tau}' = \sum_{k=1}^{B}\tau'_k \Big/ B, \qquad [15]$$

where FV_k and QV_k are calculated on the 3-D data **F** and **Q** as Equations (16) and (17), respectively.

$$FV_k = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\nabla_z f_{i,j,k}\right)^2}, \qquad [16]$$

$$QV_k = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\nabla_z q_{i,j,k}\right)^2}, \qquad [17]$$

where $q_{i,j,k}$ denotes the pixel of **Q** with coordinates (i, j, k); **Q** is acquired by applying a simple 2-D mean filter [1/9 1/9 1/9; 1/9 1/9 1/9; 1/9 1/9 1/9] to **F** in the spatial dimensions, as a coarse-noise attenuated result. By adopting the weight W'_k, the TV regularization term in the spectral dimension can also automatically adjust the penalizing strength in a 3-D manner, and thus, better results can be expected.

The Combined Spatial and Spectral Weighted Hyperspectral TV Model

An HSI can be denoised with model (2) by the use of 1 of the 2 regularization terms discussed, in the spatial dimensions:

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}} \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{B}(u_{i,j,k} - f_{i,j,k})^2 + \lambda \mathrm{WTV}_{\mathrm{Spatial}}(\mathbf{U}); \qquad [18]$$

or in the spectral dimension:

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}} \frac{1}{2}\sum_{k=1}^{B}\sum_{i=1}^{M}\sum_{j=1}^{N}(u_{i,j,k} - f_{i,j,k})^2 + \lambda \mathrm{WTV}_{\mathrm{Spectral}}(\mathbf{U}). \qquad [19]$$

It is evident that better results can be expected by denoising the HSI with Equations (18) and (19) sequentially, similar to the strategy in Chen and Qian (2011) and Othman and Qian (2006), or by fusing the 2 denoised results produced by using Equations (18) and (19), respectively, similar to the strategy in (Yuan et al. 2014). However, we cannot guarantee that the combination of 2 local optimal results is the global best result. Because Equations (18) and (19) share the same data fidelity term, the prior terms can be easily combined into a unified regularization framework, and the CSSWHTV model is proposed as

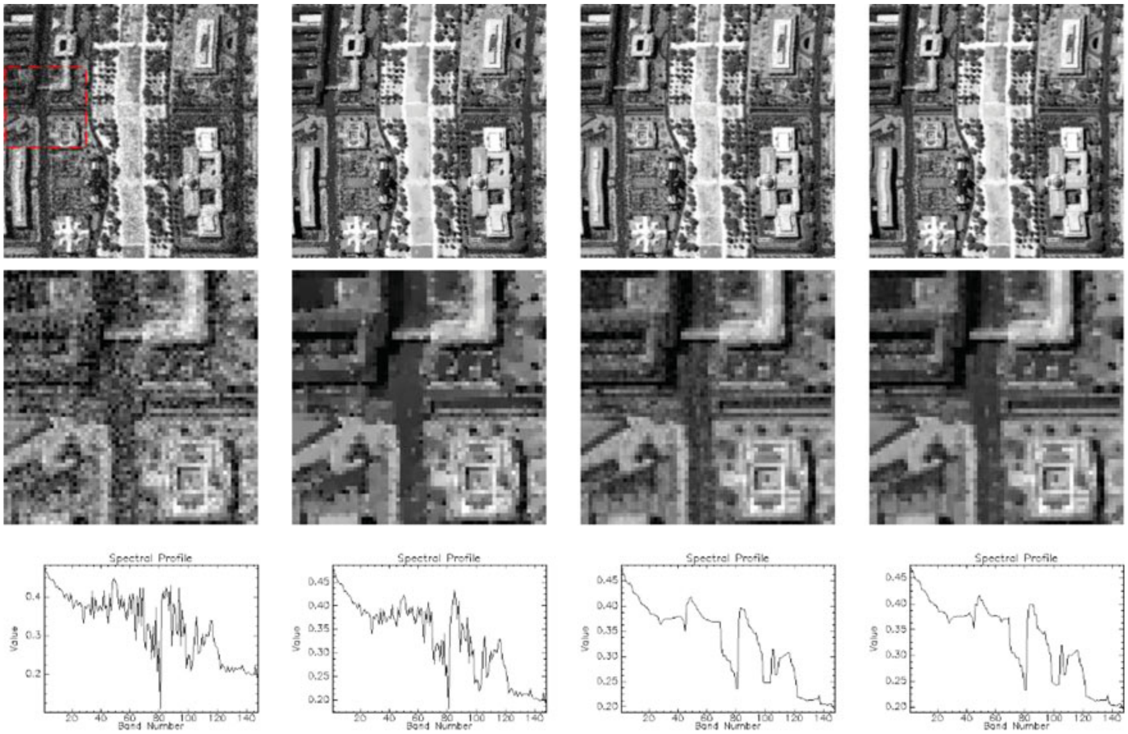
$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}} \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{B}(u_{i,j,k} - f_{i,j,k})^2 + \lambda_1 \sum_{i=1}^{M}\sum_{j=1}^{N} W_{i,j}\sqrt{\sum_{k=1}^{B}\left[\left(\nabla_x u_{i,j,k}\right)^2 + \left(\nabla_y u_{i,j,k}\right)^2\right]} + \lambda_2 \sum_{k=1}^{B} W'_k \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\nabla_z u_{i,j,k}\right)^2}, \qquad [20]$$

where λ_1 and λ_2 are the regularization parameters. With the 2 regularization terms, the noise in the HSI can be removed from both the spatial and spectral dimensions. By combining the 2 regularization terms, which are complementary in the spatial and spectral perspectives, into a unified framework, the TV model is effectively extended to a 3-D TV model, which is suitable for 3-D HSI data.

In the following, we present an example to help with the understanding of the complementary nature of the 2 regularization terms. In Figure 2, with the Washington DC Mall dataset as an example, the denoising results using models (18), (19), and (20) are presented, respectively. The data we used were the same as the data used in "Experiments and Discussion," and the σ Equation (34) was set as 0.4. Note that for all the methods, the regularization parameters were adjusted to achieve the highest SNR. The first row represents the denoised band 100 (1569.46 nm), the magnified local results in the red rectangle area are shown in the second row, and the third row shows the spectral curves of pixel (153, 124). From the results shown in Figure 2, we can see that the noise in the spatial dimensions can be effectively removed by the use of the model in (18), but there is still noise in the spectral dimension. Conversely, the noise in the spectral dimension can be effectively removed by the use of the model in (19), but there is still noise remaining in the spatial dimensions. However, by using the combined model in (20), not only is the noise in each dimension effectively removed, but the oversmoothing problem in both the spatial dimensions and the spectral dimension is alleviated.

NUMERICAL SOLUTION

HSIs are much larger than gray images, so algorithms need to be efficient. The proposed model in Equation (20) has 2 regularization terms and is difficult to solve by the use of gradient methods. Luckily, in recent years, the ADMM algorithm has proved to be able to effectively and efficiently solve many constrained and unconstrained problems in the image processing



(a) (b) (c) (d)

FIG. 2. Comparisons of the denoising results using different hyperspectral TV models. (a) The noisy data, (b) the result using model (18), (c) the result using model (19), and (d) the result using model (20).

area, e.g., imaging inverse problems (Afonso et al. 2011) and unmixing (Bioucas-Dias and Figueiredo 2010; Iordache et al. 2012, 2014; Zhao et al. 2013). In this section, we briefly review the ADMM algorithm and then describe the version of ADMM used to solve the CSSWHTV model.

ADMM Algorithm

ADMM benefits by decomposing a difficult problem into a sequence of simple ones. The nature of ADMM is to use a variable splitting procedure followed by the adoption of an augmented Lagrangian method. Consider the following unconstrained problem:

$$\min_{\mathbf{x}} f_1(\mathbf{x}) + f_2(\mathbf{G}\mathbf{x}) \qquad [21]$$

where $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{G} \in \mathbf{R}^{p\times n}$, $f_1 : \mathbf{R}^n \to \bar{\mathbf{R}}$, and $f_2 : \mathbf{R}^p \to \bar{\mathbf{R}}$. By creating a new variable,$\mathbf{v} \in \mathbf{R}^n$, Equation (21) is equivalent to:

$$\min_{\mathbf{x},\mathbf{v}} f_1(\mathbf{x}) + f_2(\mathbf{v}), \text{ subject to } \mathbf{v} = \mathbf{G}\mathbf{x}. \qquad [22]$$

Equation (22) is a variable splitting procedure, and then an augmented Lagrangian method and a nonlinear block Gauss–Seidel method are used to derive the ADMM algorithm (for more details, see Afonso et al. 2011). The ADMM algorithm used to solve Equation (21) is shown in Algorithm 1, and the convergence of the ADMM algorithm is briefly given by Theorem 1.

Theorem 1. (Eckstein–Bertsekas 1992). Considering the problem in Equation (21), suppose $\mathbf{G}$ *has a full column rank, and* f_1 *and* f_2 *are closed, proper, convex functions. For any* $\mu > 0$, *if Equation (21) has a solution, then the sequence* $\{\mathbf{x}^K\}$ *converges to it. If Equation (21) does not have a solution, then at least 1 of the 2 sequences,* $\{\mathbf{v}^K\}$ or $\{\mathbf{d}^K\}$, *diverges.*

Algorithm 1. Pseudocode of the ADMM algorithm to solve Equation (21)

```
Initialization: set $K = 0$, choose $\mu > 0$, $\mathbf{v}^0$, and $\mathbf{d}^0$
repeat:
    $\mathbf{x}^{K+1} \leftarrow \arg\min_{\mathbf{x}} f_1(\mathbf{x}) + \frac{\mu}{2}\left\|\mathbf{G}\mathbf{x} - \mathbf{v}^K - \mathbf{d}^K\right\|_2^2$
    $\mathbf{v}^{K+1} \leftarrow \arg\min_{\mathbf{v}} f_2(\mathbf{v}) + \frac{\mu}{2}\left\|\mathbf{G}\mathbf{x}^{K+1} - \mathbf{v} - \mathbf{d}^K\right\|_2^2$
    Update the Lagrange multipliers:
    $\mathbf{d}^{K+1} \leftarrow \mathbf{d}^K - (\mathbf{G}\mathbf{x}^{K+1} - \mathbf{v}^{K+1})$
    Update the iteration: $K \leftarrow K + 1$
until stop criterion is satisfied.
```

Because many terms need to be used in this section, a complete list of terms and definitions is provided in Table 1.

