Noise Removal From Hyperspectral Image With Joint Spectral–Spatial Distributed Sparse Representation

Jie Li, *Student Member, IEEE*, Qiangqiang Yuan, *Member, IEEE*,
Huanfeng Shen, *Senior Member, IEEE*, and Liangpei Zhang, *Senior Member, IEEE*

***Abstract*—Hyperspectral image (HSI) denoising is a crucial preprocessing task that is used to improve the quality of images for object detection, classification, and other subsequent applications. It has been reported that noise can be effectively removed using the sparsity in the nonnoise part of the image. With the appreciable redundancy and correlation in HSIs, the denoising performance can be greatly improved if this redundancy and correlation is utilized efficiently in the denoising process. Inspired by this observation, a noise reduction method based on joint spectral–spatial distributed sparse representation is proposed for HSIs, which exploits the intraband structure and the interband correlation in the process of joint sparse representation and joint dictionary learning. In joint spectral–spatial sparse coding, the interband correlation is exploited to capture the similar structure and maintain the spectral continuity. The intraband structure is utilized to adaptively code the spatial structure differences of the different bands. Furthermore, using a joint dictionary learning algorithm, we obtain a dictionary that simultaneously describes the content of the different bands. Experiments on both synthetic and real hyperspectral data show that the proposed method can obtain better results than the other classic methods.**

***Index Terms*—Hyperspectral image (HSI), noise reduction, sparse representation.**

I. Introduction

HYPERSPECTRAL imaging technology has attracted much research interest in the field of remote sensing because of its ability to provide abundant spatial and spectral information, which can offer different reflectance information to identify specific materials in a scene [1], [2], such as in land-surface classification [3], object recognition [4], [5], and change detection [6]. In practical applications, hyperspectral images (HSIs) are unavoidably contaminated with noise in the process of access and transmission, due to the imaging equipment and external environment [7]. A noisy image hinders the discrimination of the materials in the HSI. It is therefore important to reduce the noise before HSI analysis and interpretation.

Various HSI denoising techniques have been proposed in the last decades [8]–[32]. The simplest method of HSI denoising is to apply a traditional 2-D denoising algorithm in the spatial dimension in a band-by-band manner. However, these methods have the drawback that only the noise in the spatial dimension is removed, and the strong correlation in the spectral dimension is ignored. As a result, artifacts and distortion are produced in the spectral dimension, and the spectral signature will not be effectively restored. Therefore, a number of advanced methods jointly utilizing the spatial and spectral information have been proposed, which can be mainly classified into two categories: transform-domain-based methods [8]–[14] and spatial-domain-based methods [15]–[32].

The transform-domain-based methods attempt to separate clear signals from the noisy data by various transformations, such as principal component analysis, frequency transform, or wavelet transform. Atkinson *et al.* [8] proposed an estimator utilizing discrete Fourier transform to decorrelate the signal in the spectral domain, and a wavelet transform was used for the spatial filtering. In [9] and [10], the data were decorrelated using principal component analysis, and the noise was removed in the low-energy channels through 2-D discrete wavelet shrinkage. Wavelet-transform-based methods [8]–[11] have been also widely used in HSI denoising, as they can exploit interband correlations, and denoising can be achieved by thresholding on the wavelet coefficients. To preserve the image details in a low noise level case, reasonable assumption or prior knowledge [12]–[14] can be also applied in a wavelet method. For example, Scheunders and Backer [14] made use of Gaussian scale mixtures of the wavelet domain and introduced a noise-free image as extra prior information to enhance the quality of the image. Nonetheless, these methods are sensitive to the selection of the transform function and cannot consider the differences in the geometrical characteristics of HSIs.

Without the complex mathematical transformation, the spatial-domain-based methods seek to reduce the noise of HSIs by constructing a multidimensional spatial–spectral denoising model, and they can be further classified into filter- and regularization-based algorithms. The filter-based HSI denoising algorithms usually consider HSI data as a 3-D array and then separate the noise from the signal by multidimensional analysis

Manuscript received February 14, 2016; revised March 30, 2016; accepted May 3, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 41401383 and Grant 41422108, by the Hong Kong Scholars Program under Grant XJ2014009, and by the Cross-disciplinary Collaborative Teams Program for Science, Technology and Innovation of the Chinese Academy of Sciences. *(Corresponding author: Qiangqiang Yuan.)*

J. Li is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: aaronleecool@hotmail.com).

Q. Yuan is with the School of Geodesy and Geomatics and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: yqiang86@gmail.com).

H. Shen is with the School of Resource and Environmental Science and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

L. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

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/TGRS.2016.2564639

[15]–[22]. For example, the multidimensional filtering method based on Tucker tensor decomposition [15] reduces noise by a tensor decomposition strategy. Following this work, different tensor decomposition models have been used in HSI denoising, such as multidimensional Wiener filtering (MWF) [16], the use of the kernel trick in Tucker decomposition [18], and the rank-1 tensor decomposition (R1TD) algorithm [19]. These tensor-decomposition-based methods jointly take into account the spectral–spatial information and effectively preserve the original spectral information. However, the problem is that the application of a core tensor and tensor product can lead to information compression and loss of spatial detail. The regularization-based approaches [23]–[32] regard the HSI restoration as an ill-posed inverse problem and remove the noise of the HSI by solving the optimization function with a prior constraint. Some examples of such approaches include anisotropic diffusion [23], Markov random fields [24], total variation [25], sparsity prior [27]–[29], and low-rank models [30], [31]. These priors control the perturbation of the solutions and play an effective role in the process of denoising. Considering the high correlation and continuity along the spectral dimension of HSIs [29]–[32], the regularization-based methods can achieve promising denoising results. Hence, how to build a regularization-based model by simultaneously utilizing the spatial and spectral information is a very important issue in the HSI denoising problem.

