Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

Hyperspectral Image Denoising Using a Spatial–Spectral Monte Carlo Sampling Approach

Linlin Xu, *Member, IEEE*, Fan Li, Alexander Wong, *Member, IEEE*, and David A. Clausi, *Senior Member, IEEE*

***Abstract*—Hyperspectral image (HSI) denoising is essential for enhancing HSI quality and facilitating HSI processing tasks. However, the reduction of noise in HSI is a difficult work, primarily due to the fact that HSI consists much more spectral bands than other remote sensing images. Therefore, comparing with other image denoising jobs that rely primarily on spatial information, efficient HSI denoising requires the utilization of both spatial and spectral information. In this paper, we design an unsupervised spatial–spectral HSI denoising approach based on Monte Carlo sampling (MCS) technique. This approach allows the incorporation of both spatial and spectral information for HSI denoising. Moreover, it addresses the noise variance heterogeneity effect among different HSI bands. In the proposed HSI denoising scheme, MCS is used to estimate the posterior distribution, in order to solve a Bayesian least squares optimization problem. Based on the proposed scheme, we iterate all pixels in HIS and denoise them sequentially. A referenced pixel in hyperspectral image is denoised as follows. First, some samples are randomly drawn from image space close to the referenced pixel. Second, based on a spatial–spectral similarity likelihood, relevant samples are accepted into a sample set. Third, all samples in the accepted set will be used for calculating the estimation of posterior distribution. Finally, based on the posterior, the noise-free pixel value is estimated as the discrete conditional mean. The proposed method is tested on both simulated and real hyperspectral images, in comparison with several other popular methods. The results demonstrate that the proposed method is capable of removing the noise largely, while also preserving image details very well.**

***Index Terms*—Bayesian least squares optimization, hyperspectral imagery denoising, Monte Carlo Sampling, spatial-spectral similarity likelihood.**

I. Introduction

A HYPERSPECTRAL image (HSI) is characterized by high spectral resolution. With hundreds of spectral bands, ranging from visible to infrared bands, HSI is capable of supporting various important applications, including mineralogy, environmental monitoring and defense. Nevertheless, due to the uncertainty and complexity of the remote sensing system, HSI is unavoidable contaminated by point noise, which disturbs computer-aided image processing tasks, such as classification, spectral unmixing, and target detection. Therefore, efficient HSI denoising techniques capable of reducing noise and restoring scene signals are required.

HSI denoising has been achieved by different approaches. The transformed-domain techniques have been used for HSI denoising, e.g., Atkinson *et al.* [1] proposed a discrete Fourier transform and wavelet-based hyperspectral image denoising algorithm. Othman and Qian [2] proposed a hybrid spatial–spectral wavelet shrinkage approach to address noise level variation across different bands, where the HSI is transformed into spectral derivative space, then denoised by performing wavelet denoising in spatial and spectral domain independently. Chen and Qian [3], [4] performed wavelet denoising in 2-D image domain, and then apply PCA on denoised HSI for simultaneous dimensionality reduction and denoising of hyperspectral imagery. Chen *et al.* [5] performed HSI denoising by combining PCA and wavelet techniques, where PCA is first used to decorrelate the data, and then wavelets are used to perform denoising in low-energy PCA output channels. Chen [6] extended Sendur and Selesnick's bivariate wavelet thresholding [7] from 2-D image denoising to 3-D cube denoising.

The sparse regularization has been used for performing HSI denoising, e.g., Rasti *et al.* [8] performed HSI denoising based on sparse analysis regularization and a 3-D overcomplete wavelet dictionary. Bourguignon *et al.* [9] conducted HSI denoising by sparsely representing the spectra observations over a union of canonical and the discrete cosine transform (DCT) bases. Qian and Ye [10] proposed a nonlocal spectral–spatial structured sparse representation approach, where the HSI is first partitioned into several groups, and sparse representation with spectral–spatial structure is performed within each group to remove noise.

A spectral–spatial adaptive total-variation (TV) model has been proposed by Yuan *et al.* [11] for HSI denoising, which is capable of accounting for the noise intensity difference between different bands and spatial property differences between different pixels. A genetic kernel Tucker decomposition (GKTD) algorithm was proposed by Karami *et al.* [12] for HSI denoising, which exploits both the spectral and the spatial information in the images. An adaptive filtering approach was proposed by Phillips *et al.* [13]. A multiway filtering method based on a tensor model has been proposed by Letexier and Bourennane [14] for HSI denoising, and used to aid target detection [15] and classification [16].

In this paper, we present a novel statistical approach, where HSI denoising is formulated as a Bayesian least squares optimization problem. The posterior probability is estimated by a Monte Carlo sampling (MCS) method. MCS approach works in a nonparametric manner, therefore is more flexible than

Manuscript received September 14, 2014; revised January 19, 2015; accepted January 30, 2015. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), in part by the Canadian Space Agency (CSA), and in part by the Canada Research Chairs Program.

The authors are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: l44xu@uwaterloo.ca; alexander.wong@uwaterloo.ca; dclausi@uwaterloo.ca).

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

Digital Object Identifier 10.1109/JSTARS.2015.2402675

1939-1404 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the parametric methods. It has been applied for denoising of natural images [17] and SAR images [18]. However, both approaches are not suitable for the HSI denoising, considering the particularities of HSI.

This paper, therefore, introduces a new statistical approach based on MCS for HSI denoising. A spatial–spectral acceptance likelihood based on band-dependent noise distribution is used to aid the MCS process. The proposed acceptance likelihood is capable of capturing the signals in both spatial and spectral domain, and is robust to noise. Moreover, the method accounts for the noise variance heterogeneity across various bands, and can perform denoising adaptively, based on the noise levels of different bands. Once the posterior has been obtained, the noise-free value can be estimated as the discrete conditional mean, according to the Bayesian optimization scheme.

The contribution of this paper lies in the following aspects. 1) A Bayesian least square optimization scheme is introduced into the HSI denoising problem, leading to an efficient statistical HSI denoising approach. 2) An MCS approach that is capable of efficiently collecting samples in image space for posterior distribution estimation is designed. 3) A spatial–spectral similarity likelihood that is capable of accounting for the band-dependent noise distribution and the local patterns in both spatial and spectral domain is created.

This paper is organized as follows. Section II introduces the proposed denoising framework, as well as the MCS and the spatial–spectral similarity likelihood approaches. Section III presents the experiments results on both simulated and real HSIs. Section IV concludes the study.

II. Methodology

A. Hyperspectral Noise Model and Estimation

Here, an observed noisy pixel variable in HSI is denoted by y_{ijb}, where the indices (i,j) determine the location of y_{ijb} in image space, while b represents the band number in the spectral domain. Accordingly, the degradation model of HSI is expressed as [5], [11], [19]

$$y_{ijb} = x_{ijb} + n_b \tag{1}$$

where x_{ijb} denotes the unobservable noise-free variable, and n_b is band-dependent noise. The term band-dependent noise refers to noise which tends to assume different variances on different spectral bands. We assume that n_b satisfies zero-mean Gaussian distribution

$$P(n_b) = \frac{1}{\sigma_b\sqrt{2\pi}} exp(-n_b^2/2\sigma_b^2) \tag{2}$$

where σ_b^2 is the noise variance of the band b image.

The noise variances of different spectral bands $\sigma_b^2(b = 1,2,\ldots,B)$ are not necessarily the same. For example, in Fig. 1, the estimated noise standard deviation varies greatly across different bands in the benchmark Indian Pines image [20]. Ignoring this noise variance heterogeneity issue in denoising will lead to insufficiently removal of noise in some bands, but erasing of scene signals in some others. Therefore, it is important to design statistical distributions that are capable of

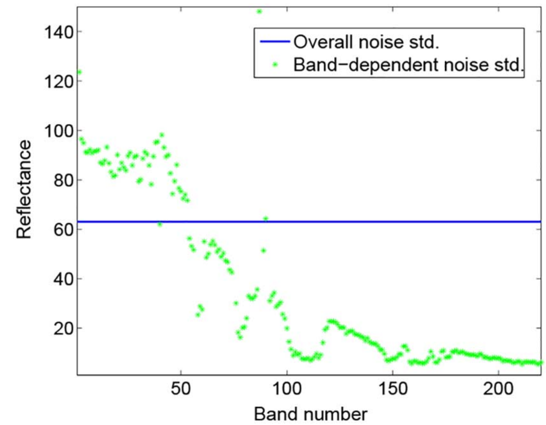


Fig. 1. Estimated noise standard deviation (std.) of different bands in the benchmark Indian Pines HSI. The overall noise std., displayed by the blue line, is estimated by assuming that noise in all bands has the same variance. The homogeneous area used for calculating the std. values is displayed on the noisy image in Fig. 18.

accounting for the noise level variation across bands. That is why we adopt the band-dependent noise distribution in (2). As we shall see in Section II-D, this noise distribution can be incorporated into the proposed MCS scheme for enhancing the denoising performance.

Since the noise variance of each of the spectral bands $\sigma_b^2(b = 1,2,\ldots,B)$ is generally unknown, they need to be estimated based on the noisy observations. Several methods can be used for HSI noise variance estimation [21], e.g., local means and local standard deviations (LMLSD) method [22], spectral and spatial decorrelation (SSDC) method [23], homogeneous regions division and spectral decorrelation (HRDSDC) method [24], and residual-scaled local standard deviation (RLSD) method [25]. Different methods tend to impose different assumptions on the spatial–spectral characteristics of signal in HSI, and the performance of a method deteriorates when its assumption is not satisfied. In this paper, considering that our main focus is the introduction of the spatial–spectral MCS model for HSI denoising, we adopt a simple approach for noise variance estimation. We estimate noise variances by identifying a homogeneous area in each image band of HSI. Then, σ_b^2 is estimated as the variance of corresponding pixel values in the bth band. This approach can yield acceptable estimate of noise variances if the signal in the identified area is fairly stationary.