TABLE 1
List of terms

Term (3-D data, vector, number)	Definition
$\mathbf{F}$,$\mathbf{f}$,$f_{i,j,k}$	The noisy data
$\mathbf{U}$,$\mathbf{u}$,$u_{i,j,k}$	The noise-free data, and $\mathbf{u}$ corresponds to $\mathbf{x}$ in Algorithm 1
$\mathbf{V}_x$,$\mathbf{v}_x$,$v_{x,(i,j,k)}$	The added variable that plays the role of the first-order difference of the noise-free data in the $x$direction
$\mathbf{V}_y$,$\mathbf{v}_y$,$v_{y,(i,j,k)}$	The added variable that plays the role of the first-order difference of the noise-free data in the $y$direction
$\mathbf{V}_z$,$\mathbf{v}_z$,$v_{z,(i,j,k)}$	The added variable that plays the role of the first-order difference of the noise-free data in the $z$direction
$\mathbf{D}_x$,$\mathbf{d}_x$,$d_{x,(i,j,k)}$	The Lagrange multiplier in the $x$direction
$\mathbf{D}_y$,$\mathbf{d}_x$,$d_{x,(i,j,k)}$	The Lagrange multiplier in the $y$direction
$\mathbf{D}_x$,$\mathbf{d}_x$,$d_{x,(i,j,k)}$	The Lagrange multiplier in the $z$direction
$W_{i,j}$	The weighting matrix in the spatial dimensions
$W\prime_k$	The weighting vector in the spectral dimension
subscripts $(i, j, :)$	Denote the vector along the spectral dimension at the spatial position (i, j) of the corresponding 3-D data
subscripts $(:, :, k)$	Denote the lexicographically ordered vector of the kth band of the corresponding 3-D data
superscript K	Denotes the corresponding variable in the Kth iteration
vector $\mathbf{v} \equiv \begin{bmatrix}\mathbf{v}_x\\ \mathbf{v}_y\\ \mathbf{v}_z\end{bmatrix}$	Corresponds to $\mathbf{v}$ in Algorithm 1
vector $\mathbf{d} \equiv \begin{bmatrix}\mathbf{d}_x\\ \mathbf{d}_y\\ \mathbf{d}_z\end{bmatrix}$	Corresponds to $\mathbf{d}$ in Algorithm 1
matrix $\mathbf{G} \equiv \begin{bmatrix}\nabla_x\\ \nabla_y\\ \nabla_z\end{bmatrix}$	The gradient operator that corresponds to $\mathbf{G}$ in Algorithm 1

By using variable splitting, Equation (20) is reformulated as:

$$\begin{aligned}\hat{\mathbf{U}} = \arg\min_{\mathbf{U}} & \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{B}(u_{i,j,k}-f_{i,j,k})^2 \\ &+\lambda_1\sum_{i=1}^{M}\sum_{j=1}^{N}W_{i,j}\sqrt{\sum_{k=1}^{B}\left[\left(v_{x,(i,j,k)}\right)^2+\left(v_{y,(i,j,k)}\right)^2\right]} \\ &+\lambda_2\sum_{k=1}^{B}W'_k\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(v_{z,(i,j,k)}\right)^2} \\ &\text{subject to } \nabla_x u_{i,j,k} = v_{x,(i,j,k)}, \\ \nabla_y u_{i,j,k} = v_{y,(i,j,k)},\ & \nabla_z u_{i,j,k} = v_{z,(i,j,k)}. \end{aligned} \qquad [23]$$

For problem (23), Step 3 in Algorithm 1 becomes Equation (24), Step 4 in Algorithm 1 becomes Equations (25) and (26), and Step 5 in Algorithm 1 becomes Equation (27), as follows:

$$\begin{aligned}\mathbf{u}^{K+1} = \arg\min_{\mathbf{u}} & \frac{1}{2}\|\mathbf{u}-\mathbf{f}\|_2^2 + \frac{\mu}{2}\left\|\nabla_x\mathbf{u}-\mathbf{v}_x^K-\mathbf{d}_x^K\right\|_2^2 \\ &+\frac{\mu}{2}\left\|\nabla_y\mathbf{u}-\mathbf{v}_y^K-\mathbf{d}_y^K\right\|_2^2+\frac{\mu}{2}\left\|\nabla_z\mathbf{u}-\mathbf{v}_z^K-\mathbf{d}_z^K\right\|_2^2,\end{aligned} \qquad [24]$$

$$\begin{aligned}(\mathbf{V}_x^{K+1},\mathbf{V}_y^{K+1}) = & \arg\min_{\mathbf{V}_x,\mathbf{V}_y} \\ &\left\{\begin{array}{l}\frac{\mu}{2}\left\|\nabla_x\mathbf{u}^{K+1}-\mathbf{v}_x-\mathbf{d}_x^K\right\|_2^2+\frac{\mu}{2}\left\|\nabla_y\mathbf{u}^{K+1}-\mathbf{v}_y-\mathbf{d}_y^K\right\|_2^2 \\ +\lambda_1\sum_{i=1}^{M}\sum_{j=1}^{N}W_{i,j}\sqrt{\sum_{k=1}^{B}\left[\left(v_{x,(i,j,k)}\right)^2+\left(v_{y,(i,j,k)}\right)^2\right]}\end{array}\right\},\end{aligned} \qquad [25]$$

$$\begin{aligned}\mathbf{V}_z^{K+1} = \arg\min_{\mathbf{V}_z} & \left\{\frac{\mu}{2}\left\|\nabla_z\mathbf{u}^{K+1}-\mathbf{v}_z-\mathbf{d}_z^K\right\|_2^2 \right. \\ &\left. +\lambda_2\sum_{k=1}^{B}W'_k\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(v_{z,(i,j,k)}\right)^2}\right\},\end{aligned} \qquad [26]$$

$$\begin{aligned}\mathbf{D}_x^{K+1} &= \mathbf{D}_x^K-(\nabla_x\mathbf{U}^{K+1}-\mathbf{V}_x^{K+1}), \\ \mathbf{D}_y^{K+1} &= \mathbf{D}_y^K-(\nabla_y\mathbf{U}^{K+1}-\mathbf{V}_y^{K+1}), \\ \mathbf{D}_z^{K+1} &= \mathbf{D}_z^K-(\nabla_z\mathbf{U}^{K+1}-\mathbf{V}_z^{K+1}).\end{aligned} \qquad [27]$$

Based on the Euler–Lagrange equation, the optimal $\mathbf{u}$ of Equation (24) can be obtained by solving the following equation:

$$\begin{aligned}&(\mathbf{I}+\mu(\nabla_x^T\nabla_x+\nabla_y^T\nabla_y+\nabla_z^T\nabla_z))\mathbf{u}^{K+1} \\ &\quad=\mathbf{f}+\mu\nabla_x^T(\mathbf{v}_x^K+\mathbf{d}_x^K)+\mu\nabla_y^T(\mathbf{v}_y^K+\mathbf{d}_y^K)+\mu\nabla_z^T(\mathbf{v}_z^K+\mathbf{d}_z^K).\end{aligned} \qquad [28]$$

Because the system is strictly diagonally dominant, the most natural choice is the Gauss–Seidel method. Similar to the method in Goldstein and Osher (2009), the numerical solution to this problem can be written component-wise as

$$
\begin{aligned}
u_{i,j,k}^{K+1} = \frac{\mu}{1+6\mu}(&u_{i-1,j,k}^{K} + u_{i+1,j,k}^{K} + u_{i,j-1,k}^{K} \\
&+u_{i,j+1,k}^{K} + u_{i,j,k-1}^{K} + u_{i,j,k+1}^{K} + v_{x,(i-1,j,k)}^{K} \\
&-v_{x,(i,j,k)}^{K} + d_{x,(i-1,j,k)}^{K} - d_{x,(i,j,k)}^{K} + v_{y,(i,j-1,k)}^{K} \\
&-v_{y,(i,j,k)}^{K} + d_{y,(i,j-1,k)}^{K} - d_{y,(i,j,k)}^{K} + v_{z,(i,j,k-1)}^{K} \\
&-v_{z,(i,j,k)}^{K} + d_{z,(i,j,k-1)}^{K} - d_{z,(i,j,k)}^{K}) + \frac{1}{1+6\mu} f_{i,j,k}.
\end{aligned}
\qquad [29]
$$

The solution Equations (25) and (26) can be obtained by the well-known vect-soft threshold. For a vector $\mathbf{a}$ and threshold τ, the vect-soft threshold is defined as:

$$
\text{vect-soft}\,\{\mathbf{a}, \tau\} = \left\{ \frac{\max(\|\mathbf{a}\|_2 - \tau, 0)}{\max(\|\mathbf{a}\|_2 - \tau, 0) + \tau} \cdot \mathbf{a} \right\}. \qquad [30]
$$

Letting $\mathbf{v}_{x,(i,j,:)}$, $\mathbf{v}_{y,(i,j,:)}$, $(\nabla_x \mathbf{u} - \mathbf{d}_x)_{(i,j,:)}$, and $(\nabla_y \mathbf{u} - \mathbf{d}_y)_{(i,j,:)}$ denote the vectors along the spectral direction at spatial position (i, j) of $\mathbf{V}_x$, $\mathbf{V}_y$, $(\nabla_x \mathbf{U} - \mathbf{D}_x)$, and $(\nabla_y \mathbf{U} - \mathbf{D}_y)$, respectively, the solution of Equation (25) is given by

$$
\begin{bmatrix} \mathbf{v}_{x,(i,j,:)}^{K+1} \\ \mathbf{v}_{y,(i,j,:)}^{K+1} \end{bmatrix} = \text{vect-soft} \left\{ \begin{bmatrix} (\nabla_x \mathbf{u}^{K+1} - \mathbf{d}_x^K)_{(i,j,:)} \\ (\nabla_y \mathbf{u}^{K+1} - \mathbf{d}_y^K)_{(i,j,:)} \end{bmatrix}, \frac{\lambda_1 W_{i,j}}{\mu} \right\}. \qquad [31]
$$

Letting $\mathbf{v}_{z,(:,:,k)}$ and $(\nabla_z \mathbf{u} - \mathbf{d}_z)_{(:,:,k)}$ denote the lexicographically ordered column vectors of the kth band of $\mathbf{V}_z$ and $(\nabla_z \mathbf{U} - \mathbf{D}_z)$, respectively, the solution of Equation (26) is presented as

$$
\mathbf{v}_{z,(:,:,k)}^{K+1} = \text{vect-soft} \left\{ (\nabla_z \mathbf{u}^{K+1} - \mathbf{d}_z^K)_{(:,:,k)}, \frac{\lambda_2 W'_k}{\mu} \right\}. \qquad [32]
$$

Finally, the proposed algorithm is summarized in Algorithm 2. The convergence of the proposed algorithm is guaranteed by Theorem 1 because it is an instance of ADMM with:

$$
\mathbf{G} \equiv \begin{bmatrix} \nabla_x \\ \nabla_y \\ \nabla_z \end{bmatrix}, \qquad [33]
$$

which is a full column rank matrix, and Equations (24)–(26) are closed, proper, convex functions.