Redundant representations and sparsity of signals is a powerful prior model and has drawn a lot of research attention in the past decade or so. Mathematically, sparsity of signals can be described by sparse representation, which assumes that signals, such as an image, can be recovered as a sparse linear coding over a prespecified dictionary of atoms [33]. By applying the sparsity of the image, this approach has performed well in 2-D image denoising [33]–[35]. However, HSIs contain many images acquired from a continuous spectrum with a narrow bandwidth. There is strong correlation in neighboring bands [36], and similar image patches of the HSI are located simultaneously in the global space and neighboring bands. The existing sparse representation framework only utilizes the sparsity of the signals in the spatial or the spectral dimension, and large spectral distortion or spatial information loss will be produced. Therefore, it is important to develop a sparse representation framework for HSI denoising that can jointly consider the spatial and spectral information.

In this paper, a joint spectral–spatial distributed sparse representation framework is proposed to recover noise-free HSIs by fully exploiting the correlation across bands and adaptively distinguishing the differences between bands. In our work, the highly correlated bands are first clustered into a group to gather more local sparsity and redundancy. Each group is then represented with the proposed spectral–spatial distributed sparse coding. The main ideas can be summarized as follows.

1) The joint sparse approximation is designed to capture both the interband correlation and the intraband structure by the distributed sparse component, the common sparse component, and the specific sparse component. Different bands with different noise levels can then also be adaptively recovered by setting variable representation error thresholdings.
2) To obtain the corresponding sparse components, joint dictionary learning is developed to train the common dictionary and the specific dictionary. The common dictionary is used to capture the common component for depicting the similar structure across bands, whereas the specific component computed by the specific dictionary is related to the unique information within each band. The experiments confirm that the proposed method outperforms many current state-of-the-art schemes in both peak signal-to-noise ratio (PSNR) and visual perception.

The remaining parts of this paper are organized as follows. In Section II, the proposed method and the optimization process are formulated. Section III contains the experimental results and discussion. Section IV concludes this paper. Moreover, some necessary notions and preliminaries are illustrated in the following. Bold symbols with lowercase letters are reserved for vectors, and bold symbols with uppercase letters are used for matrices. For a matrix with a 3-D data array $\boldsymbol{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ with a size $I_1 \times I_2 \times I_3$, the image patch of size $\sqrt{n} \times \sqrt{n}$ from $\boldsymbol{X}$ is ordered as column vectors $\boldsymbol{x}_{ijb}$ in the ith row and the jth column of the bth band. The vector $\boldsymbol{x}_{ij}.$ is formed by stacking all the ith row and the jth column vectors from the different bands into a single column vector. $\boldsymbol{X}^T$ is the transposition of $\boldsymbol{X}$.

II. Joint Spatial and Spectral Denoising Model

A. Regularization-Based HSI Denoising Framework

Assuming that the noise type is additive, then the noise degradation process of HSI $\boldsymbol{X}$ can be modeled as

$$\boldsymbol{y}_b = \boldsymbol{x}_b + \boldsymbol{v}_b \tag{1}$$

where $\boldsymbol{x}_b$ is a clear band image from HSI $\boldsymbol{X}$, with $\boldsymbol{X} = \{\boldsymbol{x}_b\}_{b=1}^{B} \in \mathbb{R}^{M \times N \times B}$; $\boldsymbol{y}_b$ is a clear band image from HSI $\boldsymbol{Y}$, with $\boldsymbol{Y} = \{\boldsymbol{y}_b\}_{b=1}^{B} \in \mathbb{R}^{M \times N \times B}$; $\boldsymbol{Y}$ denotes the noisy image; and the additive Gaussian noise is $\boldsymbol{v}_b \sim \mathcal{N}(0, \sigma_b^2)$, $1 \le b \le B$. B denotes the number of bands, and σ_b^2 means that the noise intensity varies in the different bands. Given the observation model, the denoising problem can be defined as the following regularized least squares problem:

$$\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{X}\|_2^2 + \mu R(\boldsymbol{X}) \tag{2}$$

where $\|\boldsymbol{Y} - \boldsymbol{X}\|_2^2$ is the data fidelity term denoting the loss function between the observed data and the original clear data, and $R(\boldsymbol{X})$ is the regularization item connected with the prior of the original image. μ is the regularization parameter used to balance the contribution of the two items.