B. Problem Formulation

HSI denoising is essentially an inverse problem, where the estimation of noise-free variable x from the observed noisy variable y can be formulated as a Bayesian least squares optimization problem [26]

$$\begin{aligned}\hat{x} &= \text{argmin}_x \left\{E((x-\hat{x})^2|y)\right\} \\ &= \text{argmin}_x \left\{\int (x-\hat{x})^2 p(x|y)dy\right\}.\end{aligned} \tag{3}$$

In the above equation, $p(x|y)$ represents the posterior distribution of x given y. To estimate $\hat{x}$, the derivative is taken and set equal to zero

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$$\frac{\partial}{\partial \hat{x}} \int (x-\hat{x})^2 p(x|y)dx = -\int 2(x-\hat{x})p(x|y)dx = 0 \quad (4)$$

$$\int xp(x|y)dx = \int \hat{x}p(x|y)dx. \quad (5)$$

The right-hand side of (5) can be simplified as

$$\int \hat{x}p(x|y)dx = \hat{x}\int p(x|y)dx = \hat{x}. \quad (6)$$

Therefore, the goal is to estimate the conditional mean

$$\hat{x} = \int xp(x|y)dx = E(x|y) \quad (7)$$

with the discrete form being expressed as

$$\hat{x} = \sum_x p(x|y)x. \quad (8)$$

However, the estimation of the conditional mean directly is difficult because the posterior distribution $p(x|y)$ is unknown, and it could be a complex function of y. In Section II-C, $p(x|y)$ is estimated via stochastic MCS, based on the spatial–spectral similarity likelihood in Section II-D. The weighted histogram, as the estimate of $p(x|y)$, will afterward be used to calculate $\hat{x}$, as the discrete conditional mean.

C. Posterior Estimation via Monte Carlo Sampling

This section describes an MCS approach for estimating the posterior probability in a nonparametric manner. It is more flexible than a parametric estimation approach.

Let y_0 represent a spatial–spectral variable, for which the posterior $p(x|y_0)$ is being estimated. Here, we use single index 0, instead of (i, j, b) for brevity. In MCS, $p(x|y_0)$ is estimated by drawing a sequence of samples $\Omega = \{y_1, y_2, \ldots, y_m\}$ around y_0, which are likely to be realizations of $p(x|y_0)$. Then using the samples in Ω, $\hat{p}(x|y_0)$ can be estimated by a weighted histogram approach.

More specifically, given the referenced pixel y_0, first, MCS chooses some pixels that are close to y_0 in image space. Then, $\alpha(y_k|y_0)$, the acceptance probability of the sampled pixel y_k, given y_0, is used to decide whether to accept y_k, or reject it. The acceptance probability $\alpha(y_k|y_0)$ is realized by some probabilistic similarity likelihood measures, which are detailed in Section II-D. Then, $\alpha(y_k|y_0)$ is compared with a random number into range $(0, 1)$ drawn from uniform distribution. If $\alpha(y_k|y_0)$ is bigger than the random number, y_k is accepted into Ω. Otherwise, it is rejected.

After sample pixels are selected into Ω, the importance-weighted Monte Carlo posterior estimate is computed in a nonparametric manner via the following weighted histogram approach [17]

$$\hat{p}(x|y_0) = \frac{\sum_{k\in\Omega} \alpha(y_k|y_0)\delta(x-y_k)}{T} \quad (9)$$

where $\alpha(y_k|y_0)$ represents the acceptance probability of y_k in Ω (see Section II-D for more details), $\delta(\cdot)$ is the Dirac delta function, and T is a normalization term such that $\sum_x \hat{p}(x|y_0) = 1$. Based on $\hat{p}(x|y_0)$, we can estimate noise-free image value as the discrete conditional mean, according to (8).

Note that only pixels in the same band with y_0 are considered as sampling candidates. As a result, the accepted samples in Ω belong to the same band as y_0. This choice is motivated by the fact that different spectral bands tend to capture different physical properties of the targets, and assume varying noise levels. Therefore, pixels in different bands are less likely to satisfy the same statistical distribution than pixels in the same band.

Based on the above consideration, instead of performing sampling from multiple adjacent spectral bands, we conduct band-wise sampling by drawing samples from pixels that are in the same band as y_0. The samples will be accepted based on an acceptance probability, $\alpha(y_k|y_0)$, whose realizations are detailed in Section II-D. In order to achieve accurate estimation of acceptance probability, in Section II-D, we use 3-D cubes centered at y_0 and the pixel samples for estimating $\alpha(y_k|y_0)$.

D. Spatial–Spectral Similarity Likelihood

The acceptance probability $\alpha(y_k|y_0)$ of the sampled pixel y_k determines whether it will be accepted for estimating the posterior distribution. Therefore, designing a robust and effective acceptance probability is crucial in MCS.

In practice, $\alpha(y_k|y_0)$ is realized by a probabilistic similarity likelihood [17], which measures the degree of similarity between the sampled pixel y_k and the referenced pixel y_0. In this paper, the similarity likelihood between y_0 and y_k is expressed as follows:

$$P(y_k|y_0) = \frac{1}{\sigma_b\sqrt{2\pi}} exp(-(y_k - y_0)^2/2\sigma_b^2) \quad (10)$$

where σ_b^2 is the noise variance of band b that entails both pixels y_0 and y_k. Therefore, $P(y_k|y_0)$ measures the probabilistic distance between y_0 and y_k.

Based on (10), one straightforward realization of $\alpha(y_k|y_0)$ is pixel-based likelihood

$$\alpha(y_k|y_0)_{pxl} = \{P(y_k|y_0)\}^{1/\beta} \quad (11)$$

where β is the scaling parameter. Using $\alpha(y_k|y_0)_{pxl}$ in MCS leads to an denoising approach, called pixel-MCS in this paper. However, pixel-based likelihood is very sensitive to noise.

In order to increase the robustness against noise, region-based textual similarity likelihood can be used for implementing $\alpha(y_k|y_0)$

$$\alpha(y_k|y_0)_{spa} = \left\{\prod_{ij} P(y_k(i,j)|y_0(i,j))\right\}^{1/\beta} \quad (12)$$

where indices (i, j) are used to iterate the corresponding pixel pair $y_k(i, j)$ and $y_0(i, j)$ within the sampled region R_k and the referenced region R_0. A region could be defined as a square window centered at a pixel. Note that (12) assumes that pixels in a region are independent. As we can see, in (12), two pixels are considered similar when their local spatial patterns are similar. Consequently, $\alpha(y_k|y_0)_{spa}$ is more robust to noise disturbance

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

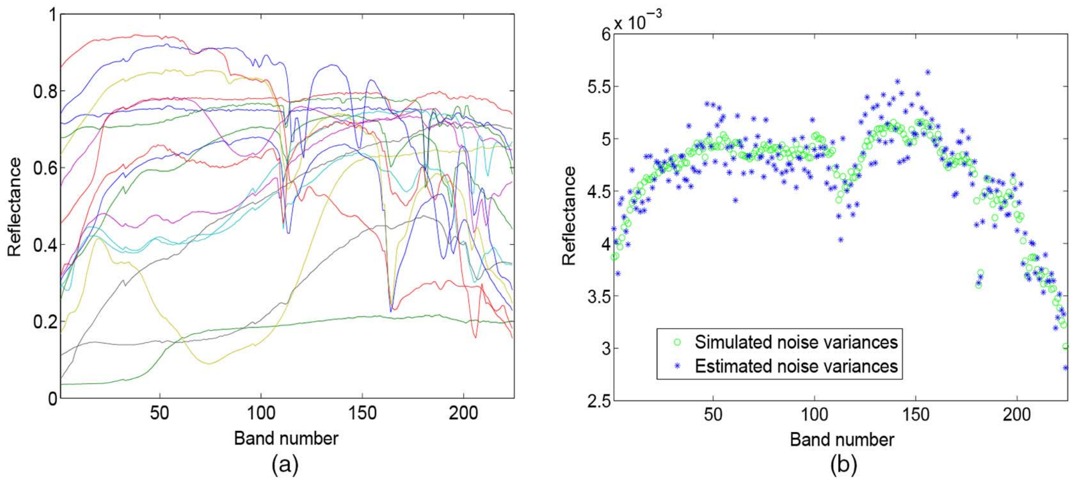


Fig. 2. (a) Spectra used for simulating data 1 based on the Indian Pines image. (b) Simulated and estimated noise variances at different bands in the simulated data 1.

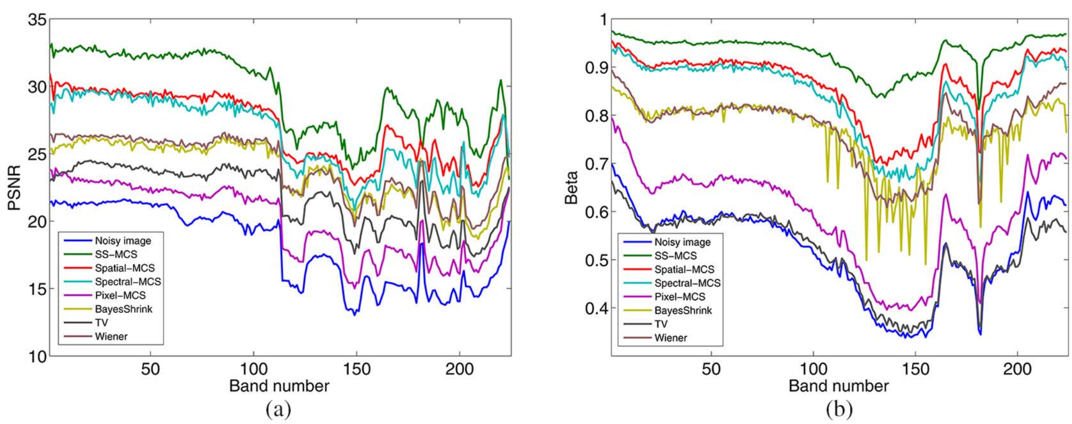


Fig. 3. (a) PSNR and (b) β achieved by different methods, at different bands of the simulated data 1. On both measures, the lines of SS-MCS are above those of the other methods, indicating that SS-MCS outperforms the other methods in terms of both noise removal and detail preservation.

than $\alpha(y_k|y_0)_{pxl}$. The resulting MCS denoising approach is called spatial-MCS.

Alternatively, $\alpha(y_k|y_0)$ can be implemented using only spectral information

$$\alpha(y_k|y_0)_{spe} = \left\{\prod_b P(y_k(b)|y_0(b))\right\}^{1/\beta} \tag{13}$$

where index (b) is used to iterate the corresponding pixel pair $y_k(b)$ and $y_0(b)$ within the sampled pixel stack V_k and the referenced stack V_0. A stack can be acquired by gathering pixels in all bands that are located at the same spatial position. In (13), two pixels are considered similar when their local spectral patterns are similar. Consequently, $\alpha(y_k|y_0)_{spe}$ is also robust to noise. The resulting MCS denoising approach is called spectral-MCS.