Algorithm 2. Pseudocode of the proposed algorithm to solve the CSSWHTV model in Equation (20)

```
1. Initialization: set $K = 0$, choose $\mu > 0$, $\mathbf{U}^0 = \mathbf{F}$, $\mathbf{V}_x^0 = \nabla_x \mathbf{F}$, $\mathbf{V}_y^0 = \nabla_y \mathbf{F}$, $\mathbf{V}_z^0 = \nabla_z \mathbf{F}$, $\mathbf{D}_x^0 = \mathbf{0}$, $\mathbf{D}_y^0 = \mathbf{0}$, and $\mathbf{D}_z^0 = \mathbf{0}$
2. repeat:
    3. Get $\mathbf{U}^{K+1}$ by calculating
    $u_{i,j,k}^{K+1} = \frac{\mu}{1+6\mu}(u_{i-1,j,k}^{K} + u_{i+1,j,k}^{K} + u_{i,j-1,k}^{K} + u_{i,j+1,k}^{K} + u_{i,j,k-1}^{K} + u_{i,j,k+1}^{K} +$
    $v_{x,(i-1,j,k)}^{K} - v_{x,(i,j,k)}^{K} + d_{x,(i-1,j,k)}^{K} - d_{x,(i,j,k)}^{K} + v_{y,(i,j-1,k)}^{K} - v_{y,(i,j,k)}^{K} + d_{y,(i,j-1,k)}^{K} - d_{y,(i,j,k)}^{K}$
    $+v_{z,(i,j,k-1)}^{K} - v_{z,(i,j,k)}^{K} + d_{z,(i,j,k-1)}^{K} - d_{z,(i,j,k)}^{K}) + \frac{1}{1+6\mu} f_{i,j,k}$
    for every $(i, j, k)$
    4. Get $\mathbf{V}_x^{K+1}$ and $\mathbf{V}_y^{K+1}$ by calculating
    $\begin{bmatrix} \mathbf{v}_{x,(i,j,:)}^{K+1} \\ \mathbf{v}_{y,(i,j,:)}^{K+1} \end{bmatrix} = \text{vect-soft} \left\{ \begin{bmatrix} (\nabla_x \mathbf{u}^{K+1} - \mathbf{d}_x^K)_{(i,j,:)} \\ (\nabla_y \mathbf{u}^{K+1} - \mathbf{d}_y^K)_{(i,j,:)} \end{bmatrix}, \frac{\lambda_1 W_{i,j}}{\mu} \right\}$ for every $(i, j)$
    5. Get $\mathbf{V}_z^{K+1}$ by calculating
    $\mathbf{v}_{z,(:,:,k)}^{K+1} = \text{vect-soft} \left\{ (\nabla_z \mathbf{u}^{K+1} - \mathbf{d}_z^K)_{(:,:,k)}, \frac{\lambda_2 W'_k}{\mu} \right\}$ for every $k$
    6. Update the Lagrange multipliers:
    $\mathbf{D}_x^{K+1} = \mathbf{D}_x^K - (\nabla_x \mathbf{U}^{K+1} - \mathbf{V}_x^{K+1})$
    $\mathbf{D}_y^{K+1} = \mathbf{D}_y^K - (\nabla_y \mathbf{U}^{K+1} - \mathbf{V}_y^{K+1})$
    $\mathbf{D}_z^{K+1} = \mathbf{D}_z^K - (\nabla_z \mathbf{U}^{K+1} - \mathbf{V}_z^{K+1})$
    7. Update the iteration: $K \leftarrow K + 1$
8. until $\sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{B} (u_{i,j,k}^{K} - u_{i,j,k}^{K-1})^2 < e$.
```

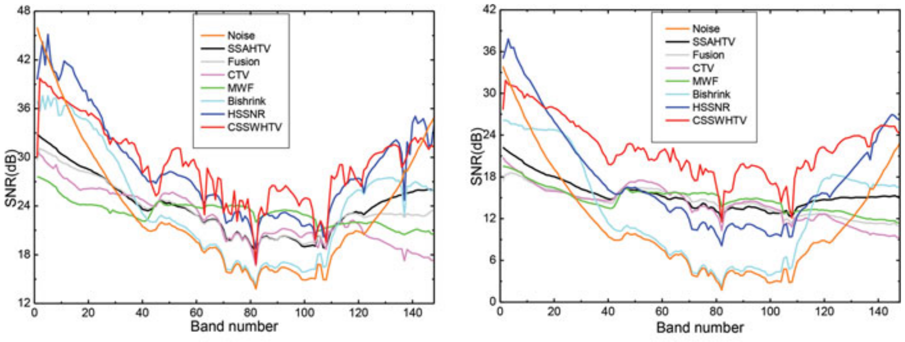
TABLE 2

Performance of the denoising methods for different noise levels on a portion of the Washington DC Mall image

σ	Index	Noise	SSAHTV	Fusion	CTV	MWF	BiShr.	HSSNR	CSSWHTV
0.1	SNR	31.75	32.69	32.60	32.81	32.49	32.22	**36.35**	34.71
	σSNR	0.0035	0.0030	0.0035	0.0039	0.0206	0.0059	0.0059	0.0048
	MSA	1.94	1.62	1.69	1.61	1.21	1.73	**1.10**	1.29
	σMSA	0.0009	0.0008	0.0008	0.0009	0.0024	0.0013	0.0009	0.0010
	PARMs	—	1/41	1/41,1/42	1/360,100	195,195,15	—	—	1/120,1
0.2	SNR	25.74	27.59	27.51	27.71	28.40	26.42	**31.50**	30.84
	σSNR	0.0022	0.0033	0.0030	0.0043	0.0212	0.0031	0.0092	0.0063
	MSA	3.88	2.75	2.92	2.79	**1.83**	3.52	1.95	1.94
	σMSA	0.0012	0.0013	0.0014	0.0019	0.0035	0.0015	0.0020	0.0018
	PARMs	—	1/14	1/14,1/14	1/140,100	195,195,15	—	—	1/46,2
0.3	SNR	22.21	24.84	24.78	24.87	25.70	22.91	28.37	**28.72**
	σSNR	0.0026	0.0033	0.0029	0.0024	0.0260	0.0027	0.0106	0.0044
	MSA	5.79	3.64	3.82	3.72	**2.34**	5.30	2.80	2.44
	σMSA	0.0023	0.0020	0.0020	0.0022	0.0056	0.0023	0.0041	0.0015
	PARMs	—	1/8	1/8,1/7	1/80,100	195,195,10	—	—	1/25,4
0.4	SNR	19.71	22.97	22.89	22.93	23.75	20.41	26.06	**27.21**
	σSNR	0.0018	0.0042	0.0033	0.0034	0.0153	0.0033	0.0053	0.0052
	MSA	7.69	4.48	4.70	4.55	**2.83**	7.05	3.6527	2.86
	σMSA	0.0019	0.0019	0.0013	0.0020	0.0044	0.0029	0.0020	0.0021
	PARMs	—	1/6	1/6,1/5	1/56,100	195,195,10	—	—	1/18,5
0.6	SNR	16.19	20.49	20.55	20.34	21.12	16.91	22.70	**25.16**
	σSNR	0.0027	0.0045	0.0036	0.0037	0.0104	0.0040	0.0093	0.0047
	MSA	11.37	5.34	5.72	5.86	3.68	10.44	5.36	**3.52**
	σMSA	0.0041	0.0026	0.0035	0.0023	0.0045	0.0048	0.0067	0.0035
	PARMs	—	1/3	1/3,1/3	1/33,100	195,195,10	—	—	1/12,8
0.8	SNR	13.69	18.76	18.95	18.62	19.36	14.45	20.28	**23.83**
	σSNR	0.0047	0.0049	0.0046	0.0039	0.0177	0.0051	0.0085	0.0074
	MSA	14.91	6.06	6.48	6.76	4.40	13.66	7.06	**3.99**
	σMSA	0.0098	0.0053	0.0050	0.0049	0.0070	0.0093	0.0089	0.0051
	PARMs	—	1/2	1/2,1/2	1/22,100	195,195,10	—	—	1/11,14
1.2	SNR	10.17	16.26	16.64	16.42	16.97	10.97	16.82	**22.06**
	σSNR	0.0032	0.0060	0.0040	0.0045	0.0125	0.0035	0.0090	0.0136
	MSA	21.43	9.64	8.64	8.13	5.72	19.54	10.41	**4.80**
	σMSA	0.0065	0.0069	0.0060	0.0073	0.0080	0.0087	0.0114	0.0076
	PARMs	—	1/2	1/2,1	1/13,100	195,195,10	—	—	1/8,24
1.6	SNR	7.67	15.36	15.16	15.02	15.35	8.52	14.34	**20.72**
	σSNR	0.0042	0.0036	0.0057	0.0043	0.0102	0.0046	0.0080	0.0137
	MSA	27.21	8.56	7.92	9.07	6.94	24.97	13.65	**5.52**
	σMSA	0.0109	0.0064	0.0055	0.0065	0.0068	0.0103	0.0139	0.0088
	PARMs	—	1	1,2	1/9,100	195,195,5	—	—	1/6,31

Computational Complexity Analysis

Here, we analyze the complexity of the proposed Algorithm 2. For an HSI of size $M \times N \times B$, the computational time of Step 3 is $O(21MNB)$. For a vector of length$n$, the vect-soft threshold of the vector requires $O(3n+4)$ time. Therefore, the complexity of Step 4 is$O(MN(10B+4))$, and the complexity of Step 5 is $O(B(5MN+4))$. Finally, Step 6 requires $O(6MNB)$ time. This leads to a total complexity of $O(K(42MNB+4MN+4B))$.



$\sigma = 0.4$ $\sigma = 1.6$

FIG. 3. SNR versus band number for the different methods on a portion of the Washington DC Mall image.

Experiments and Discussion

To validate the effectiveness of the proposed method, we performed both simulated experiments and real data experiments. The proposed method was compared with the method in Yuan et al. (2012) named SSAHTV, the method in Yuan et al. (2014) named Fusion, the method in Zhang (2012) named cubic total variation (CTV), the multidimensional Wiener filtering (MWF) method in Muti and Bourennane (2007), the hybrid spatial–spectral noise-reduction (HSSNR) method in Othman and Qian (2006), and the method called BiShrink, which uses bivariate wavelet thresholding and a complex 3-D dual-tree wavelet transform (Chen et al. 2011). The HSSNR and BiShrink methods are both parameter free. For the HSSNR method, third-order Coiflet wavelets and BayesShrink were used. For the proposed method, λ_1 was chosen from [1/150, 1/140,. . .,1/110, 1/100, 1/99,. . .,1], and λ_2 was chosen from [1/5, 1/4, 1/3, 1/2,1, 2, 3,. . .,59, 60], for all the datasets.