B. Basic Sparse Representation Framework

The regularization item $R(\boldsymbol{X})$ plays a very important role in the denoising process. It controls the perturbation of the solution, solves the ill-posed problem, and guarantees a stable noise-free image estimation. As discussed in Section I, sparsity-based prior models have been successfully applied in image

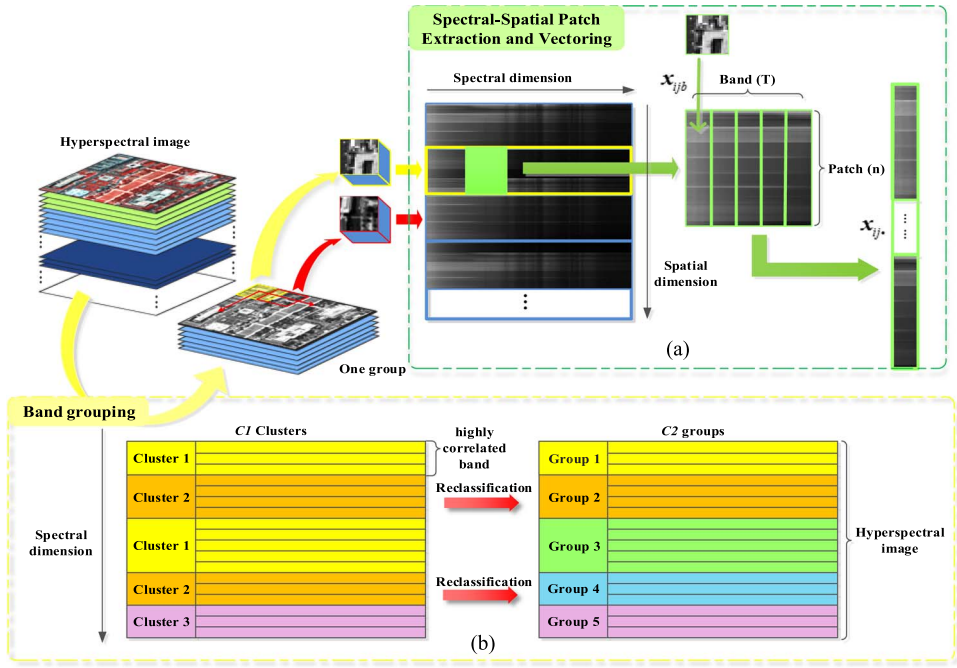


Fig. 1. (a) Method of spectral–spatial vector extraction from an HSI in the proposed model. (b) Band grouping.

denoising and are effective in removing the noise component. Mathematically, the sparse representation model assumes that column vectors $\boldsymbol{x} \in \mathbb{R}^n$ can be sparsely represented as $\boldsymbol{D\alpha}$ [33]–[35]. $\boldsymbol{D}$ is an overcomplete dictionary, such as discrete cosine bases, wavelet bases, or the dictionary learned from a set of examples, and the entries in the sparse vector $\boldsymbol{\alpha}$ are mostly close to zero. To make the ill-posed denoising problem more precise, the sparse approximation problem can be described as

$$\{\widehat{\boldsymbol{D}}, \widehat{\boldsymbol{\alpha}}\} = \arg\min_{\boldsymbol{D},\boldsymbol{\alpha}} \|\boldsymbol{x} - \boldsymbol{D\alpha}\|_2^2 + \mu\|\boldsymbol{\alpha}\|_0 \tag{3}$$

where $\|\boldsymbol{\alpha}\|_0$ is the sparse regularization, which represents the count of the nonzero entries of $\boldsymbol{\alpha}$. Although the problem in (3) is an NP-hard problem, some approximation strategies [37]–[39] can be used to guarantee an effective solution, such as basis pursuit [38] or orthogonal matching pursuit (OMP) [39]. However, the band-by-band approach used in the basic sparse representation framework for HSI denoising ignores the sparsity of the spectral information and therefore leads to noise residuals in the spectral dimension. For HSIs, our ultimate aim is to construct an HSI sparse representation model that can simultaneously code the spectral–spatial information for the different bands. In Section II-C, the proposed joint spectral–spatial distributed sparse representation model is presented in detail.

C. Joint Spectral–Spatial Distributed Sparse Representation Model

1) Hyperspectral Image Spectral–Spatial Sparse Denoising Model With Noise Adaptivity: Due to the high correlation and the close spectral reflectance values in the neighborhood or highly correlated bands, patches coming from the same spatial location will be regarded as similar structural information. It is assumed that the similar structural information generates similar sparse patterns. Thus, the highly correlated bands can be simultaneously coded by the proposed spectral–spatial sparse representation model, which is introduced in detail as follows.

First, the highly correlated bands are gathered into $C1$ clusters by the correlation-measure-based k-means. As shown in Fig. 1(b), the $C1$ clusters are then reclassified to $C2$ groups based on the spectral position of the different clusters. For one group, the similar patches across the bands are stacked into a vector for sparse representation, as described in Fig. 1(a). A brief description of the patch vectorization is as follows.

1) The 3-D patches are extracted from the groups and arranged into a new 2-D matrix. As shown in Fig. 1(a), each 3-D patch is reshaped into a submatrix of the matrix, with a width a. The neighborhood columns in the matrix have similar sparse pattern.
2) Patches with the size of $n \times T$ from the matrix are stacked into a vector $\boldsymbol{x}_{ij}$ and then jointly coded. Here, T represents that there are T similar spatial vectors $\boldsymbol{x}_{ijb}$. When applying sparse representation to the vector, these patches can keep the similar sparsity by sharing the same atoms. Moreover, owing to the spectral information in the vector, the spectral noise can be simultaneously removed in the processing.