Since, in HSI, signal that is discriminative against noise lies in both spatial and spectral domain, instead of using solely spatial information or solely spectral information, we should incorporate both spatial and spectral information for building effective similarity likelihood. Accordingly, we expressed the acceptance probability $\alpha(y_k|y_0)_{ss}$ based on spatial–spectral similarity likelihood as follows:

$$\alpha(y_k|y_0)_{ss} = \left\{\prod_{ijb} P(y_k(i,j,b)|y_0(i,j,b))\right\}^{1/\beta} \tag{14}$$

where indices (i,j,b) are used to iterate the corresponding pixel pair $y_k(i,j,b)$ and $y_0(i,j,b)$ within a sampled 3-D cube T_k and referenced cube T_0. The size of the 3-D cube is characterized by region size in image space and the bandwidth in spectral domain. The resulting MCS denoising approach is called spatial–spectral-MCS (SS-MCS).

Therefore, comparing with spatial-MCS that considers only spatial information, and spectral-MCS that utilizes only spectral information, SS-MCS takes into account both spatial and spectral information for obtaining similar pixels. Consequently, SS-MCS is supposed to be more robust to noise, and more capable of finding pixels that are due to the same distribution.

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

TABLE I
STATISTICS ACHIEVED BY DIFFERENT METHODS ON SIMULATED IMAGES (BEST RESULTS ARE HIGHLIGHTED IN BOLD)

	Simulated data 1						Simulated data 2					
	PSNR			Beta			PSNR			Beta		
	Mean	Median	Std.	Mean	Median	Std.	Mean	Median	Std.	Mean	Median	Std.
Original	18.0	19.0	2.8	0.518	0.547	0.092	16.7	17.3	**1.0**	0.265	0.282	0.056
BayesShrink	23.6	24.7	**2.1**	0.765	0.796	0.074	22.3	22.7	1.3	0.387	0.377	0.060
TV	21.7	22.9	2.2	0.517	0.550	0.080	21.4	22.1	1.2	0.285	0.281	**0.031**
Wiener	24.0	25.0	2.2	0.775	0.801	0.069	22.7	23.5	1.4	0.446	0.457	0.083
Pixel-MCS	20.0	21.2	2.6	0.594	0.629	0.101	18.2	18.8	1.3	0.296	0.312	0.069
Spectral-MCS	26.2	27.3	2.7	0.841	0.880	0.079	22.5	23.3	1.7	0.490	0.516	0.112
Spatial-MCS	27.0	27.5	2.5	0.863	0.895	0.069	23.0	23.6	1.5	0.495	0.527	0.111
SS-MCS	**29.6**	**30.0**	2.8	**0.930**	**0.947**	**0.035**	**24.3**	**24.6**	1.3	**0.617**	**0.640**	0.083

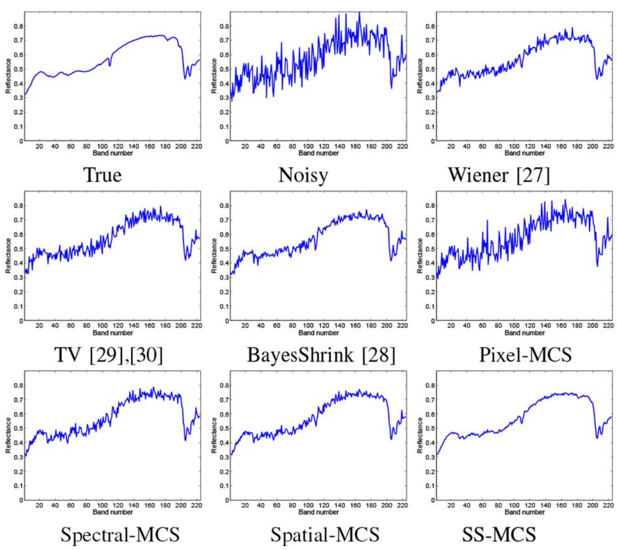


Fig. 4. Spectra at pixel (70,70) of the simulated data 1, before and after denoising by different methods. SS-MCS produce the most similar spectra to the true image. The spectra of spatial-MCS also seems to have less fluctuations than other methods. Pixel-MCS fails to significantly reduce the noise.

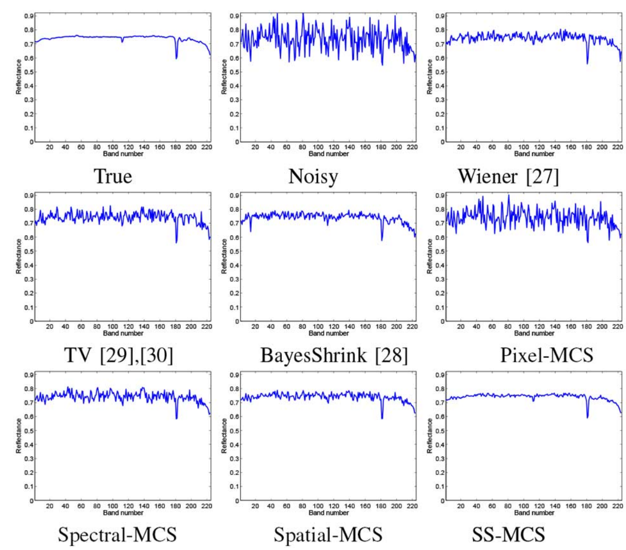


Fig. 5. Spectra at pixel (130,130) of the simulated data 1, before and after denoising by different methods.

E. Summary of the Proposed Algorithm

The proposed algorithm is summarized as the following steps:

1) For each pixel in the HSI, randomly draw M pixel-samples in the same image band, from a search area around the referenced pixel y_0 using MCS. For each pixel-sample y_0, obtain a 3-D cube sample T_k centered at y_k. Obtain also a 3-D cube T_0 centered at y_0.
2) For each 3-D cube-sample T_k, calculate the probabilistic similarity of each pixel-pair $y_k(i,j,b)$ and $y_0(i,j,b)$ at location (i,j,b) of the 3-D cube, using (10). Then calculate $\alpha(y_k|y_0)_{ss}$ in (14).
3) Generate a value u randomly from a uniform distribution $U(0,1)$. Accept y_k into the sample set Ω, if $u \leq \alpha(y_k|y_0)_{ss}$; otherwise, discard.
4) After processing the M samples by repeating (ii)–(iv), the accepted samples in Ω will be used to estimate the posterior distribution $\hat{p}(x|y_0)$ in (9), as described in Section II-C.
5) Given $\hat{p}(x|y_0)$, compute the noise-free estimate $\hat{x}_0$ of the reference pixel y_0 as the discrete conditional mean, according to (8).

III. EXPERIMENTS AND DISCUSSION

The proposed method is tested on both simulated and real HSI, in comparison with several other popular methods. In the simulated study, two benchmark images, i.e., Indian Pines image [20] and Pavia U image are used for simulating noisy hyperspectral images. The first simulation uses only the ground-truth label information of Indian Pines image, while the second is achieved by treating the PaviaU image as clean observations. Given the true image values in the simulated study, some numerical measures can be derived for performance assessment. The real image adopted is the Indian Pines image. The evaluation is primarily based on the visual inspection and some posterior measures, such as signal-to-noise ratio (SNR) and the classification accuracy using the denoised images.

Methods in all experiments include some popular denoising techniques, i.e., the Wiener filter [27], wavelet-based BayesShrink algorithm [28], and the total variation (TV) methods [29], [30]. Moreover, in order to test the influence of spectral and spatial information on the proposed method, we incorporate for comparison also three variants introduced in Section II-D, i.e., pixel-MCS, spatial-MCS, and spectral-MCS, besides the proposed SS-MCS algorithm summarized in Section II-E.

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

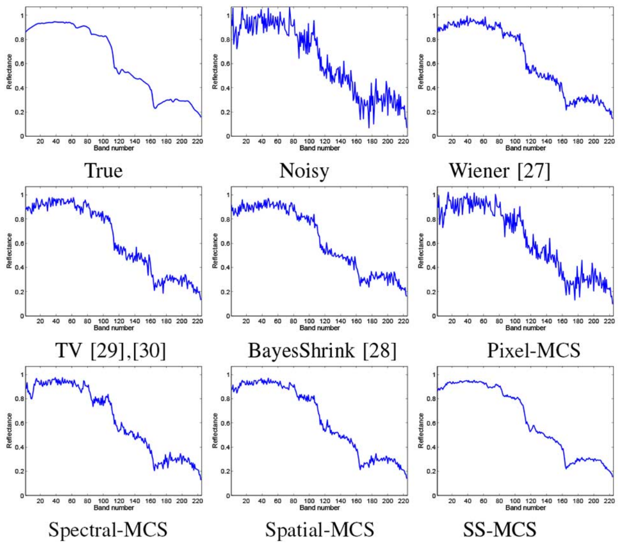


Fig. 6. Spectra at pixel (60,100) of the simulated data 1, before and after denoising by different methods.

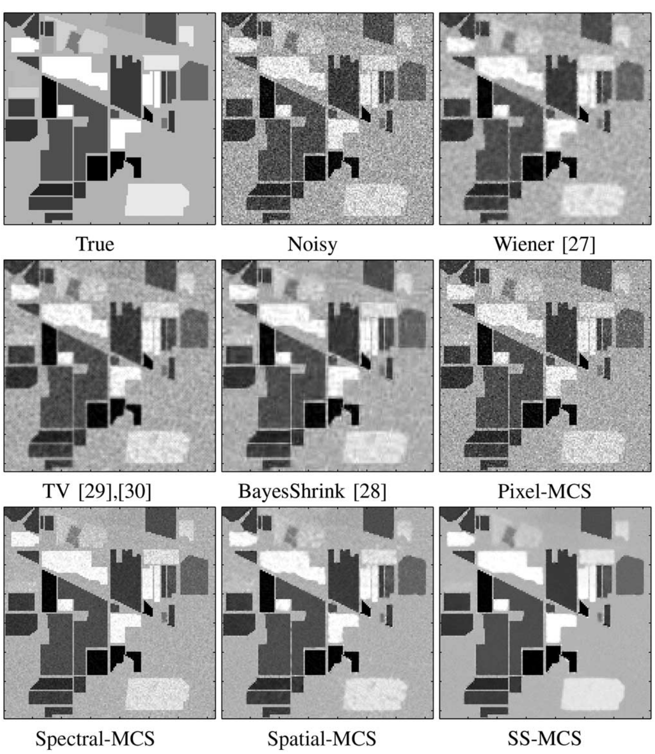


Fig. 7. Denoising results achieved by different methods, on band 30 of simulated data 1. Denoised image by the proposed SS-MCS method is the most similar one to the true image. The other methods tend to either preserve undesirable artifacts or blur image boundaries and weak signal.