Simulated Experiments

For the simulated experiments, we first added Gaussian noise to the original HSI, and the different methods were then performed on the noisy HSI. In this way, the original HSI can be used to quantitatively assess the denoising results. A clear part of the widely used Hyperspectral Digital Imagery Collection Experiment (HYDICE) image of the Washington DC Mall of size 200 × 200 × 148 was used for the simulated experiments. The original data of size 1280 × 307 contain 191 bands after removing the water absorption bands. A rectangular region from

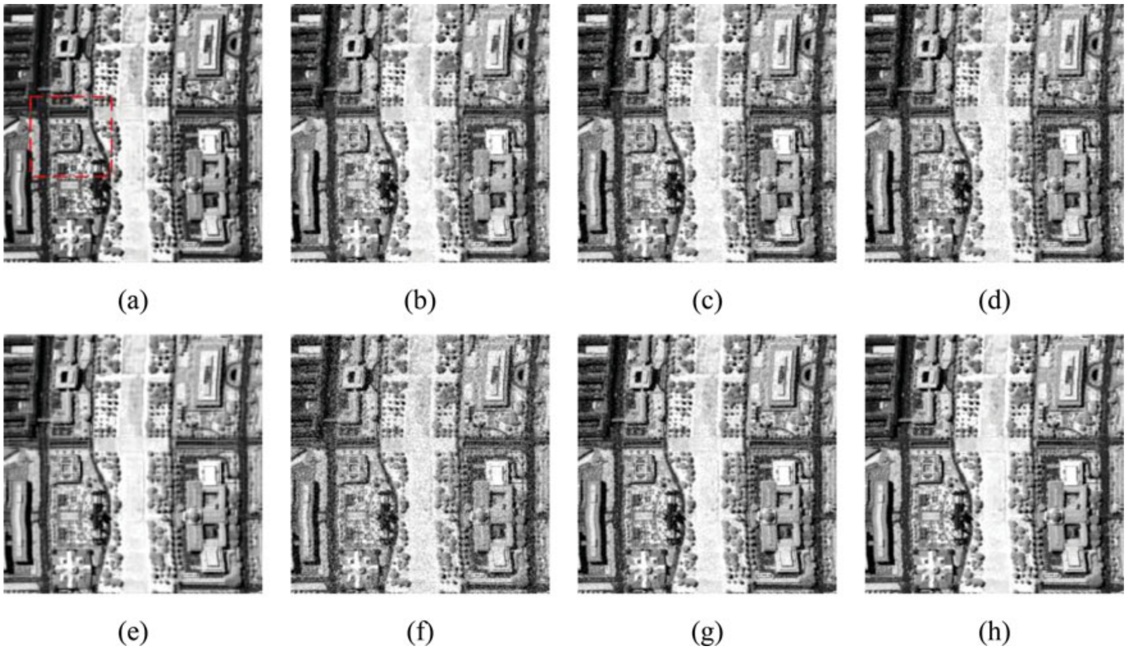


FIG. 4. Denoising results in the simulated experiment when $\sigma = 0.4$: (a) original band 74 (1174.65 nm), (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

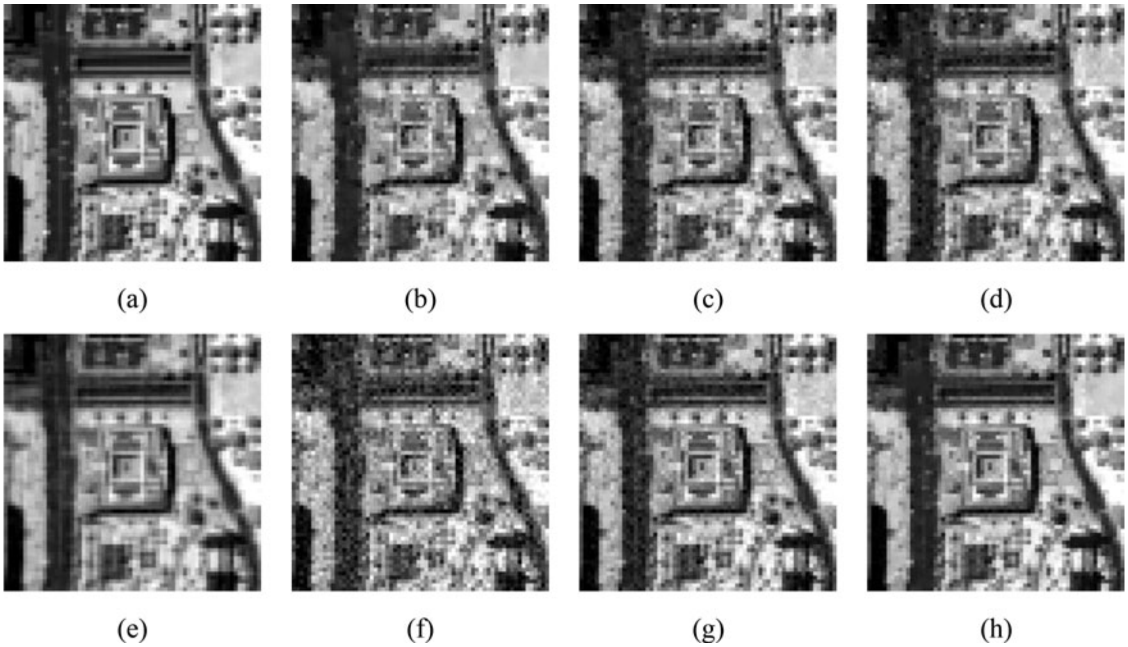


FIG. 5. Magnified results in the red rectangle of Figure 4: (a) original image, (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

line 567 to 766, column 71 to 270, was cropped from the original data for use in the experiments. To get rid of the disturbance introduced by the initial noise, the bands of 1∼10, 19, 85∼89, 101∼108, 132∼142, and 187∼191, which contain visually detectable noise, were also discarded. The gray values were then normalized between [0, 1] before adding the noise. For every band, the noise was additive Gaussian noise with zero-mean, and the noise variance was varied along the spectral dimension

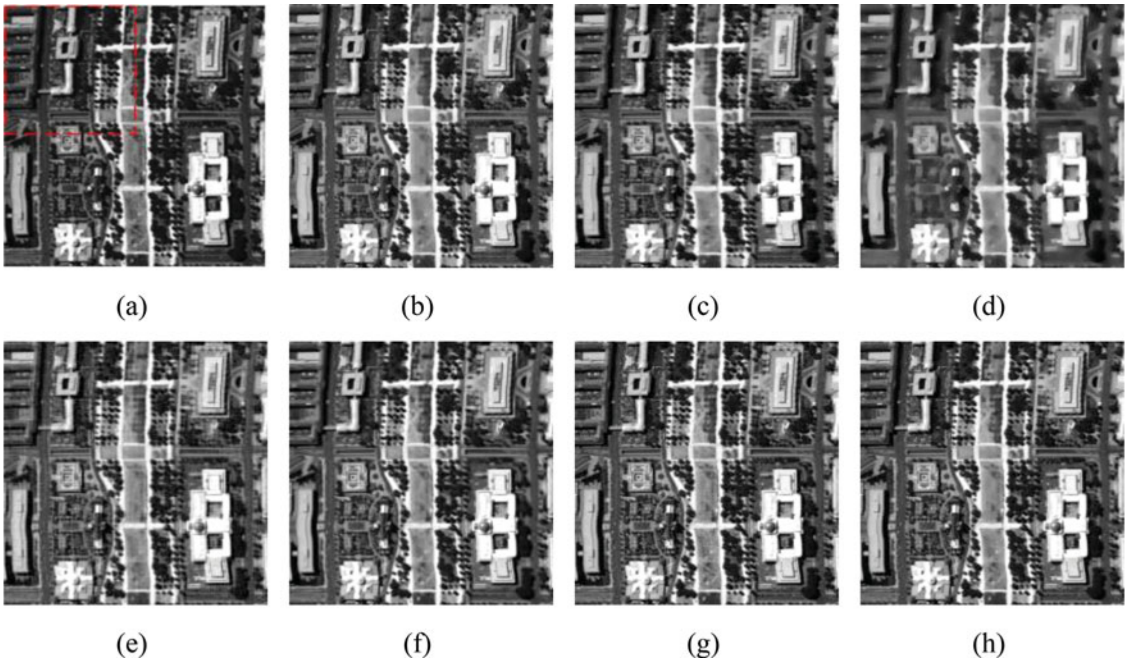


FIG. 6. Denoising results in the simulated experiment when $\sigma = 0.4$: (a) original band 146 (2414.49 nm), (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

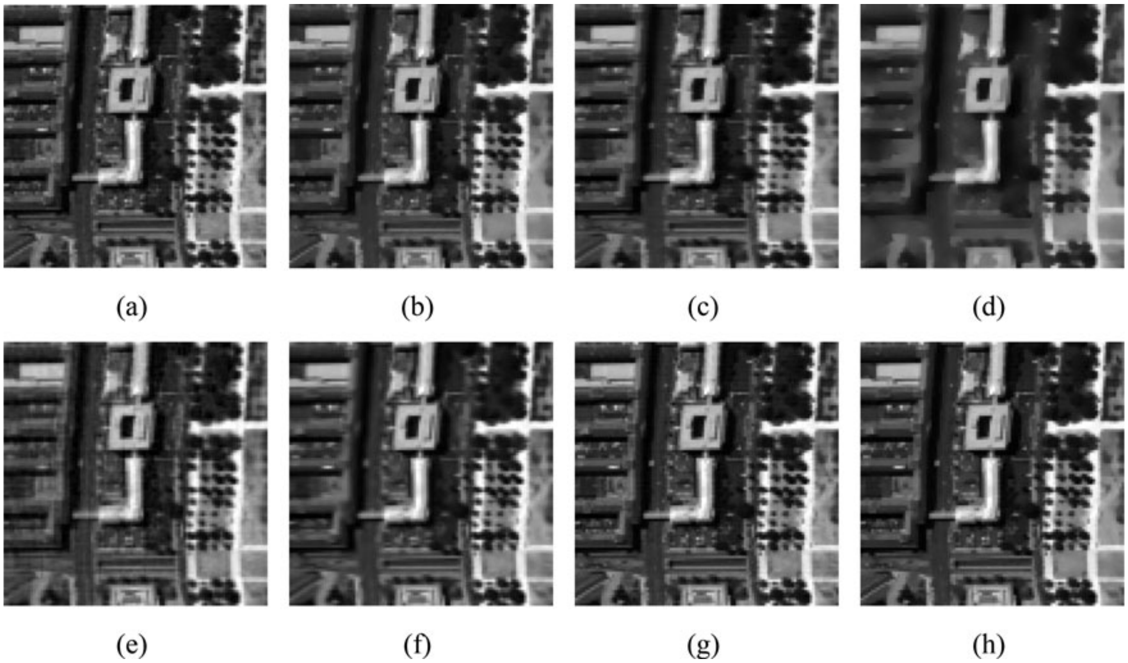


FIG. 7. Magnified results in the red rectangle of Figure 6: (a) original image, (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

(σ_k^2) as a Gaussian shape centered at the middle band $(B/2)$ as

$$\sigma_k^2 = \sigma^2 \cdot e^{-\frac{(k-B/2)^2}{2\eta^2}} \Big/ \sum\nolimits_{l=1}^{B} e^{-\frac{(l-B/2)^2}{2\eta^2}}, \qquad [34]$$

where the power of the noise is controlled by σ, and η controls the Gaussian curve's shape. In this study, η was set as 20. We should note that this setting does not equal a real case and is not critical for the design of the proposed method, but was used only to test the proposed method.