Using the spectral–spatial sparse coding strategy introduced earlier, the objective function for HSIs can be written as

$$\{\widehat{\boldsymbol{X}}, \widehat{\boldsymbol{D}}, \widehat{\alpha}\} = \arg\min_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{\alpha}} \lambda\|\boldsymbol{X} - \boldsymbol{Y}\|_F^2 + \sum_{ij} \mu_{ij}\|\alpha_{ij.}\|_0 + \sum_{ij} \|\boldsymbol{D}\boldsymbol{\alpha}_{ij.} - \boldsymbol{R}_{ij}\boldsymbol{X}\|_2^2. \tag{4}$$

In (4) $\boldsymbol{R}_{ij}.$ is a matrix that extracts the vector $\boldsymbol{x}_{ij}. \in \mathbb{R}^{nT\times 1}$ from the image $\boldsymbol{X}$, and $\boldsymbol{\alpha}_{ij}.$ is the sparse vector according to the vector $\boldsymbol{x}_{ij}.$. In this expression, $\|\boldsymbol{\alpha}_{ij}.\|_0$ is the sparsity prior, and $\sum_{ij}\|\boldsymbol{D}\boldsymbol{\alpha}_{ij.}-\boldsymbol{R}_{ij}.\boldsymbol{X}\|_2^2$ is the global prior [33].

In the sparse model in (4), we allow a representation error $\varepsilon_{ij}^2. = \|\boldsymbol{D}\boldsymbol{\alpha}_{ij}. - \boldsymbol{R}_{ij}.\boldsymbol{X}\|_2^2$ as noise exists in the image. Due to the noise varying from band to band, the error $\varepsilon_{ij}^2. = \|\boldsymbol{D}\boldsymbol{\alpha}_{ij}. - \boldsymbol{R}_{ij}.\boldsymbol{X}\|_2^2$ for $\boldsymbol{x}_{ij}.$ should be below its corresponding threshold $\varepsilon_{ij}. \leq \text{each}(\text{Const}\cdot\sigma_b)$ $(b=1,\ldots,T)$. This means that $\varepsilon_{ijb}(b=1,\ldots,T)$ from $\varepsilon_{ij}.$ should satisfy the corresponding τ_b, which is equal to $\text{Const}\cdot\sigma_b$. Furthermore, the choice of μ_{ij} is handled implicitly by the count of the nonzero entries of $\boldsymbol{\alpha}_{ij}$, and hence, it depends on the representation error of each band. A smaller representation error allows a larger number of nonzero entries and implies a low μ_{ij}. Conversely, it implies a high μ_{ij}. In the preceding way, the different noise intensities in the different bands can be adaptively reduced.

Using the model in (4), the sparsity of the spatial and spectral domains can be simultaneously represented, and a common sparse component across the bands can be effectively coded. However, these highly correlated patches also contain specific components because of the tiny differences in the spectral reflectance. How to sparse code the common information and specific information simultaneously is a critical problem. To solve this issue, a joint spectral–spatial sparse representation and dictionary strategy is proposed. The task can be achieved by alternating between two steps: joint sparse approximation, which captures the common sparse information and the specific sparse information, and joint dictionary learning, which learns a joint spectral–spatial dictionary from the corrupted image itself.

2) Joint Sparse Approximation: Recent studies [40]–[42] have demonstrated that the joint reconstruction of several correlated signals with sparse representation can outperform the results achieved when recovering each signal one by one. For an HSI, these highly correlated bands have a similar structural pattern, and they should be coded jointly to preserve the spatial structure and the spectral continuity. Meanwhile, the differences between bands in a group should be also carefully considered in the sparse representation process. Therefore, in the joint sparse approximation stage, we assume not only that the highly correlated bands share a common sparse component, which is reflected in the interband correlation across bands, but also that each individual patch contains a specific sparse component, which is reflected in the interband correlation within bands. The common sparse component means that the coefficient is the same for all the bands, and the specific sparse component represents the unique structural information of each band.

To simplify the description of the formulation in Sections II-C2 and C3, we use $\boldsymbol{x}$ with sparse coefficient $\boldsymbol{\alpha}$ to replace the extracted vector $\boldsymbol{x}_{ij}.$ with sparse coefficient $\boldsymbol{\alpha}_{ij}.$, $\boldsymbol{x}_b$ with sparse coefficient $\boldsymbol{\alpha}_b$ to replace vector $\boldsymbol{x}_{ijb}$ from the bth band with sparse coefficient $\boldsymbol{\alpha}_{ijb}$, and μ to replace μ_{ij} for vector $\boldsymbol{x}_{ij}.$. The joint sparse coding of one patch in the bth band then has the form

$$\boldsymbol{x}_b = \boldsymbol{D}_b\boldsymbol{\alpha}_b = \boldsymbol{D}_{b,c}\boldsymbol{z}_c + \boldsymbol{D}_{b,b}\boldsymbol{z}_b \tag{5}$$

where $\boldsymbol{\alpha}_b$ $(1\leq b\leq B)$ is the sparse representation coefficient. From the joint sparse idea aforementioned, $\boldsymbol{\alpha}_b$ can be decomposed into two parts: $\boldsymbol{z}_c$ and $\boldsymbol{z}_b$. $\boldsymbol{z}_c$ is the common sparse component with sparsity S_c, which is the same for all the correlated vectors $\boldsymbol{x}_b$. $\boldsymbol{z}_b$ with sparsity S_b is the specific sparse component of sparse representation $\boldsymbol{\alpha}_b$, which captures the intraband correlation to preserve the unique feature for the bth band.