In all experiments, the parameters of referenced methods are set by following the suggestions of the respective authors. For the MCS-based methods, a region-size of 3×3 is used for spatial-MCS and SS-MCS; a band width of 5 is used for spectral-MCS and SS-MCS; the search area of all methods is set to be 21×21, with approximately half of pixels being sampled; the β parameter of all methods is set to be the dimensionality of the variable used for calculating similarity likelihood, e.g., 9 for spatial-MCS.

In the proposed methods, in order to obtain region samples at the boundary areas, before obtaining samples, we perform image padding in spatial dimensions. Therefore, region-samples centered at image boundaries will be full-sized, since padded pixels can be utilized. But, we do not perform padding in the spectral dimension. So, band-stack-samples centered at the ends of spectral band are half-sized, since no padding was performed at the end of spectral band.

A. Experiments With Simulated Data

In this experiment, we simulate two noisy HSI by polluting the clean data with band-dependent Gaussian noise, with band variance being determined by the following rule:

$$\sigma_b^2 = \frac{\sum_{ij} y_{ijb}^2}{10^{SNR/10} MN} \tag{15}$$

where M and N are, respectively, the numbers of rows and columns in the image, and SNR is the signal-to-noise ratio, given which the noise variances can be calculated.

For performance evaluation, we use peak signal to noise ratio (PSNR) to measure the degree of noise removal. The PSNR of band b is formulated as

$$PSNR_b = \frac{MN}{\sum_{ij}(\hat{x}_{ijb} - x_{ijb})^2}. \tag{16}$$

We use β to measure the image detail preservation [31], which has been widely used in image denoising, e.g., [32]. The β of band b is formulated as

$$\beta_b = \frac{t(S_b - \bar{S}_b, \hat{S}_b - \bar{\hat{S}}_b)}{\sqrt{t(S_b - \bar{S}_b, S_b - \bar{S}_b)t(\hat{S}_b - \bar{\hat{S}}_b, \hat{S}_b - \bar{\hat{S}}_b)}} \tag{17}$$

where S_b and $\hat{S}_b$ are the high-pass filtered version of the clean image X_b and denoised image $\hat{X}_b$ in band b, obtained using 3×3 standard approximation of the Laplacian operator, $\bar{S}_b$ and $\bar{\hat{S}}_b$ are the mean values of S_b and $\hat{S}_b$ and

$$t(A_b, B_b) = \sum_{ij} A_{ijb} B_{ijb} \tag{18}$$

β should be close to unity for the optimal effect of detail preservation.

1) Simulated Data 1: The benchmark Indian Pines image is used for simulating the first noisy hyperspectral data. The Indian Pines image was captured by airborne visible/infrared imaging spectrometer (AVIRIS) over a vegetation area in northwestern Indiana, USA, with a spatial resolution of 20 m, consisting of 145×145 pixels of 16 ground-truth classes and 220 spectral reactance bands.

Only the labeling information of Indian Pines image is used in simulation. We substitute the pixels in 17 ground-true classes

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

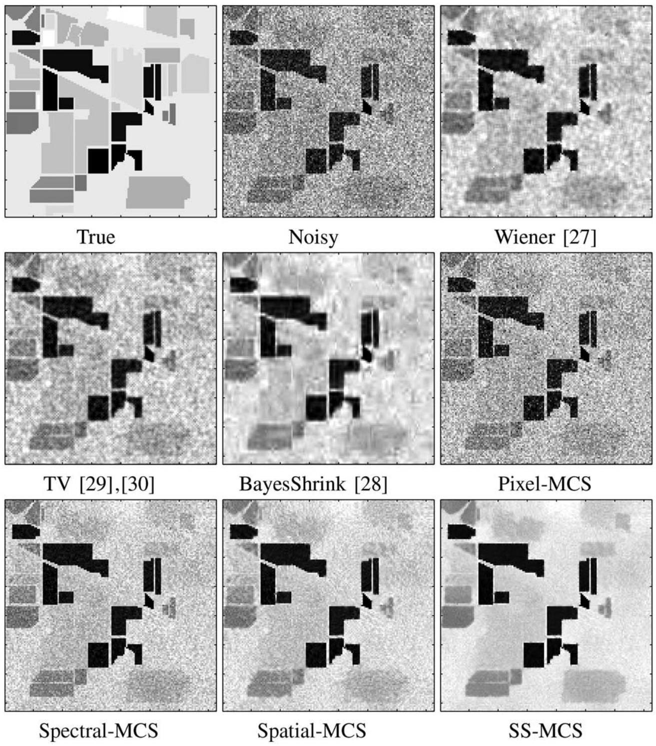


Fig. 8. Denoising results achieved by different methods, on band 150 of simulated data 1.

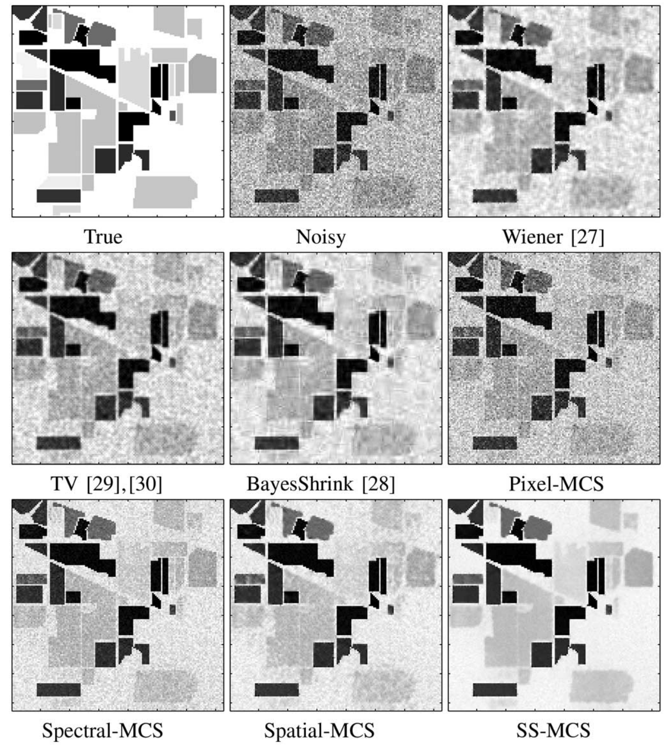


Fig. 9. Denoising results achieved by different methods, on band 200 of simulated data 1.

(the zero-labeled class are also utilized here) with 17 spectra of 224 spectral bands randomly chosen from the USGS spectra library [33] [see Fig. 2(a)]. The resulting image is considered a clean image. We then degrade this image with band-dependent Gaussian noise, whose variance in band b is determined according to (19). In this simulated study, we set $SNR = 20$. Accordingly, $\sigma_b^2(b = 1, 2, \ldots, B)$ varies, depending on the signal strength of band [see Fig. 2(b) for the simulated and estimated noise variances of all bands].

After testing all methods on the simulated image, in Fig. 3, we plot the $PSNR_b$ and β_b as a function of band number, for each method. As we can see, for both measures, the lines of SS-MCS are above those of the other methods, indicating that SS-MCS outperforms the other methods in terms of both noise removal and detail preservation. The second best method seems to be spatial-MCS, followed closely by spectral-MCS. Due to its sensitivity to noise, pixel-MCS achieves the lowest PSNR line among all methods. The Wiener and BayesShrink methods achieve comparable performances, but BayesShrink has large variation on β. The TV method produces the lowest β line, which is probably because TV tends to oversmooth image. The lines of all denoising methods are above the lines of the noisy image.

Table I shows the statistics of different methods on $PSNR$ and β. They are basically consistent with Fig. 3. SS-MCS achieved the highest mean and median $PSNR$ and β values, followed by spatial-MCS and spectral-MCS. Pixel-MCS produces very low values. Wiener and BayesShrink achieve comparable statistics, which are better than those of TV method.

Figs. 4–6 show the spectra of clean, noisy, and denoised images produced by different methods, at three different locations of the image. They suggest that SS-MCS produces the most similar spectra to the true images. The spectra of spatial-MCS also seem to have less fluctuation than other methods. Pixel-MCS fails to significantly reduce the noise.

Figs. 7–9 display the clean, noisy, and denoised images of three different bands achieved by different methods. As we can see, the denoised image by SS-MCS is the most similar one to the true image. Spatial-MCS also achieves good balance between noise removal and edge preservation. But it still tends to blur the boundaries of some weak blocks, comparing with spectral-MCS, which, on the other hand, tends to preserve undesirable noise-like artifacts. Pixel-MCS does not remove noise efficiently. Comparing with Wiener and TV, BayesShrink achieves better balance between noise removal and detail preservation. TV method tends to blur the boundaries of blocks. Wiener preserves boundaries of dark blocks very well, but erases those of bright ones.

An understanding of the SS-MCS method as a function of the size of the 3-D cube is derivable. That is, we would like to know how the variation of the spatial region size and spectral bandwidth affect the similarity likelihood measurement. For this purpose, in Fig. 10, we plot the performance of SS-MCS, measured by $PSNR$ and β values, as functions of bandwidth and region size, respectively. The increase of the size of the 3-D cube sample, either by band width or by region size, would lead to first increase, then decrease in numerical measures. The peak bandwidth for β is 7, and for $PSNR$ it is 13, while the peak region sizes are 5×5 for both measures. The size of 3-D

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

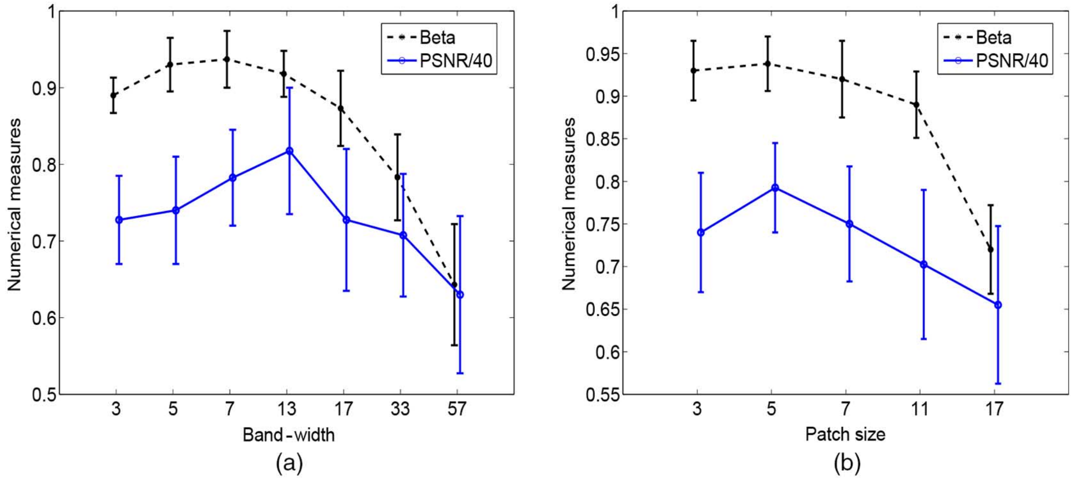


Fig. 10. (a) Error bar plot of numerical measures achieved by SS-MCS, i.e., $PSNR$ and β, as functions of bandwidth of the 3-D cube sample, on the simulated data 1, with all the other parameters being fixed at the default values. (b) Plot of numerical measures as functions of the region sizes of 3-D cube sample, with all the other parameters fixed at the default values. The center of the bar represents the mean value of numerical measures over all bands, while the height represents standard deviation.