In addition to the visual evaluation, we used the SNR index and the mean spectral angle (MSA) index to quantitatively evaluate the quality of the results. For the different methods, the parameters were set to achieve the best values of the SNR index, and the used parameter values are listed sequentially in the tables.

Let $\mathbf{X}_j^i$ and $\mathbf{Y}_j^i$ denote the ith pixel in the jth band of the original HSI $\mathbf{X}$ and the denoised result $\mathbf{Y}$, respectively. The SNR index in decibels is then defined as

$$\mathrm{SNR} = 10\log_{10}\left[\sum_{j=1}^{B}\sum_{i=1}^{N}(\mathbf{X}_j^i)^2 \Big/ \sum_{j=1}^{B}\sum_{i=1}^{N}(\mathbf{X}_j^i - \mathbf{Y}_j^i)^2\right], \qquad [35]$$

where N is the total number of pixels in 1 band of the HSI.

The MSA index is the mean of the spectral angle (SA), which measures the absolute angle between the spectral vectors of the reference and denoised images. It reflects the spectral distortion and is defined as

$$\mathbf{MSA} = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left|\arccos\frac{\sum_{j=1}^{B}\mathbf{X}_j^i\mathbf{Y}_j^i}{\sqrt{\sum_{j=1}^{B}(\mathbf{X}_j^i)^2\sum_{j=1}^{B}(\mathbf{Y}_j^i)^2}}\right|. \qquad [36]$$

The results of the different methods when σ in Equation (34) was varied from 0.1 to 1.6 are shown in Table 2. In the table, for different values ofσ, all the experiments were repeated 10 times. The mean values of the SNR and MSA were calculated, and the standard deviations of the SNR and MSA are written as $\sigma\, SNR$ and$\sigma\, MSA$, respectively. The best results for the SNR and MSA are marked in bold and the second best are underlined. The results indicate that CSSWHTV outperforms the other methods in most cases. For the case when the noisy HSI already has a high SNR ($\sigma = 0.1$), i.e., the noise in the HSI is generally not visually detectable, HSSNR performs better than CSSWHTV due to its effective preprocessing step of elevating the noise. For the other cases, when the HSI contains moderate or high levels of noise, the results indicate a better denoising capability for CSSWHTV. Moreover, CSSWHTV is a relatively pure denoising method when compared to HSSNR, which means there is a greater potential for improvement.

To show the stability of all the methods in different bands, the SNR was calculated separately for each band. To save space, the results are shown in Figure 3 for the cases when the HSI contained moderate noise ($\sigma = 0.4$) and heavy noise ($\sigma = 1.6$). Note that the spectral band numbers in Figure 3 and hereafter are not the numbers in the original data, but the recoded numbers after discarding the noisy bands. As shown in Figure 3,

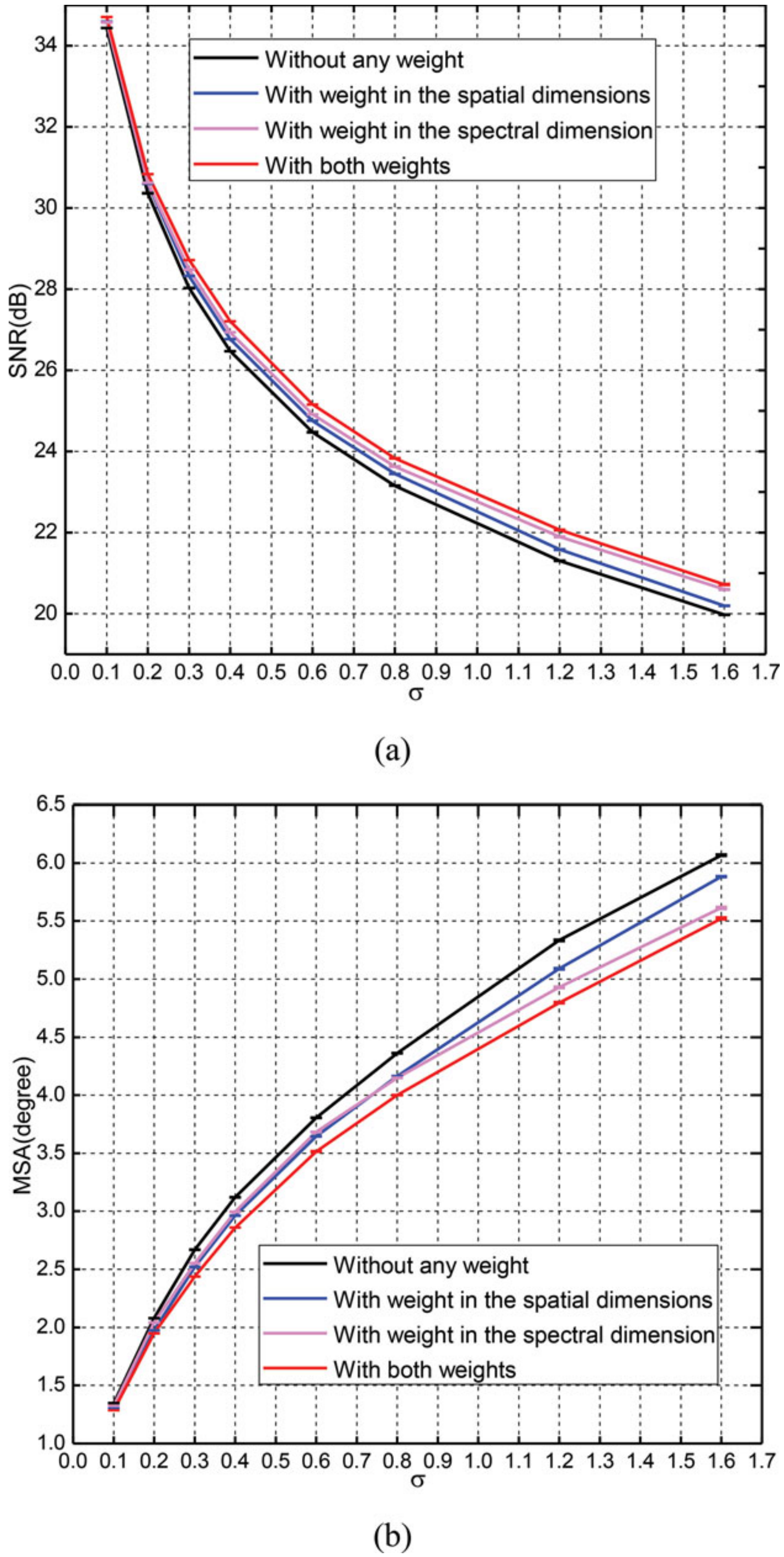


FIG. 8. Comparison of the SNR and MSA values of the methods adopting different combinations of the weights for different levels of noise: (a) SNR, (b) MSA.

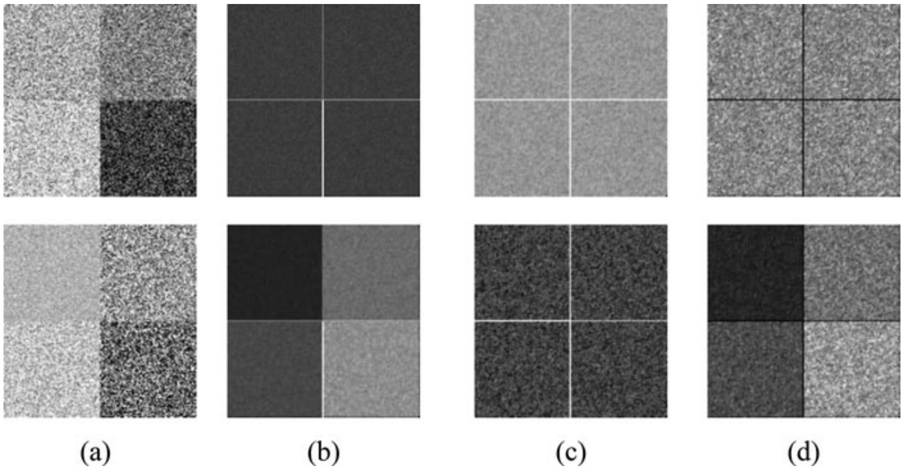


FIG. 9. The experimental results using a piecewise constant HSI. The first row shows the results under Case 1, and the second row shows the results under Case 2: (a) band 210 (988.00 nm) of the noisy HSI, (b) $FV_{i,j}$, (c) $PV_{i,j}/FV_{i,j}$, and (d) the proposed weight.

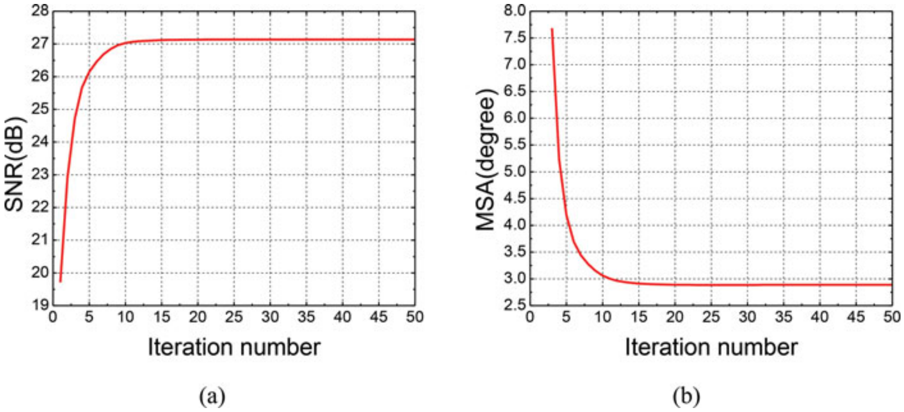


FIG. 10. The results using the CSSWHTV method with different iteration numbers: (a) SNR, (b) MSA.