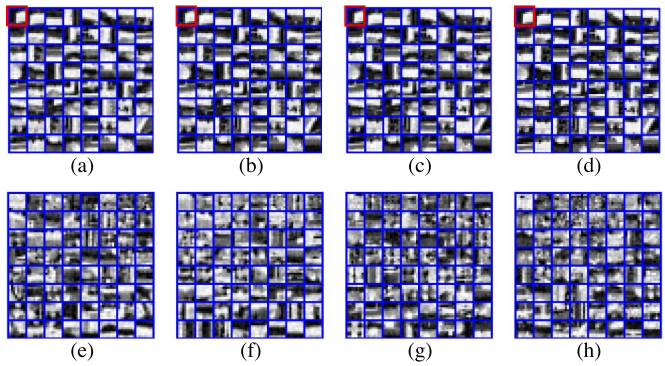


Fig. 2. Dictionary trained in the joint spectral–spatial dictionary learning. (Top row) Common dictionary $\boldsymbol{D_c}$ of the different bands. (Bottom row) Specific dictionary $\boldsymbol{D}_I$ of the different bands.

Supposing the dictionary $\boldsymbol{D}_b$ is fixed, then the sparse representation process in the joint sparse stage involves solving the following optimization problem:

$$\{\widehat{\boldsymbol{z}}_c, \widehat{\boldsymbol{z}}_b\} = \arg\min_{\boldsymbol{z}_c,\boldsymbol{z}_b}\sum_{b=1}^{T}\|(\boldsymbol{D}_{b,c}\boldsymbol{z}_c+\boldsymbol{D}_{b,b}\boldsymbol{z}_b)-\boldsymbol{x}_b\|_2^2 + \mu\left(\sum_{b=1}^{T}\|\boldsymbol{z}_c\|_0+\|\boldsymbol{z}_b\|_0\right). \tag{6}$$

Given the T correlated patches $\boldsymbol{x}=[\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T]^T\in\mathbb{R}^{n\times T}$, then the sparse representation coefficients, which involve a common component and T unique components, can be denoted by $\boldsymbol{z}=[\boldsymbol{z}_c,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_T]^T\in\mathbb{R}^{K\times(T+1)}$. The aforementioned optimization task can be changed to

$$\{\widehat{\boldsymbol{z}}\} = \arg\min_{\boldsymbol{z}}\|\boldsymbol{D}\boldsymbol{z}-\boldsymbol{x}\|_2^2+\mu\|\boldsymbol{z}\|_0 \tag{7}$$

where $\|\boldsymbol{z}\|_0 = S_c+\sum_{t=1}^{T}S_t$, and dictionary $\boldsymbol{D}\in\mathbb{R}^{nT\times K(T+1)}$ can be extended as

$$\boldsymbol{D}=\begin{bmatrix}\boldsymbol{D}_1\\ \vdots\\ \boldsymbol{D_T}\end{bmatrix}=\begin{bmatrix}\boldsymbol{D_{1,c}} & \boldsymbol{D_{1,1}} & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{D_{T,c}} & \boldsymbol{0} & \cdots & \boldsymbol{D_{T,T}}\end{bmatrix}=[\boldsymbol{D_c}\boldsymbol{D_I}] \tag{8}$$

where $\boldsymbol{D_c}=[\boldsymbol{D_{1,c}},\ldots,\boldsymbol{D}_{T,c}]^T$ is the common dictionary for the common sparse component, and $\boldsymbol{D}_{b,\boldsymbol{c}}$ is only a part of the common dictionary for the bth band. All of $\boldsymbol{D}_{b,\boldsymbol{c}}$ $(1\leq b\leq T)$ have almost the same structure. $\boldsymbol{D}_I=\text{diag}(\boldsymbol{D}_{1,1},\ldots,\boldsymbol{D}_{T,T})$ is the set of specific subdictionaries for each band. $\boldsymbol{D}_{b,b}$ contains the particular information for the bth band, and each of $\boldsymbol{D}_{b,b}$ has a different structure. To illustrate the difference between $\boldsymbol{D_c}$ and $\boldsymbol{D}_I$, their structures are shown in Fig. 2, and we give an explanation in the joint dictionary learning part.

Because of the simplicity and efficiency [39] of OMP for the L_0 problem, we adopt the joint OMP algorithm to optimize problem (7). Due to the noise intensity being different in the different bands, the algorithm stops when each band satisfies a

certain criterion. The optimization problem is depicted in the following procedure.

The optimization of joint OMP

Input: an extracted vector $\boldsymbol{x} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]^{\mathrm{T}}$, a fixed dictionary $\boldsymbol{D}$, a constant c, the noise standard deviation $\{\sigma_b\}_{b=1}^{T}$, and the maximum number of iterations *maxIter*.
Output: The sparse representation coefficients $\boldsymbol{z}$, involving a common component $\boldsymbol{z}_c$ and T specific components $\{\boldsymbol{z}_b\}_{b=1}^{T}$.
Initialization: the residual $\boldsymbol{r}_b = \boldsymbol{x}_b (1 \le b \le T)$ and $\boldsymbol{r} = \boldsymbol{x}$, the solution support $S = \text{Support}\{\boldsymbol{z}\} = \emptyset$, the solution $\boldsymbol{z}_c = 0$ and $\boldsymbol{z}_b = 0$.
Procedure: Do