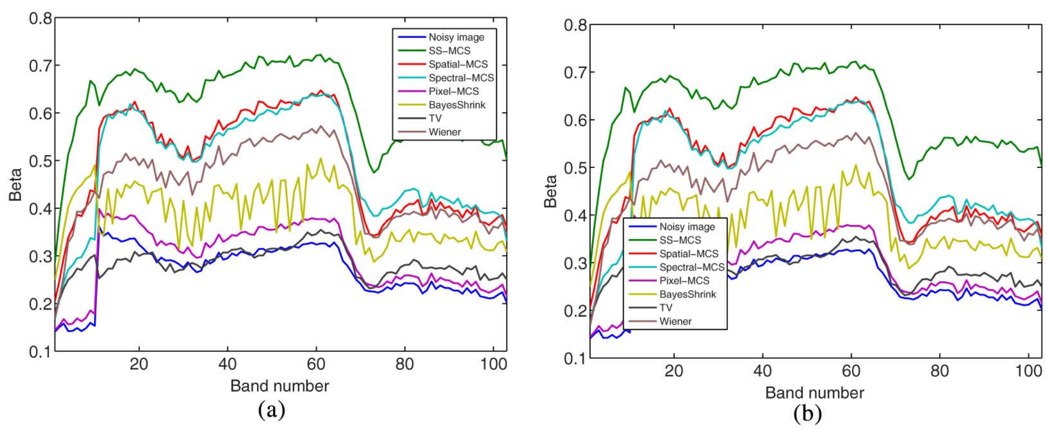


Fig. 11. (a) PSNR and (b) β achieved by different methods, at different bands of the simulated data 2. Proposed SS-MCS method achieves the highest values in most bands, significantly outperforming spatial-MCS and spectral-MCS, which are usually the second and third best.

cube sample determines the extent of spatial–spectral correlation considered when calculating similarity likelihood. Bigger size explains larger scale similarities, thereby may be more restrict toward accepting samples. Given an insufficient number of accepted samples, the denoising performance may deteriorate. That may explain why decreased numerical measures are observed with the increase in cube size.

2) Simulated Data 2: The second data is simulated based on the Pavia University (PaviaU) image, which is an urban image, centered at the University of Pavia, consisting of 610×340 pixels, acquired by the reflective optics system imaging spectrometer, with a spatial resolution of 1.3 m, consisting of 103 spectral bands after removing 12 noisy bands. In this experiment, the image used for simulation is a 128×128 subset of the PaviaU image, containing both homogeneous area and structured area.

We simulate noisy image by treating the PaviaU image as a clean image, and polluting it by band-dependent Gaussian noise. Since most bands in PaviaU image have very high SNR, they can be treated as clean observations. However, for band 1–10 that contain higher noise level than the other bands, we apply a 3×3 mean filter on them to reduce noise, before using them as clean image. The variances of the simulated noise are band-dependent, generated according to (19) by using $SNR = 10$.

The true image used here is more heterogeneous in the spatial domain than the simulated data 1 in Section III-A1. Moreover, the SNR is also lower. Therefore, the simulated image is more challenging for the denoising methods.

In Fig. 11, the results are quite consistent with the results in Section III-A1. The proposed SS-MCS method achieves the highest values in most bands, significantly outperforming spatial-MCS and spectral-MCS, demonstrating the importance and benefits of utilizing both spatial and spectral information in MCS. Similar to the experiment in Section III-A1, the overall performance of spatial-MCS is better than spectral-MCS, which, however, achieves higher β values on bands 70–103. The Wiener method outperforms BayesShrink on both measures. The pixel-MCS achieves the very low values on most bands.

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

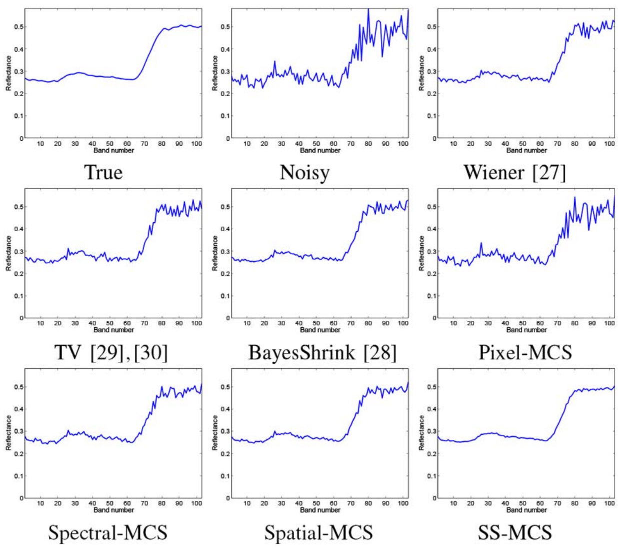


Fig. 12. Spectra at pixel (38,42) of the simulated data 2, before and after denoising by different methods. The proposed SS-MCS method produce smooth spectra that are the most similar to the true ones. The noise has higher variances in band 80–103. Nevertheless, SS-MCS method handles this noise heterogeneity issue very well. But the other methods tend to preserve intense noise on band 80–130.

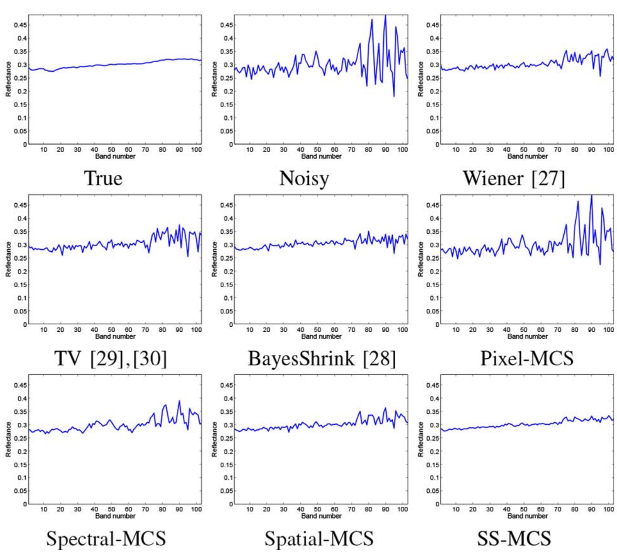


Fig. 13. Spectra at pixel (67,39) of the simulated data 2, before and after denoising by different methods.

In Table I, according to the mean $PSNR$ values, SS-MCS outperforms spatial-MCS by 5.3%, and spectral-MCS by 7.4%. According to mean β value, they are 19.7% and 20.6%, respectively. In both measures, the median values produce the same rank of methods as the mean values. In terms of mean $PSNR$ value, Wiener slightly outperforms spectral-MCS, which nevertheless achieves higher mean β value.

In Figs. 12 and 13, similar to Section III-A1, the proposed SS-MCS method produce smooth spectra that are the most similar to the true ones. The noise has higher variances in band 80–103. Nevertheless, SS-MCS method handles this noise heterogeneity issue very well. But the other methods tend to preserve intense noise on band 80–130.

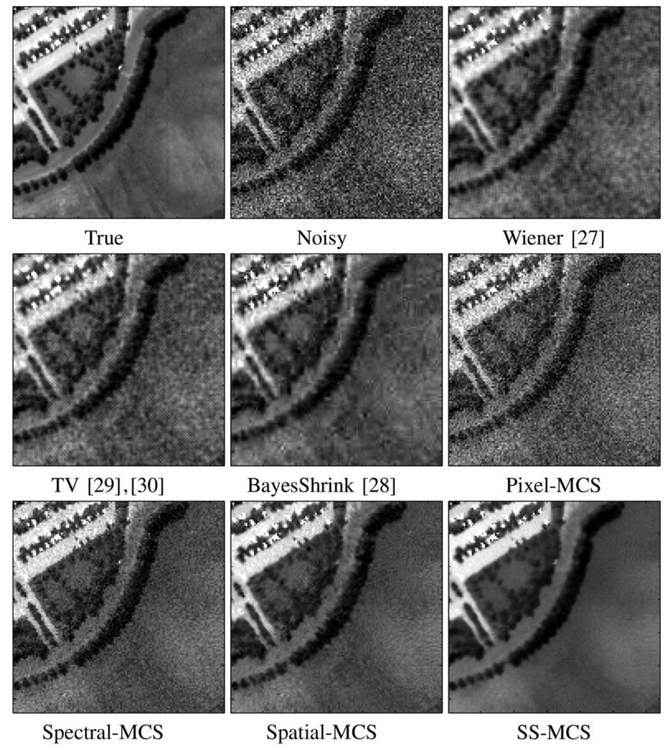


Fig. 14. Denoising results achieved by different methods, on band 30 of simulated data 2. The proposed SS-MCS method is capable of preserving image details very well, e.g., trees and constructions in up-left area are clearly delineated, in the meantime, reducing greatly the noise artifacts, e.g., the homogeneous area is sufficiently smoothed. The other methods tend to either keep undesirable artifacts or remove scene signals.

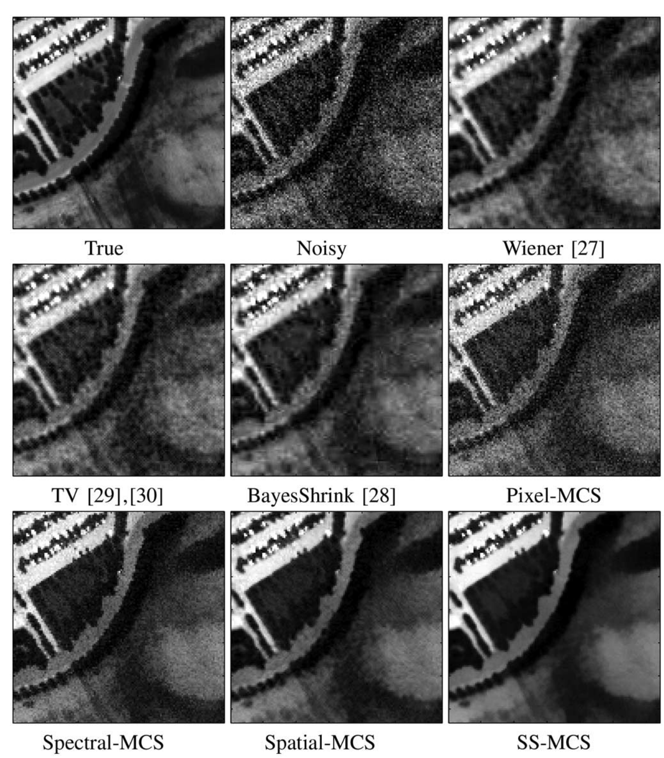


Fig. 15. Denoising results achieved by different methods, on band 60 of simulated data 2.