CSSWHTV is robust for all the bands and shows an obvious advantage over the other methods in the bands with heavy noise. For a more thorough evaluation of the different methods, the results and the magnified results of band 74, which contained relatively heavy noise, and those of band 146, which contained a relatively low level of noise, are shown in Figures 4–7, respectively. From the results, it is clear that the results of the proposed CSSWHTV method look sharper and clearer, and it appears to be the best method because it can remove most of the noise and maintain the details without introducing any artifacts. Comparing the results of CTV and CSSWHTV, which are both 3-D TV denoising models, it is easy to see the ad-

TABLE 3
Values of the parameters for the different methods in the real data experiments

Data	SSAHTV	Fusion	CTV	MWF	BiShr.	HSSNR	CSSWHTV
Urban	$\lambda = 1$	$\lambda_1 = 1$ $\lambda_2 = 1$	λ=1/20 $\beta = 100$	$k_1 = 100$ $k_2 = 100$ $k_3 = 10$	—	—	$\lambda_1 = 1/5$ $\lambda_2 = 1$
Indian Pines	$\lambda = 1/4$	$\lambda_1 = 1/4$ $\lambda_2 = 1/4$	λ=1/22 $\beta = 100$	$k_1 = 80$ $k_2 = 80$ $k_3 = 10$	—	—	$\lambda_1 = 1/4$ $\lambda_2 = 50$

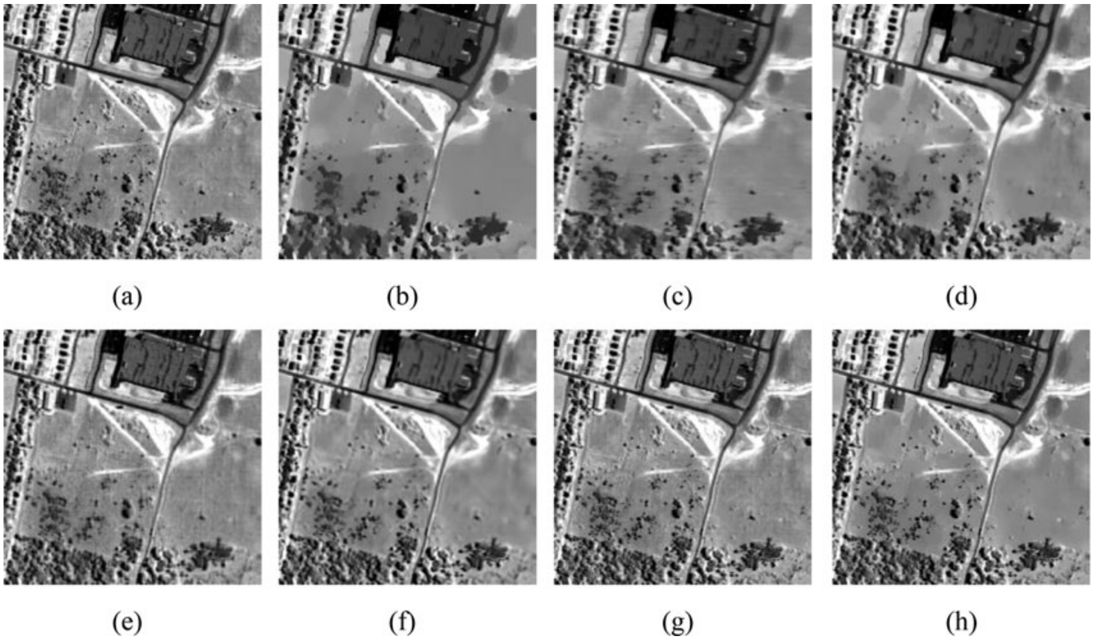


FIG. 11. Denoising results of the Urban dataset in the real data experiments: (a) noise band 100 (1321.80 nm), (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

vantages of CSSWHTV, which can automatically adjust the penalizing strength for each pixel in both the spatial and spectral dimensions.

Based on the combination of the vectorial TV norm and the weight, the CSSWHTV method can automatically adjust the penalizing strength in a 3-D manner in both the spatial and spectral dimensions. Because the weights $W_{i,j}$ and W_k' are used in both the spatial and spectral dimensions, as in Equation (20), the effect on the SNR and MSA of using only 1 of them, neither of them, and both of them, were compared and are

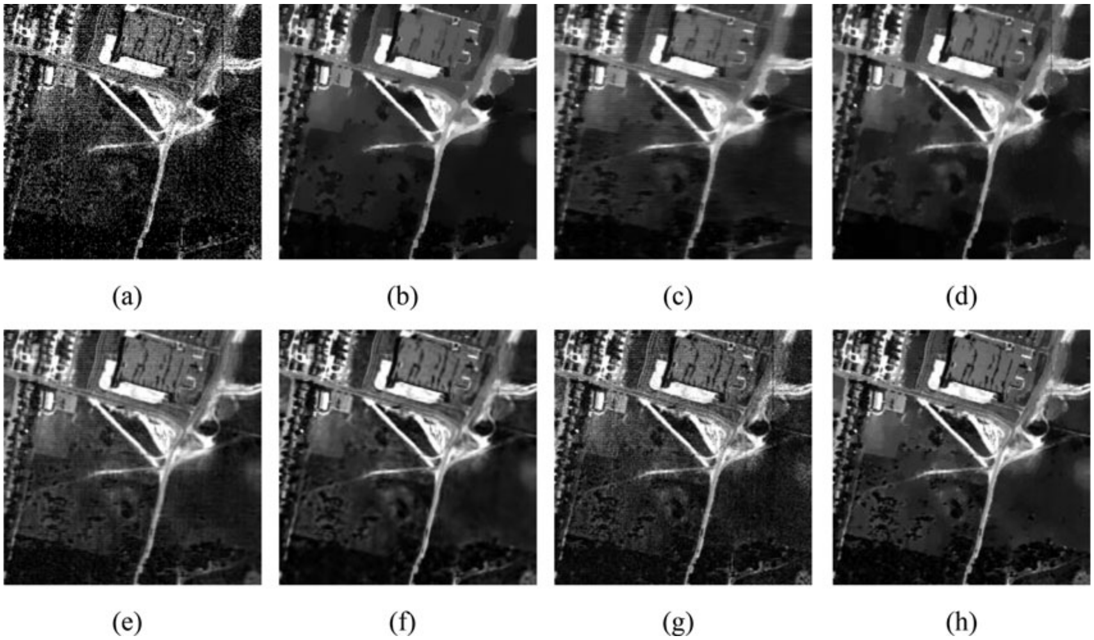


FIG. 12. Denoising results of the Urban dataset in the real data experiments: (a) noise band 192 (2456.57 nm), (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

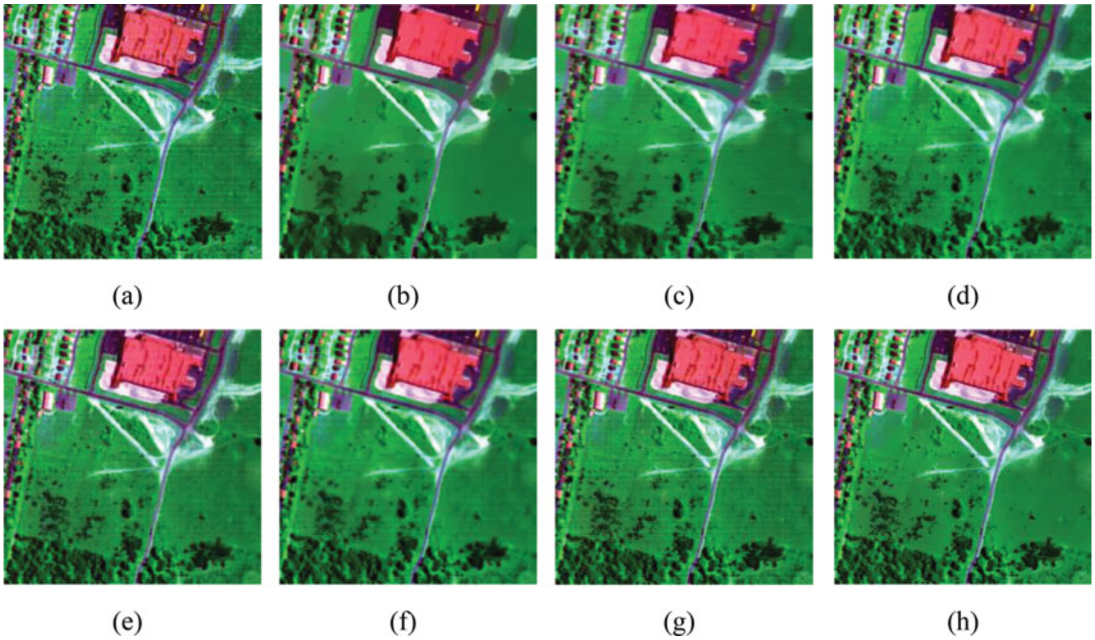


FIG. 13. Denoising results of the Urban dataset in the real data experiments: (a) combination of noise bands 1 (401.29 nm), 103 (1364.81 nm), and 192 (2456.57 nm); (b) SSAHTV; (c) Fusion; (d) CTV; (e) MWF; (f) BiShrink; (g) HSSNR; and (h) CSSWHTV.

depicted in Figure 8. The regularization parameters for all the cases were adjusted to achieve the highest SNR values. The results were averaged over 10 experiments, and the error bars show the standard deviations for the experiments. As shown in Figure 8, for the different levels of noise, better results can be acquired by using $W_{i,j}$ or W_k' than by not using either of them, and the best results are achieved when both $W_{i,j}$ and W_k' are used.

In the following, the reason behind the good performance of the weight will be further illustrated with a simulated piecewise

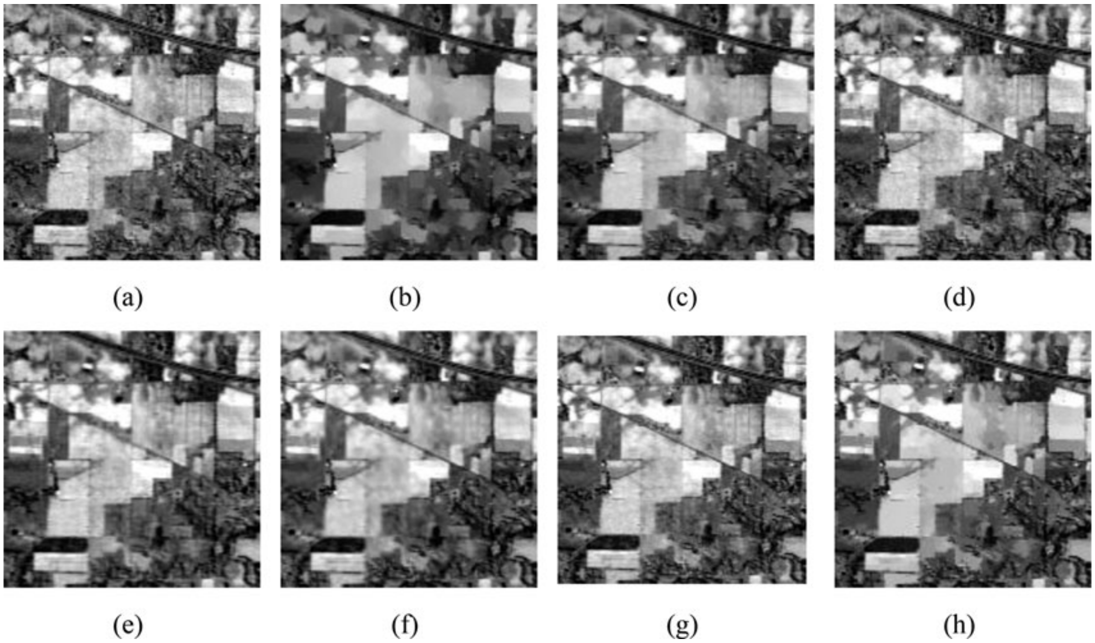


FIG. 14. Denoising results of the Indian Pines dataset in the real data experiments: (a) noise band 102 (1342.73 nm), (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