(a) Compute $\boldsymbol{w}_c = \boldsymbol{D}_c^{\mathrm{T}} \boldsymbol{r}$ and $\boldsymbol{w}_b = \boldsymbol{D}_{b,b}^{\mathrm{T}} T \boldsymbol{r}_b (1 \le b \le T)$.
(b) Find a minimizer: if $\max_{k_c} |w_{c,k_c}| \ge \max_{k_b} |w_{b,k_b}|$, $k_c^* = \arg\max_{k_c} |w_{c,k_c}|$ and $S_c = S_c \cup \{k_c^*\}$; otherwise, for any $\max_{k_c} |w_{c,k_c}| < \max_{k_b} |w_{b,k_b}|$, then $\{k_c^*, k_b^*\} = \{\arg\max_k |w_{c,k_c}|, \arg\max_k |w_{c,k_b}|\}$, $S_c = S_c \cup \{k_c^*\}$, and $S_b = S_b \cup \{k_b^*\}$.
(c) Update the provisional solution: compute $\boldsymbol{z} = [\boldsymbol{z}_c, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_T]^{\mathrm{T}}$, the minimizer of $\|\boldsymbol{D}\boldsymbol{z} - \boldsymbol{x}\|_2^2$, subject to $S = S_c \cup \{S_b\}_{b=1}^{T} = \text{Support}\{\boldsymbol{z}\}$.
(d) Update the residual: compute $\boldsymbol{r}_b = \boldsymbol{x}_b - \boldsymbol{D}_{S_c} \boldsymbol{z}_{c,k_c} - \boldsymbol{D}_{S_b} \boldsymbol{z}_{b,k_b}$ then $\boldsymbol{r} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T]^{\mathrm{T}}$.
(e) Stopping rule: each $\|\boldsymbol{r}_b\|_2^2 < (c \cdot \sigma_b)^2$ or achieve the maximum iteration *maxIter*.

End

3) Joint Dictionary Learning: In the proposed method, the dictionary $\boldsymbol{D}$ is composed of the common dictionary $\boldsymbol{D}_{\boldsymbol{c}}$, which includes the similar spatial structure and the continuous spectral information across the bands, and the specific dictionary $\boldsymbol{D}_I$, which includes the unique information of the intraband structure. When fixing the sparse vector $\boldsymbol{\alpha}$, we propose joint spectral–spatial dictionary learning to optimize the atoms from the dictionary. Similar to K-SVD [33], only one column of the dictionary and the nonzero entries in the associated row are updated each time. However, for different components of the sparse coding, the utilized atom is only extracted from the related part of $\boldsymbol{D}$.

For any column k in $\boldsymbol{D}$, we compute its representation error $\boldsymbol{e}^k$ for the different situations, i.e.,

$$\begin{cases} \boldsymbol{e}^k = \boldsymbol{x} - \sum\limits_{m \ne k} \boldsymbol{d}_c^m \alpha(m) - \sum\limits_{m \ne k} \boldsymbol{d}_b^m \alpha(m), & k \in S_c \\ \boldsymbol{e}^k = \left[\underbrace{0, \ldots, 0}_{1,\ldots,\, b-1}, \boldsymbol{x}_b - \sum\limits_{m \ne k} \boldsymbol{d}_{b,c}^m \alpha(m) - \sum\limits_{m \ne k} \boldsymbol{d}_{b,b}^m \alpha(m), \underbrace{0, \ldots, 0}_{b+1,\ldots,T} \right]^T \\ \quad k \in S_b \end{cases} \tag{9}$$

where $\boldsymbol{x}$ is one of the patches using the kth atom, $\alpha(m)$ is the mth entry of the vector $\boldsymbol{\alpha}$, and $\boldsymbol{e}^k$ is the column vector from the error matrix $\boldsymbol{E}^k$. $\boldsymbol{d}_c^m$ is the mth column of dictionary $\boldsymbol{D}_{\boldsymbol{c}}$, and $\boldsymbol{d}_{b,c}^m$ is the mth column of $\boldsymbol{D}_{b,\boldsymbol{c}}$. $\boldsymbol{d}_b^m$ is the mth column of $\boldsymbol{D}_I$, and $\boldsymbol{d}_{b,b}^m$ is the mth column of $\boldsymbol{D}_{b,b}$.

After obtaining the representation error, the resulting problem is given by

$$\{\boldsymbol{d}^k, \alpha_{\cdot,k}\} = \arg\min_{\alpha_{ij\cdot}} \left\| \boldsymbol{E}^k - \boldsymbol{d}^k \boldsymbol{\alpha}_{\cdot,k}^T \right\|_F^2 + \mu_{ij} \|\boldsymbol{\alpha}_{ij\cdot}\|_0 \tag{10}$$

where $\boldsymbol{d}^k$ and $\boldsymbol{\alpha}_{\cdot,k}$ are the updated atom and coefficient row, respectively. The problem can be solved directly via singular value decomposition [43].

In summary, the proposed approach is to solve the dictionary learning problem by alternating between two steps: 1) we first find the common components and the specific components of the sparse coefficients with a fixed dictionary by the use of joint sparse coding, and 2) we then learn the spectral–spatial dictionary by the use of the joint dictionary learning approach. Fig. 2 shows the difference between the common dictionary $\boldsymbol{D}_{\boldsymbol{c}}$ and the specific dictionary $\boldsymbol{D}_I$, as obtained by joint dictionary learning.