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

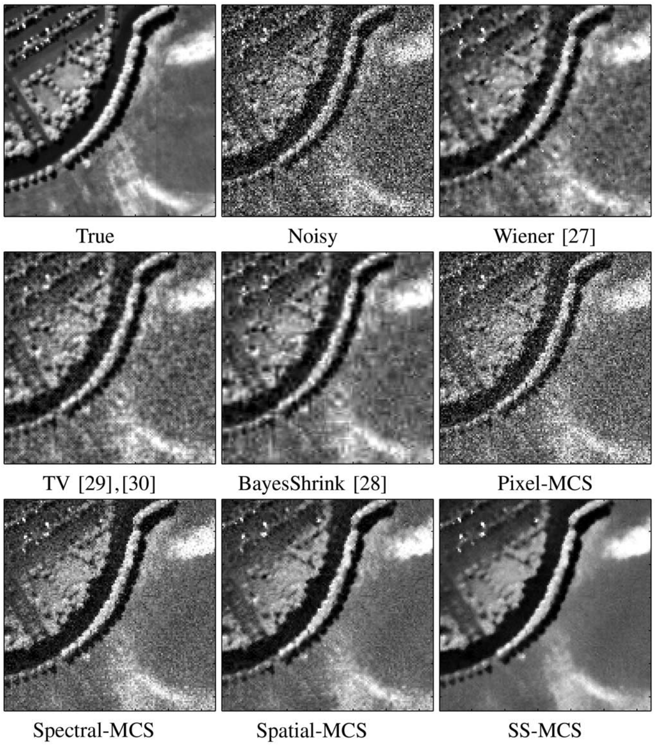


Fig. 16. Denoising results achieved by different methods, on band 90 of simulated data 2.

Figs. 14–16 also demonstrate that SS-MCS is capable of preserving image details very well, e.g., trees and constructions in up-left area are clearly delineated. Also, noise artifacts are greatly removed, e.g., the homogeneous area is sufficiently smoothed. The other methods tend to either keep undesirable artifacts or remove scene signals.

B. Experiments With Real Hyperspectral Data

In this experiment, all methods are tested on the benchmark Indian Pines image, introduced in Section III-A1. The image contains 220 spectral bands. Some bands contain more noise than the other bands. In this experiment, all bands are used for testing, in order to examine the adaptability of the proposed method.

1) SNR and Visual Evaluation: We use the SNR as numerical measure of denoising performance. The SNR of band b is expressed as follows [34]–[36]:

$$SNR_b = 10\log_{10}\frac{\sum_{ij}\hat{x}_{ijb}^2}{\sum_{ij}(\hat{x}_{ijb}-m_b)^2} \tag{19}$$

where $\hat{x}_{ijb}$ is the denoised pixel value and m_b is the mean value of $\{\hat{x}_{ijb}\}$ in a homogeneous area. The estimation of SNR relies on a homogeneous area. Different selections of the homogeneous area may lead to different SNR values. In order to reduce this variability, we adopt the class labels for identifying homogeneous areas. Since pixels in the same class are more similar to each other than pixels in different classes, a homogeneous area is set to contain all pixels belonging to the same class. SNR_b is estimated as the mean of SNR values estimated using labels of different classes. Since all pixels are used for estimating SNR, this approach is capable of reducing the bias caused by a particular selection of homogeneous area in the image.

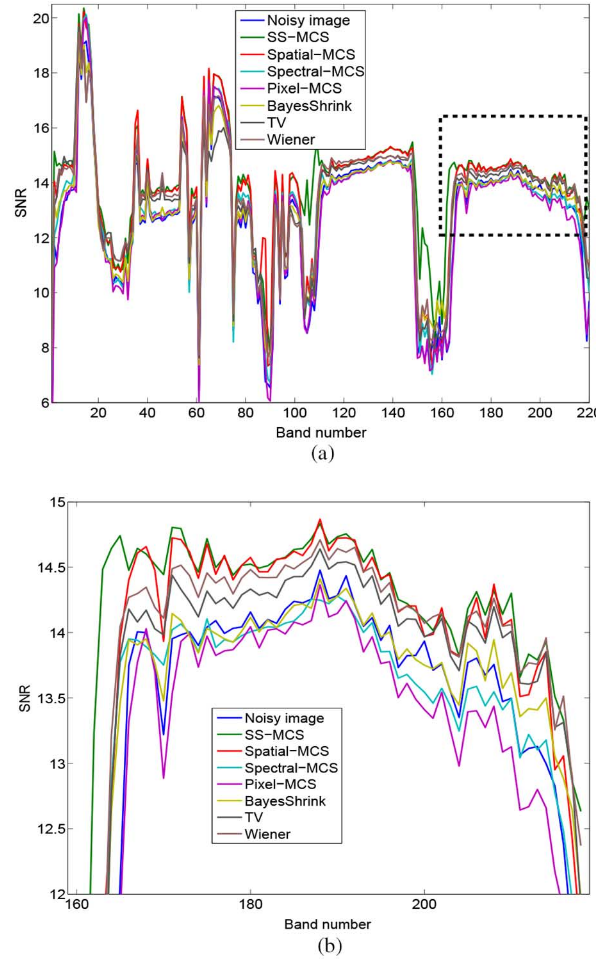


Fig. 17. (a) SNR achieved by different methods on Indian Pines image. (b) Zoom-in plot of SNR in the highlighted region in (a). The proposed SS-MCS achieves the highest SNR values on most bands, followed by spatial-MCS and Wiener. On high-noise-level bands, i.e., 104–108 and 150–160, SS-MCS performs especially better than the other methods.

TABLE II
SNR ACHIEVED BY DIFFERENT METHODS ON INDIAN PINES IMAGES (BEST RESULTS ARE HIGHLIGHTED IN BOLD)

	SNR		
	Mean	Median	Std.
Original	12.8	13.4	2.5
BayesShrink	13.1	13.7	2.1
TV	13.3	14.0	2.2
Wiener	13.4	14.1	2.2
Pixel-MCS	12.8	13.4	2.6
Spectral-MCS	13.1	13.6	2.4
Spatial-MCS	13.7	14.2	2.3
SS-MCS	**14.0**	**14.3**	**2.0**

Fig. 17 shows the plot of SNR achieved by different methods as a function of band number. It indicates that SS-MCS achieves the highest SNR values on most bands, followed by spatial-MCS and Wiener. On high-noise-level bands, i.e., 104–108 and 150–160, SS-MCS performs especially better than the other methods. Pixel-MCS achieves higher SNR values when

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

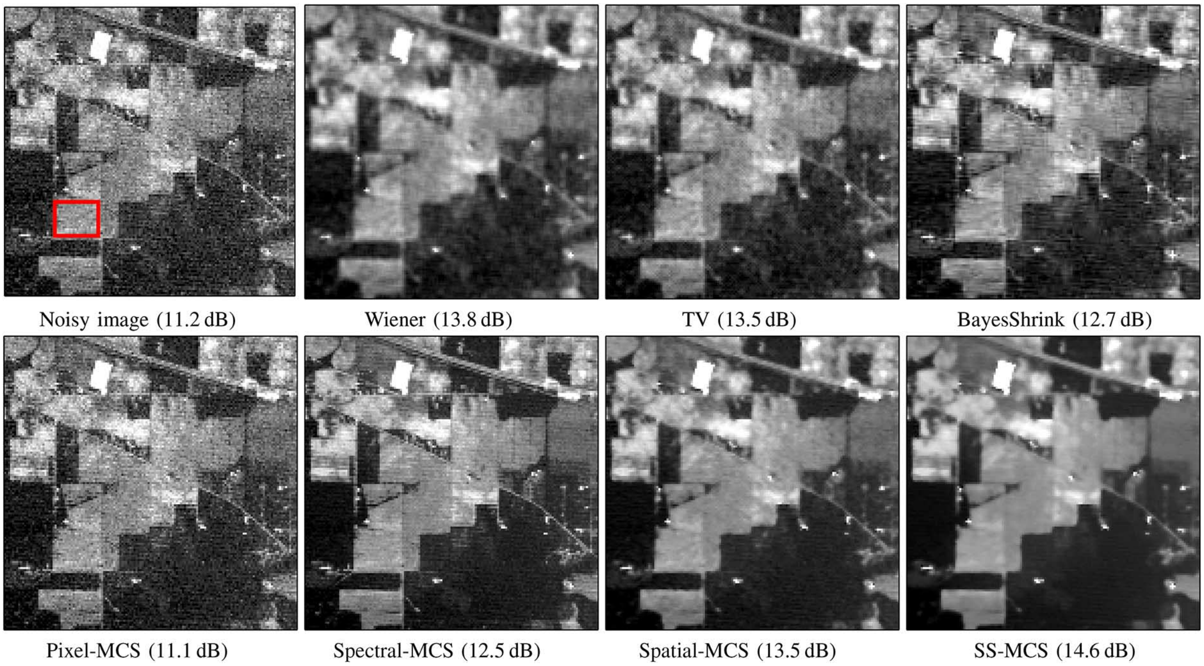


Fig. 18. Denoising results achieved by different methods, on band 3 of Indian Pines image. The highlighted rectangular in noisy image shows the homogeneous area used for calculating SNR of all bands in Fig. 1. The SNR values are shown in parenthesis. The proposed SS-MCS method increases the SNR of the noisy image by 3.4 dB, the largest among all methods. Moreover, it reduces both random noise and systematic strip noise efficiently, while also preserving image details, e.g., bright point targets very well. Other methods tend to either keep undesirable artifacts or oversmooth the image.