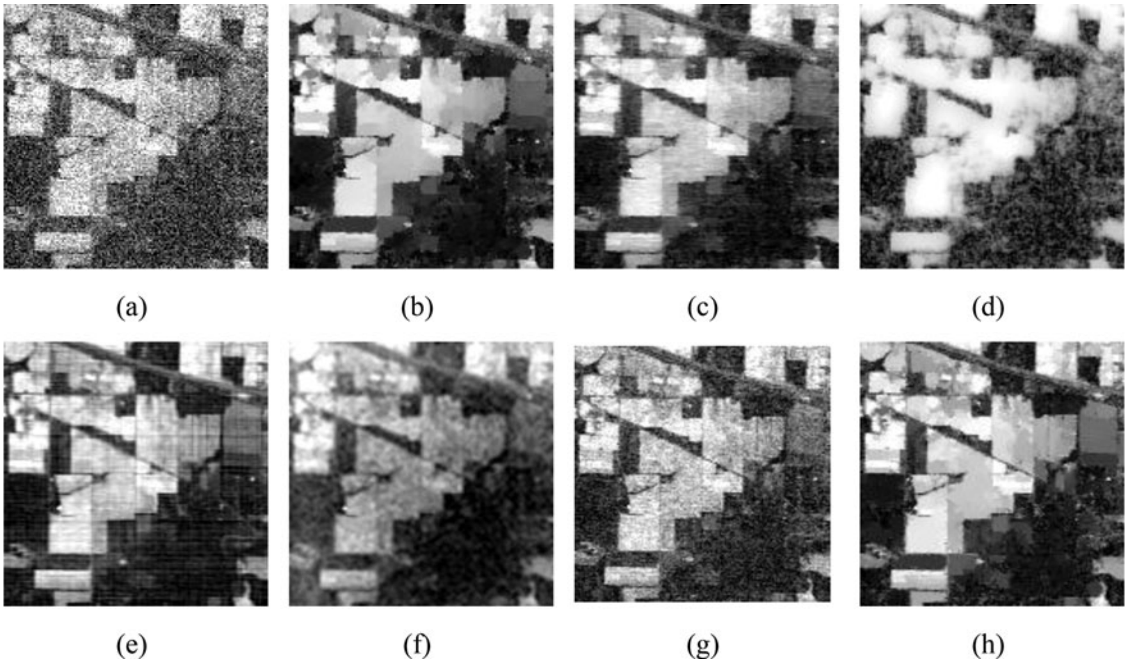


FIG. 15. Denoising results of the Indian Pines dataset in the real data experiments: (a) noise band 204 (2479.25 nm), (b) SSAHTV, (c) Fusion, (d) CTV, (e) MWF, (f) BiShrink, (g) HSSNR, and (h) CSSWHTV.

constant HSI. Four types of mineral spectra from the United States Geological Survey (USGS) spectral library were used to generate the simulated HSI. The simulated HSI consisted of 420 bands, 128 samples of 128 lines, and 4 blocks. In the order of the top-left, bottom-left, top-right, and bottom-right blocks, every pixel in the same block had the same spectra of alunite, andradite, calcite, and chlorite, respectively. Gaussian random noise was added to the simulated data, and the noise variances follows Equation (34). Here, 2 cases were considered to test the proposed weight. In Case 1, we directly added noise to the whole data, and the SNR of the noisy HSI was 18.89 dB. In Case 2, we independently added Gaussian noise to the 4 blocks, and the SNR of the whole noisy HSI was 14.23 dB, and the SNR values of the top-left, bottom-left, top-right, and bottom-right blocks were 23.37 dB, 19.58 dB, 12.49 dB, and 5.20 dB, respectively. Note that the configuration in Case 2 does not imply that the noise variance varies as much within a band as in a real case, but is to help analyze and show the changes of $FV_{i,j}$, $PV_{i,j}/FV_{i,j}$, and the weight $W_{i,j}$ with respect to the magnitude of the noise intensity in the image.

In Figure 9, the derivation process of the proposed weight is illustrated. Comparing Figure 9(a) and Figure 9(b), it is clear that the value of $FV_{i,j}$ is positively related to the noise intensity and the signal variation. In Figure 9(c), it is shown that the value of $PV_{i,j}/FV_{i,j}$ can effectively represent the signal variation. As shown in Figure 9(d), the proposed weight is positively correlated with the noise intensity and negatively correlated with the variation of the pixel.

The convergence of the proposed CSSWHTV method was proved, and here we only give the experimental analysis. To save space, only the results when $\sigma = 0.4$ are reported, because the other results are consistent with these observations. Figure 10 shows the SNR and MSA values of the results produced by the CSSWHTV method with different iteration numbers, and it is shown that the CSSWHTV method can converge in only a few iterations.

TABLE 4
Classification accuracies using SVM on different denoising results of Indian Pines dataset

	Original image	SSAHTV	Fusion	CTV	MWF	BiShrink	HSSNR	CSSWHTV
OA	74.74%	89.33%	90.29%	88.92%	80.21%	84.84%	78.42%	91.99%
Kappa	0.7075	0.8755	0.8868	0.8709	0.7704	0.8235	0.7498	0.9064

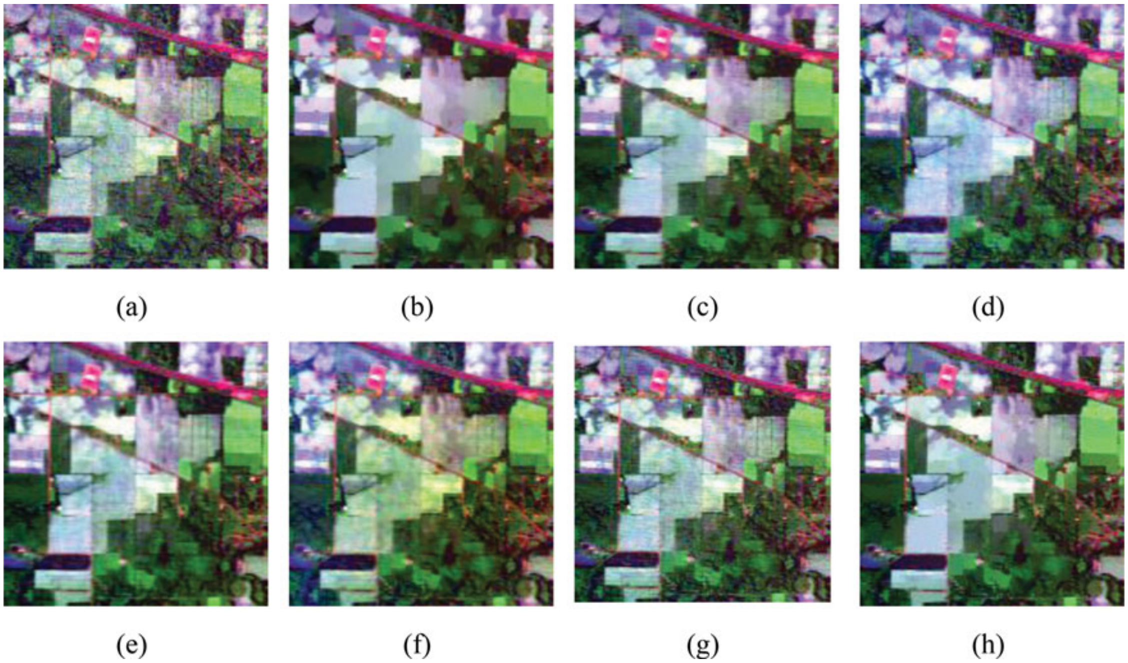


FIG. 16. Denoising results of the Indian Pines dataset in the real data experiments: (a) combination of noise bands 3 (419.62 nm), 102 (1342.73 nm), and 204 (2479.25 nm); (b) SSAHTV; (c) Fusion; (d) CTV; (e) MWF; (f) BiShrink; (g) HSSNR; and (h) CSSWHTV.

Real Data Experiments

To further confirm the effectiveness of the proposed method, 2 real data experiments were performed. In the first experiment, we used the well-known HYDICE Urban dataset, which has a size of 307 × 307 × 210. A portion of the data of size 200 × 202 × 210 from line 93 to 292, column 95 to 296, was used. In the other experiment, we used the well-known AVIRIS Indian Pines dataset, which has a size of 145 × 145 × 220. For the different methods, the parameters were set empirically to achieve the best visual quality, and the parameter values are listed in Table 3.

For the Urban data, the bands 140–151 and 206–210, which are atmospheric and water absorption bands, have been removed from the original hyperspectral image. Therefore, there were only 193 bands used in the experiment. Different levels of noise are contained in the image, as shown in Figure 11(a), and Figure 12(a). For band 100, the noise is relatively low, and for band 192, there are stripes and random noise. From the results shown in Figure 11, we can see that the CSSWHTV method can effectively preserve the edges and textures for the band with a low level of noise. From the results shown in Figure 12, the CSSWHTV method shows its ability to remove random and striping noise, and the result of the CSSWHTV method looks very sharp and clear, without any artifacts. Figure 13 also shows the color results with bands 1, 103, and 192 as the red, green, and blue colors, respectively, and the effective performance of the proposed method is clear. From the above results, it appears that the proposed method can perform even better with the real Urban data. Due to the varying noise strength, it is difficult for the MWF, BiShrink, and CTV methods to effectively remove the stripes and preserve the details. For the SSAHTV method, the regularization parameter has to be set as a large value to remove all the stripes, at the cost of blurred edges and textures. For the CSSWHTV method, the TV regularization term in the spectral dimension can give a smooth prior along the spectral direction, which makes it relatively easy to remove the stripes.

For the Indian Pines data, bands 150–163, which are atmospheric and water absorption bands, have been removed from the original hyperspectral image. Therefore, there were only 206 bands used in the experiment. Figure 14(a) and Figure 15(a) show the noise bands 102 and 204 with relatively low and high noise levels, respectively. From the results, we can see that the CSSWHTV method outperforms the other methods for the bands with different noise cases. Figure 16 shows the color results composed of bands 3, 102, and 204 for the different methods, which further validate the performance of the CSSWHTV method. Because the ground truth data is available for the Indian Pines dataset, we also performed classification on the results. By using cross-validation, 9 classes of corn-notill, corn-mintill, grass/pasture, grass/trees, hay-windrowed, soybeans-notill, soybeans-mintill, soybeans-clean, and woods were used in the experiment. Fifty training samples were randomly chosen for each class, and the Support Vector Machine (SVM) method is used for the classification. The classification process is repeated 10 times, and the mean overall accuracy (OA) and mean kappa coefficient are reported in Table 4. From the results, it shows that the improvement of the classification accuracy after image denoising is obvious, and the proposed method achieves the best results.

CONCLUSIONS

In this article, a combined spatial and spectral weighted hyperspectral total variation (CSSWHTV) model has been proposed for denoising HSIs. By viewing the HSI as a 3-D cube, the new model combines the TV regularizations in both the spatial and spectral dimensions into a unified framework. Moreover, to clearly remove the noise and retain the edges, the regularization terms have been designed to be able to automatically control the penalizing strength for each pixel in both the spatial and spectral dimensions. To solve the proposed new model, we have proposed a new algorithm, which is a version of the well-known ADMM procedure, which is fast and has a good convergence property. Both simulated and real data experiments were conducted to validate the effectiveness of the proposed method. The consistent improvements when compared with the other well-known methods in different noise cases show that the proposed method is robust and can effectively remove noise while preserving the details, without introducing any new artifacts.