In Fig. 2, the top row shows the 64 atoms from the four subdictionaries $\boldsymbol{D}_{b,\boldsymbol{c}}$ from the common dictionary $\boldsymbol{D}_{\boldsymbol{c}}$, and the atoms are arranged by the standard deviation from a large value to a small value. The bottom row shows the specific subdictionaries $\boldsymbol{D}_{b,b}$ corresponding to $\boldsymbol{D}_{b,\boldsymbol{c}}$. From the top row, the atom with the maximum standard deviation in $\boldsymbol{D}_{\boldsymbol{c}}$ is separated into four subatoms, which are shown with red blocks in Fig. 2(a)–(d). The similar pattern in these subatoms indicates that $\boldsymbol{D}_{\boldsymbol{c}}$ can provide the similar structure for the highly correlated bands. On the other hand, Fig. 2(e)–(h) shows that the texture structure in the atoms of the specific subdictionaries is different, which reflects the unique information of the different bands.

After obtaining the sparse coefficients $\boldsymbol{\alpha}_{ij\cdot}$ and the dictionary $\boldsymbol{D}$, returning to (4), we can solve the HSI denoising problem as follows:

$$\widehat{\boldsymbol{X}} = \left(\lambda \boldsymbol{I} + \sum_{ij} R_{ij}^T . \boldsymbol{R}_{ij} \cdot \right)^{-1} \left(\lambda \boldsymbol{Y} + \sum_{ij} \boldsymbol{R}_{ij}^T . \boldsymbol{D} \boldsymbol{\alpha}_{ij\cdot} \right). \tag{11}$$

Finally, the flowchart of the overall denoising process is summarized in Fig. 3

III. Experimental Results and Analysis

The performance of the proposed method was evaluated by applying the different methods to both simulated and real data. Three data sets were employed in the experiments.

1) The first data set was the Washington DC Mall HSI from the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne sensor, which consists of 200 lines and 200 columns, as well as 191 bands. This data set was used in the simulated experiments.
2) The second data set was the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines hyperspectral data set with the size of $145 \times 145 \times 220$, which was used for the real-data experiments. A total of 206 bands was used in the experiments after removing bands 150–163, which are disturbed by the atmosphere and water.

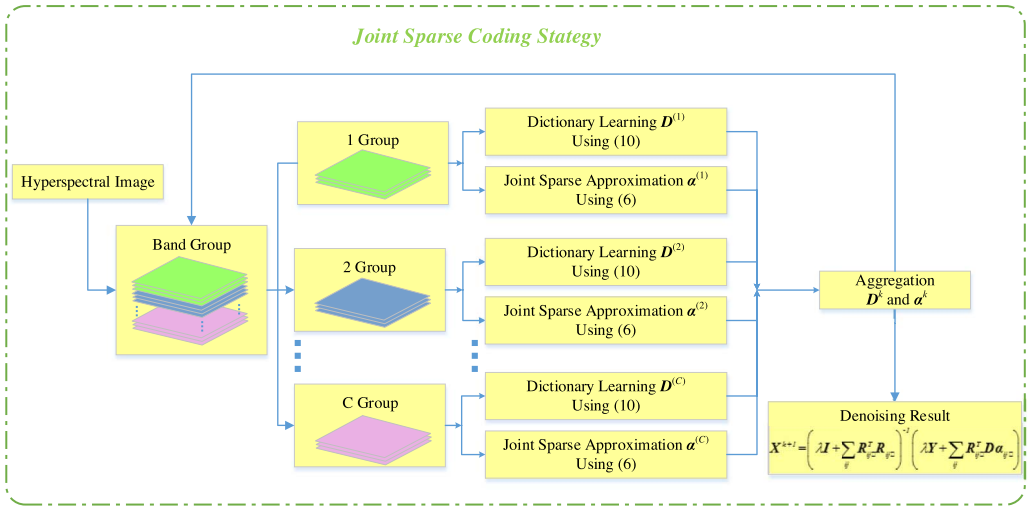


Fig. 3. Flowchart of the proposed joint spectral–spatial denoising method.

TABLE I
QUANTITATIVE EVALUATION OF THE DENOISING RESULTS OF THE SIMULATED EXPERIMENTS

Noise level	Evaluation index	HSSNR	MWF	SSAHTV	K-SVD	SRLR	LRMR	Proposed method (one iteration)	**Proposed method (multiple iterations)**
$\sigma_b = 5$	MPSNR	40.808	37.414	33.514	36.098	43.606	**44.669**	42.052	42.485
	MSSIM	0.986	0.977	0.939	0.963	0.994	**0.994**	0.991	0.992
	MSA(degree)	2.040	2.150	4.162	3.258	1.420	**1.267**	1.679	1.505
$\sigma_b = 15$	MPSNR	32.000	30.617	29.727	29.636	36.400	36.904	33.688	**37.428**
	MSSIM	0.912	0.904	0.844	0.858	0.971	0.974	0.941	**0.977**
	MSA(degree)	5.555	4.329	6.055	5.933	3.023	2.955	3.928	**2.486**
$\sigma_b = 25$	MPSNR	28.116	27.684	25.866	26.977	31.058	34.251	31.411	**34.731**
	MSSIM	0.813	0.826	0.704	0.754	0.893	0.953	0.901	**0.960**
	MSA(degree)	8.577	5.924	7.990	7.534	6.222	3.990	4.746	**3.211**
$\sigma_b = Gau(200,30)$	MPSNR	35.510	30.539	30.019	32.039	35.321	**38.869**	37.488	38.618
	MSSIM	0.943	0.898	0.862	0.894	0.939	**0.981**	0.967	0.980
	MSA(degree)	5.092	5.048	7.095	6.405	5.502	2.731	3.290	**2.569**

3) The third data set was acquired by the airborne Reflective Optics System Imaging Spectrometer (ROSIS) and covers the University of Pavia, Italy. The image scene is of 200 × 200 pixels. After removing 12 water absorption bands, 103 spectral channels remained.