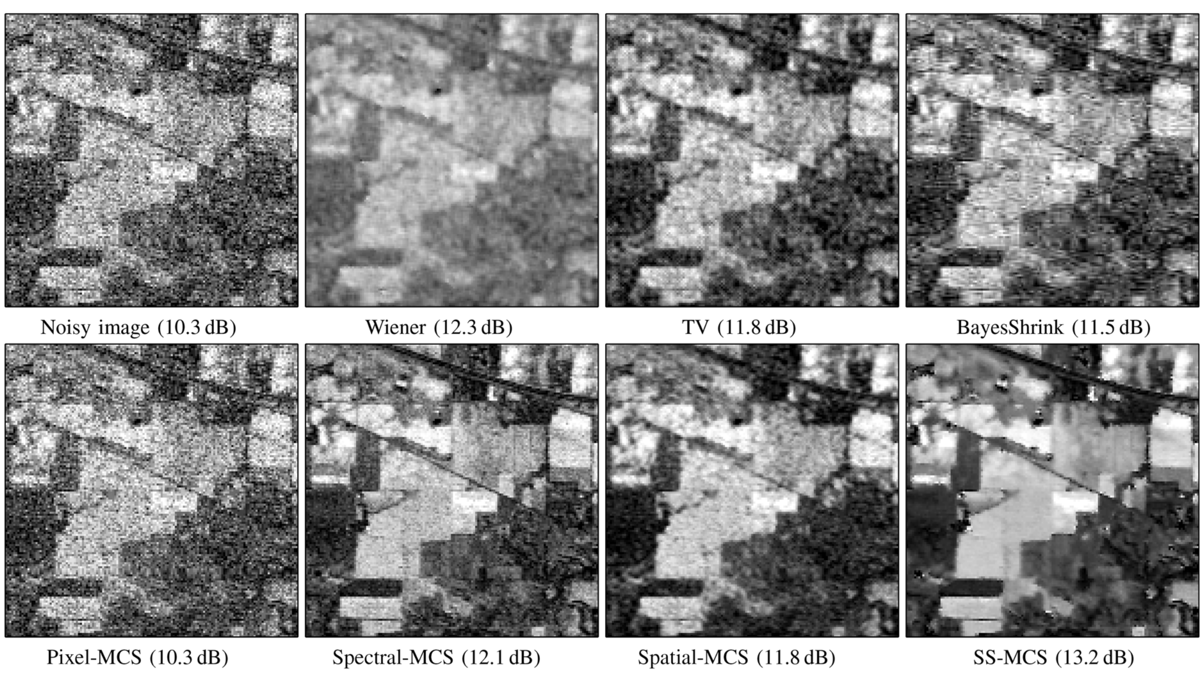


Fig. 19. Denoising results achieved by different methods, on band 103 of Indian Pines image. The SNR values are shown in parenthesis. The proposed SS-MCS method increases the SNR of noisy image by 2.9 dB. In the image denoised by this method, the image details, such as point targets and line targets, are well preserved and highlighted, while the noise in homogeneous areas is largely removed.

the SNR of noisy image is high, e.g., on bands 120–140. Wiener tends to outperform TV and BayesShrink. All denoising methods are able to increase the SNR of the noisy image.

The statistics of SNR in Table II suggests consistent results. Comparing with the noisy image, SS-MCS increases the SNR by 1.15 db on average, which is about 9.0% of the SNR of the noisy image. Moreover, SS-MCS achieves lowest variation across image bands than the other algorithms.

Figs. 18–20 show the denoised images at three bands, achieved by different methods. In Fig. 18, the proposed SS-MCS method increases the SNR of the noisy image by 3.4 dB, the largest among all methods. It reduces both random noise

Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

TABLE III
OA, AA, AND KAPPA COEFFICIENT (KAPPA) ACHIEVED BY DIFFERENT METHODS ON THE INDIAN PINES IMAGE
(BEST RESULTS ARE HIGHLIGHTED IN BOLD)

	SVM			LDA		
	OA (%)	AA (%)	Kappa	OA (%)	AA (%)	Kappa
Original	76.7	64.7	0.733	75.4	68.2	0.721
BayesShrink	85.3	76.3	0.832	79.2	77.1	0.763
TV	87.1	81.3	0.852	86.0	83.7	0.841
Wiener	89.5	81.4	0.881	87.3	85.7	0.856
Pixel-MCS	78.3	65.5	0.752	75.3	69.1	0.718
Spectral-MCS	82.8	72.6	0.804	87.0	85.0	0.853
Spatial-MCS	89.4	81.3	0.879	86.7	82.0	0.849
SS-MCS	**93.7**	**85.5**	**0.928**	**91.8**	**88.5**	**0.906**

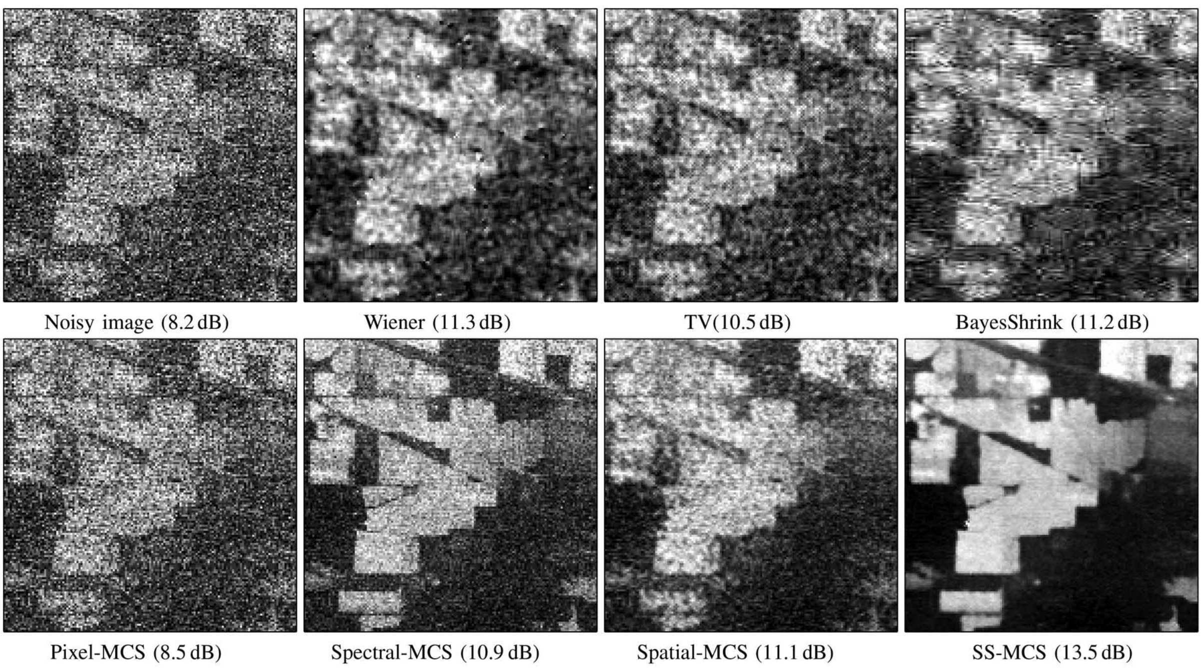


Fig. 20. Denoising results achieved by different methods, on band 219 of Indian Pines image. The SNR values are shown in parenthesis. The proposed SS-MCS method increases the SNR of noisy image dramatically by 5.3 dB. Moreover, it recovers the scene signal from intense noise pollution. Using information in adjacent channels, spectral-MCS also highlights the signals, but also preserves large amount of noise.

and systematic strip noise efficiently, while also preserving image details, e.g., bright point targets very well. Other methods tend to either keep undesirable artifacts or oversmooth the image. In Fig. 19, the SS-MCS method increases the SNR of noisy image dramatically by 2.9 dB. The image denoised by this method demonstrates fine image details, i.e., point targets and line targets, as well as smooth and clean homogeneous areas. In Fig. 20, SS-MCS increases the SNR of noisy image dramatically by 5.3 dB. Moreover, it recovers the scene signal from intense noise pollution. Using information in adjacent channels, spectral-MCS also highlights the signals, but preserves large amount of noise.

2) Classification-Aided Assessment: The denoised images achieved by different methods are used for performing supervised classification, in order to shed light on the relative performance of denoising methods. However, it is important to keep in mind that other image processing tasks, such as target detection and spectral unmixing, could also benefit from the denoising operation. In this experiment, we select two classifiers, i.e., SVM [37] and linear discriminant analysis (LDA) [38], for performing supervised classification. SVM represents the discriminative machine learning technique, while LDA is a classical statistical generative model. All the 220 bands in the denoised images are used to feed the classifiers. In the training stage, 10% of the labeled pixels in each of the 16 classes are randomly selected as training samples. The rest of the pixels are used for testing. The numerical measures include overall accuracy (OA), averaged accuracy (AA), and the Kappa coefficient.

Table III shows the statistics of both classifiers on the denoised images of different denoising methods. Overall, the proposed SS-MCS method performs better than the other methods in terms of all three measures, on both classifiers. It achieves OA of 93.7% on SVM and 91.8% on LDA, outperforming the second best method, i.e., Wiener, by 4.4 percentage points on average, and the noisy image by 16.7 percentage points on average. In terms of OA and Kappa coefficient, spatial-MCS ranks third according to SVM, but