Despite the effective performance of the proposed method, an investigation into what extent it could improve the precision of the subsequent applications would be interesting. The automatic and efficient selection of the regularization parameters also needs further study.

FUNDING

This work was supported in part by the 863 program under Grant 2013AA12A301, by the National Natural Science Foundation of China under Grants 41571362 and 41431175, and by the NASG Key Laboratory of Land Environment and Disaster Monitoring (No. LEDM2014B01).

REFERENCES

Acito, N., Diani, M., and Corsini, G. 2010. “Hyperspectral signal subspace identification in the presence of rare signal components.” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48(No. 4): pp. 1940–1954.

Acito, N., Diani, M., and Corsini, G. 2011a. “Signal-dependent noise modeling and model parameter estimation in hyperspectral images.” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49(No. 8): pp. 2957–2971.

Acito, N., Diani, M., and Corsini, G. 2011b. “Subspace-based striping noise reduction in hyperspectral images.” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49(No. 4): pp. 1325–1342.

Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. 2011. “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems.” *IEEE Transactions on Image Processing*, Vol. 20(No. 3): pp. 681–695.

Alparone, L., Selva, M., Aiazzi, B., Baronti, S., Butera, F., and Chiarantini, L. 2009. “Signal-dependent noise modelling and estimation of new-generation imaging spectrometers.” Paper presented at IEEE 1st Workshop on Hyperspectral Image & Signal Processing: Evolution in Remote Sensing (WHISPERS), Grenoble, France, August 2009.

Bioucas-Dias, J. M., and Figueiredo, M. A. 2010. “Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing.” Paper presented at IEEE 2nd Workshop on Hyperspectral Image & Signal Processing: Evolution in Remote Sensing (WHISPERS), Reykjavik, Iceland, June 2010.

Bioucas-Dias, J. M., Figueiredo, M. A., and Oliveira, J. P. 2006. “Total variation-based image deconvolution: a majorization-minimization approach.” Paper presented at International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, May 2006.

Bresson, X., and Chan, T. F. 2008. “Fast dual minimization of the vectorial total variation norm and applications to color image processing.” *Inverse Problems and Imaging*, Vol. 2(No. 4): pp. 455–484.

Buades, A., Coll, B., and Morel, J.-M. 2005. “A non-local algorithm for image denoising.” Paper presented at IEEE Computer Vision and Pattern Recognition, San Diego, CA, June 2005.

Chen, G., Bui, T. D., and Krzyzak, A. 2011. “Denoising of three-dimensional data cube using bivariate wavelet shrinking.” *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 25(No. 3): pp. 403–413.

Chen, G., Bui, T. D., Quach, K. G., and Qian, S.-E. 2014. “Denoising hyperspectral imagery using principal component analysis and block-matching 4D filtering.” *Canadian Journal of Remote Sensing*, Vol. 40(No. 1): pp. 60–66.

Chen, G., and Qian, S.-E. 2011. “Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage.” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49(No. 3): pp. 973–980.

Chen, G., Qian, S.-E., and Gleason, S. 2012. “Denoising of hyperspectral imagery by combining PCA with block-matching 3-D filtering.” *Canadian Journal of Remote Sensing*, Vol. 37(No. 6): pp. 590–595.

Eckstein, J., and Bertsekas, D. P. 1992. “On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators.” *Mathematical Programming*, Vol. 55(No. 1-3): pp. 293–318.

Elad, M., and Aharon, M. 2006. “Image denoising via sparse and redundant representations over learned dictionaries.” *IEEE Transactions on Image Processing*, Vol. 15(No. 12): pp. 3736–3745.

Gabay, D., and Mercier, B. 1976. “A dual algorithm for the solution of nonlinear variational problems via finite element approximation.” *Computers & Mathematics with Applications*, Vol. 2(No. 1): pp. 17–40.

Goldstein, T., and Osher, S. 2009. “The split Bregman method for L1-regularized problems.” *SIAM Journal on Imaging Sciences*, Vol. 2(No. 2): pp. 323–343.

Guo, X., Huang, X., Zhang, L., and Zhang, L. 2013. “Hyperspectral image noise reduction based on rank-1 tensor decomposition.” *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 83(No. 9): pp. 50–63.

Harris, J., Ponomarev, P., Shang, J., and Rogge, D. 2006. “Noise reduction and best band selection techniques for improving classification results using hyperspectral data: application to lithological mapping in Canada’s Arctic.” *Canadian Journal of Remote Sensing*, Vol. 32(No. 5): pp. 341–354.

He, W., Zhang, H., Zhang, L., and Shen, H. 2016. “ Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration.” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54(No. 1): pp. 178–188.

Iordache, M.-D., Bioucas-Dias, J. M., and Plaza, A. 2012. “Total variation spatial regularization for sparse hyperspectral unmixing.” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50(No. 11): pp. 4484–4502.

Iordache, M.-D., Bioucas-Dias, J. M., and Plaza, A. 2014. "Collaborative sparse regression for hyperspectral unmixing." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52(No. 1): pp. 341–354.

Jiang, C., Zhang, H., Shen, H., and Zhang, L. 2014. "Two-step sparse coding for the pan-sharpening of remote sensing images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7(No. 5): pp. 1792–1805.

Kuybeda, O., Malah, D., and Barzohar, M. 2007. "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors." *IEEE Transactions on Signal Processing*, Vol. 55(No. 12): pp. 5579–5592.

Letexier, D., and Bourennane, S. 2008. "Noise removal from hyperspectral images by multidimensional filtering." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46(No. 7): pp. 2061–2069.

Li, T., Chen, X.-M., Xue, B., Li, Q.-Q., and Ni, G.-Q. 2010. "A total variation denoising algorithm for hyperspectral data." Paper presented at Photonics Asia, Beijing, China, October 2010.

Lin, T., and Bourennane, S. 2013a. "Hyperspectral image processing by jointly filtering wavelet component tensor." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 51(No. 6): pp. 3529–3541.

Lin, T., and Bourennane, S. 2013b. "Survey of hyperspectral image denoising methods based on tensor decompositions." *EURASIP Journal on Advances in Signal Processing*, Vol. 2013(No. 1): pp. 1–11.

Liu, X., Bourennane, S., and Fossati, C. 2012. "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50(No. 10): pp. 3717–3724.

Méndez-Rial, R., Calvino-Cancela, M., and Martín-Herrero, J. 2010. "Accurate implementation of anisotropic diffusion in the hypercube." *IEEE Geoscience and Remote Sensing Letters*, Vol. 7(No. 4): pp. 870–874.

Maggioni, M., Katkovnik, V., Egiazarian, K., and Foi, A. 2013. "Nonlocal transform-domain filter for volumetric data denoising and reconstruction." *IEEE Transactions on Image Processing*, Vol. 22(No. 1): pp. 119–133.

Martín-Herrero, J. 2007. "Anisotropic diffusion in the hypercube." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45(No. 5): pp. 1386–1398.

Mendez-Rial, R., and Martin-Herrero, J. 2012. "Efficiency of semi-implicit schemes for anisotropic diffusion in the hypercube." *IEEE Transactions on Image Processing*, Vol. 21(No. 5): pp. 2389–2398.

Muti, D., and Bourennane, S. 2007. "Survey on tensor signal algebraic filtering." *Signal Processing*, Vol. 87(No. 2): pp. 237–249.

Othman, H., and Qian, S.-E. 2006. "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44(No. 2): pp. 397–408.

Perona, P., and Malik, J. 1990. "Scale-space and edge detection using anisotropic diffusion." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12(No. 7): pp. 629–639.

Portilla, J., Strela, V., Wainwright, M. J., and Simoncelli, E. P. 2003. "Image denoising using scale mixtures of Gaussians in the wavelet domain." *IEEE Transactions on Image Processing*, Vol. 12(No. 11): pp. 1338–1351.

Qian, Y., Shen, Y., Ye, M., and Wang, Q. 2012. "3-D nonlocal means filter with noise estimation for hyperspectral imagery denoising." Paper presented at IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, July 2012.

Qian, Y., and Ye, M. 2013. "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* Vol. 6(No. 2): pp. 499–515.

Rasti, B., Sveinsson, J. R., and Ulfarsson, M. O. 2014. "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52(No. 10): pp. 6688–6698.

Rudin, L. I., Osher, S., and Fatemi, E. 1992. "Nonlinear total variation based noise removal algorithms." *Physica D: Nonlinear Phenomena*, Vol. 60(No. 1): pp. 259–268.

Selesnick, I. W. 2002. "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency." *IEEE Transactions on Signal Processing*, Vol. 50(No. 11): pp. 2744–2756.

Şendur, L., and Selesnick, I. W. 2002. "Bivariate shrinkage with local variance estimation." *IEEE Signal Processing Letters*, Vol. 9(No. 12): pp. 438–441.

Wang, Y., Niu, R., and Yu, X. 2010. "Anisotropic diffusion for hyperspectral imagery enhancement." *IEEE Sensors Journal*, Vol. 10(No. 3): pp. 469–477.

Yang, J., Yin, W., Zhang, Y., and Wang, Y. 2009. "A fast algorithm for edge-preserving variational multichannel image restoration." *SIAM Journal on Imaging Sciences*, Vol. 2(No. 2): pp. 569–592.

Yuan, Q., Zhang, L., and Shen, H. 2012. "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50(No. 10): pp. 3660–3677.

Yuan, Q., Zhang, L., and Shen, H. 2014. "Hyperspectral image denoising with a spatial–spectral view fusion strategy." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52(No. 5): pp. 2314–2325.

Zelinski, A. C., and Goyal, V. K. 2006. "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation." Paper presented at IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Denver, Colorado, USA, August 2006.

Zhang, H. 2012. "Hyperspectral image denoising with cubic total variation model." Paper presented at ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Melbourne, Australia, August 2012.

Zhang, H., He, W., Zhang, L., Shen, H., and Yuan, Q. 2014. "Hyperspectral image restoration using low-rank matrix recovery." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52(No. 8): pp. 4729–4743.

Zhang, H., Zhang, L., and Shen, H. 2012. "A super-resolution reconstruction algorithm for hyperspectral images." *Signal Processing*, Vol. 92(No. 9): pp. 2082–2096.

Zhang, H., Zhai, H., Zhang, L., and Li, P. 2016. "Spectral-Spatial Sparse Subspace Clustering for Hyperspectral Remote Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2016.2524557.

Zhao, X.-L., Wang, F., Huang, T.-Z., Ng, M. K., and Plemmons, R. J. 2013. "Deblurring and sparse unmixing for hyperspectral images." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 51(No. 7): pp. 4045–4058.

Zhao, Y.-Q., and Yang, J. 2015. "Hyperspectral image denoising via sparse representation and low-rank constraint." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53(No. 1): pp. 296–308.