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

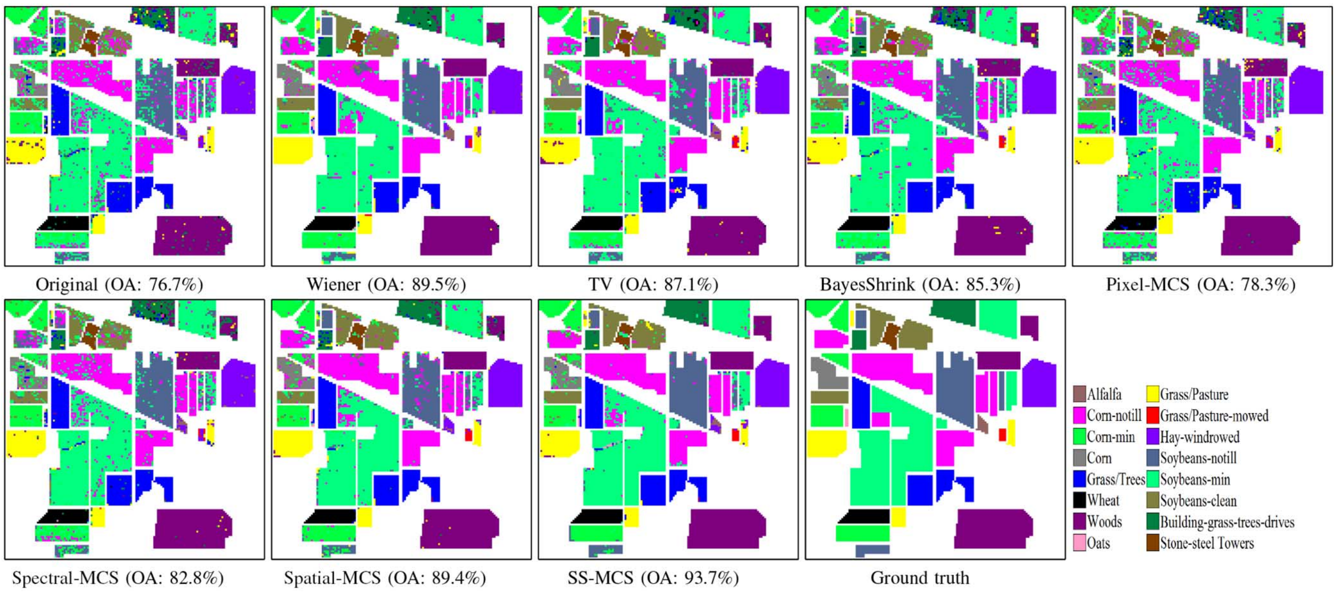


Fig. 21. Classification maps achieved by different methods on the Indian Pines image. The overall accuracies are included. Classification on the orignal image tends to produce intense artifacts in the classification map. All denoised images lead to maps of reduced artifacts to different degrees. Generally speaking, the maps produced by SS-MCS demonstrate less inner-class artifacts than all the other denoising methods.

fourth according to LDA. In contrast, spectral-MCS ranks sixth according to SVM, but third according to LDA. All denoised methods are able to increase the classification accuracy over the noisy image. Pixel-MCS achieves accuracies that are only slightly higher than those of the noisy image. SVM produces higher OA and Kappa coefficients, while LDA produces higher AA.

The classification maps in Fig. 21 indicate consistent results with the statistics in Table III. Classification on the original image tends to produce intense artifacts in the classification map, due to the existence of noise in images. All denoised images lead to maps of reduced artifacts to different degrees. Generally speaking, the maps produced by SS-MCS demonstrate less inner-class artifacts than all the other denoising methods.

IV. Conclusion

This paper presented a novel hypersepctral image denoising technique based on MCS and spectral–spatial similarity likelihood. The stochastic MCS was used to estimate the posterior probability. Also, in order to utilize both the spatial and spectral information in hyperspectral image, a spatial–spectral probabilistic similarity measurement based on band-dependent noise distribution was used to calculate the acceptance probability in MCS. The proposed method was tested on two simulated hyperspectral images and the benchmark Indian Pines image, in comparison with several other methods. Both the numerical and visual results demonstrate that the proposed method has strong capability in terms of both noise removal and image details preservation.

Accurate estimation of noise variances is a difficult issue, which relies on a system that is capable of fully capturing the spatial–spectral correlation effect in HSI. In this paper, we adopt a simple approach by identifying a homogeneous area in the image. However, it will potentially be difficult to select a region where the variation of pixel values is only due to noise when the spatial resolution of the image is very low. Therefore, in our future research, we will focus on developing more advanced methods for noise level estimation in order to further improve the denoising performance.

References

[1] I. Atkinson, F. Kamalabadi, and D. L. Jones, "Wavelet-based hyperspectral image estimation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'03)*, 2003, vol. 2, pp. 743–745.

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

[3] G. Chen and S.-E. Qian, "Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis," *Can. J. Remote Sens.*, vol. 34, no. 5, pp. 447–454, 2008.

[4] G. Chen and S.-E. Qian, "Denoising and dimensionality reduction of hyperspectral imagery using wavelet packets, neighbour shrinking and principal component analysis," *Int. J. Remote Sens.*, vol. 30, no. 18, pp. 4889–4895, 2009.

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

[6] G. Chen, T. D. Bui, and A. Krzyzak, "Denoising of three-dimensional data cube using bivariate wavelet shrinking," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 25, no. 03, pp. 403–413, 2011.

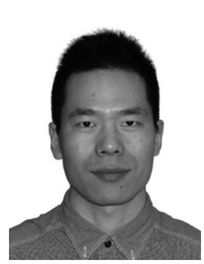
[7] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 438–441, Dec. 2002.

[8] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using 3D wavelets," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'12)*, 2012, pp. 1349–1352.

[9] S. Bourguignon, D. Mary, and E. Slezak, "Sparsity-based denoising of hyperspectral astrophysical data with colored noise: Application to the muse instrument," in *Proc. 2nd Workshop Hyperspectral Image Signal Process. Evol. Remote Sens. (WHISPERS)*, 2010, pp. 1–4.

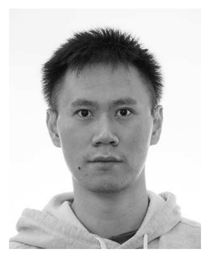
Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

[10] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.
[11] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.
[12] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel non-negative tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.
[13] R. D. Phillips, C. E. Blinn, L. T. Watson, and R. H. Wynne, "An adaptive noise-filtering algorithm for aviris data with implications for classification accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3168–3179, Sep. 2009.
[14] J. Martn-Herrero, "Anisotropic diffusion in the hypercube," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1386–1398, May 2007.
[15] N. Renard and S. Bourennane, "Improvement of target detection methods by multiway filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2407–2417, Aug. 2008.
[16] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of classification for hyperspectral images based on tensor modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 801–805, Oct. 2010.
[17] A. Wong, A. Mishra, W. Zhang, P. Fieguth, and D. A. Clausi, "Stochastic image denoising based on markov-chain monte carlo sampling," *Signal Process.*, vol. 91, no. 8, pp. 2112–2120, 2011.
[18] J. Glaister, A. Wong, and D. A. Clausi, "Despeckling of synthetic aperture radar images using monte carlo texture likelihood sampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1238–1248, Feb. 2014.
[19] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.
[20] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, vol. 29. Hoboken, NJ, USA: Wiley, 2005.
[21] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 488–498, Apr. 2013.
[22] B.-C. Gao, "An operational method for estimating signal to noise ratios from data acquired with imaging spectrometers," *Remote Sens. Environ.*, vol. 43, no. 1, pp. 23–33, 1993.
[23] R. Roger and J. Arnold, "Reliably estimating the noise in aviris hyperspectral images," *Int. J. Remote Sens.*, vol. 17, no. 10, pp. 1951–1962, 1996.
[24] L.-R. Gao, B. Zhang, X. Zhang, W.-J. Zhang, and Q.-X. Tong, "A new operational method for estimating noise in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 83–87, Jan. 2008.
[25] L. Gao, B. Zhang, J. Wen, and Q. Ran, "Residual-scaled local standard deviations method for estimating noise in hyperspectral images," in *Proc. Int. Symp. Multispectral Image Process. Pattern Recognit.*, Int. Soc. Optics Photonics, 2007, p. 678713.
[26] P. Fieguth, Statistical Image Processing and Multidimensional Modeling. New York, NY, USA: Springer, 2011.
[27] Wiener. *2-D adaptive noise-removal filtering* [Online]. Available: http://www.mathworks.com/help/images/ref/wiener2.html, accessed on Sep. 30, 2014.
[28] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
[29] G. Gilboa. (2003, Sep.). *Total Variation Denoising* [Online]. Available: http://visl.technion.ac.il/ gilboa/PDE-filt/tv_denoising.html
[30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D Nonlinear Phenom.*, vol. 60, no. 1, pp. 259–268, 1992.
[31] F. Sattar, L. Floreby, G. Salomonsson, and B. Lovstrom, "Image enhancement based on a nonlinear multiscale method," *IEEE Trans. Image Process.*, vol. 6, no. 6, pp. 888–895, Jun. 1997.
[32] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 8, pp. 1773–1784, Aug. 2003.
[33] R. N. Clark *et al*. (2007). *Usgs Digital Spectral Library splib06a* [Online]. Available: http://speclab.cr.usgs.gov/spectral-lib.html
[34] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
[35] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[36] J. M. Nascimento and J. M. Bioucas Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[37] J. M. Moguerza and A. Muñoz, "Support vector machines with applications," *Stat. Sci.*, vol. 21, pp. 322–336, 2006.
[38] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, vol. 2. no. 1, New York, NY, USA: Springer, 2009.



Linlin Xu (M'14) received the B.Eng. and M.Sc. degrees in geomatics engineering from China University of Geosciences, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in geography from the University of Waterloo, Waterloo, ON, Canada.

His research interests include the areas of hyperspectral and SAR image processing.



Fan Li received the B.S. degree in geographical information system from Sun Yat-sen University, Guangzhou, China, and the M.E. degree in pattern recognition and intelligent system from Wuhan University, Wuhan, China. He is currently working toward the Ph.D degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada.

His research interests include the application of machine learning and computer vision techniques to the remote sensing field.

Alexander Wong (M'05) received the B.A.Sc. degree in computer engineering, the M.A.Sc. degree in electrical and computer engineering, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005, 2007, and 2010, respectively.

He is currently an Assistant Professor with the Department of Systems Design Engineering, University of Waterloo. He has authored refereed journal and conference papers, as well as patents, in various fields, such as computer vision, graphics, image processing, multimedia systems, and wireless communications. His research interests include image processing, computer vision, pattern recognition, and cognitive radio networks, with a focus on biomedical and remote sensing image processing and analysis such as image registration, image denoising and reconstruction, image super-resolution, image segmentation, tracking, and image and video coding and transmission.

Dr. Wong was the recipient of an Outstanding Performance Award, an Engineering Research Excellence Award, an Early Researcher Award from the Ministry of Economic Development and Innovation, a Best Paper Award by the Canadian Image Processing and Pattern Recognition Society, and the Alumni Gold Medal.

David A. Clausi (S'93–M'96–SM'03) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1990, 1992, and 1996, respectively.

After completing his Ph.D., he worked in the medical imaging field with Mitra Imaging Inc., Waterloo, ON, Canada. He worked as an Assistant Professor of Geomatics Engineering with the University of Calgary, Calgary, AB, Canada, in 1997. He returned to his alma mater in 1999 and was awarded tenure and promotion to Associate Professor in 2003. He is an active interdisciplinary and multidisciplinary researcher. He has an extensive publication record, publishing refereed journal and conference papers on remote sensing, computer vision, algorithm design, and biomechanics. His research interest include the automated interpretation of synthetic aperture radar (SAR) sea ice imagery, in support of operational activities of the Canadian Ice Service.

Dr. Clausi is the recipient of numerous scholarships, conference paper awards, and two Teaching Excellence Awards.