For the simulated experiments, we evaluated the performance with the mean PSNR (MPSNR) index and the mean structural similarity (MSSIM) index, as used in [25], which are denoted as the average of the PSNR index and the SSIM index for all the bands, respectively. To evaluate the spectral fidelity of the results, the average of the spectral angle (MSA) was also introduced in the simulated experiments. Moreover, in the real-data analysis, as there was no reference image available to compute PSNR and SSIM, the classification accuracy improvement after denoising was employed to further assess the effectiveness of the proposed method.

To verify the performance of the proposed denoising method, it was compared with hybrid spatial–spectral noise reduction (HSSNR) [11], MWF [16], spectral–spatial adaptive hyperspectral total variation (SSAHTV) [25], K-SVD [33], sparse representation and low-rank constraint (SRLR) [30], and low-rank matrix recovery (LRMR) [31]. The proposed method was also used to verify the effectiveness of one iteration and multiple iterations. In the simulated process, we simulated the additional noise in the following two cases.

1) For different bands, the noise intensity is equal. For example, $\sigma_b = \mathrm{cons}\tan t$ and is from 5 to 25.
2) For different bands, the noise intensity is different.

The noise variance σ_b^2 was added along the spectral axis and was varied like a Gaussian curve centered at the middle band $(B/2)$ [44] as

$$\sigma_b^2 = \beta^2 \frac{\exp\left\{-(b-B/2)^2/2\eta^2\right\}}{\sum_{b=1}^{B}\exp\left\{-(b-B/2)^2/2\eta^2\right\}} \tag{12}$$

where the power of the noise is controlled by β, and η behaves like the standard deviation for the Gaussian curve. In the simulated experiments, $\beta = 200$, $\eta = 30$, and the noise was defined as $\sigma_b = \mathrm{Gau}(\beta, \eta)$.

A. Simulated-Data Experiments

To allow a comprehensive comparison, quantitative evaluation, visual comparison, and the spectral difference between the original and denoised data were used to analyze the results of the different methods. The contrasting results of the two

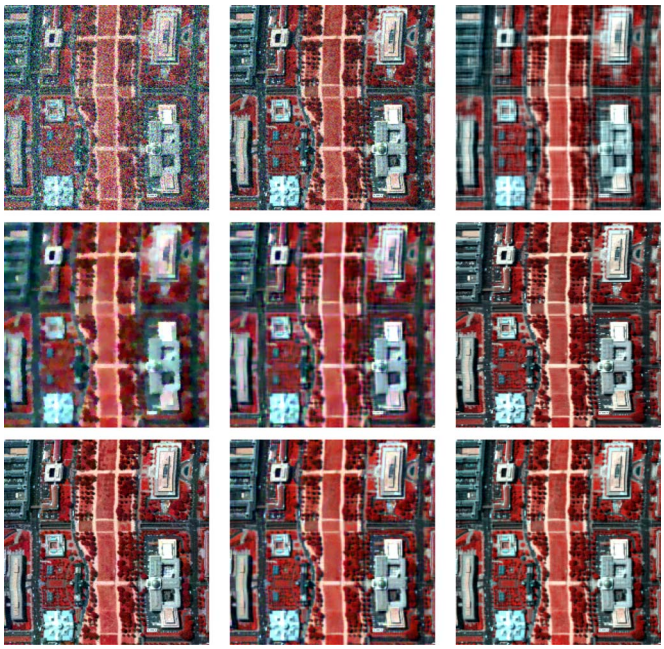


Fig. 4. Results of the Washington DC Mall image with $\sigma_b = 25$ in Case 1. (Top row) Noisy band (17, 27, 57), HSSNR, and MWF. (Middle row) SSAHTV, K-SVD, and SRLR. (Bottom row) LRMR, K-SVD, the proposed method with one iteration, and the proposed method with multiple iterations.

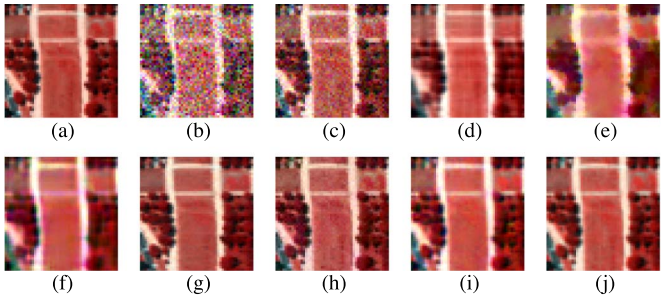


Fig. 5. $\times 4$ magnified results of the Washington DC Mall image with $\sigma_b = 25$ in Case 1. (a) Clean image (17, 27, 57). (b) Noisy image (17, 27, 57). (c) HSSNR. (d) MWF. (e) SSAHTV. (f) K-SVD. (g) SRLR. (h) LRMR. (i) Proposed method with one iteration. (j) Proposed method with multiple iterations.

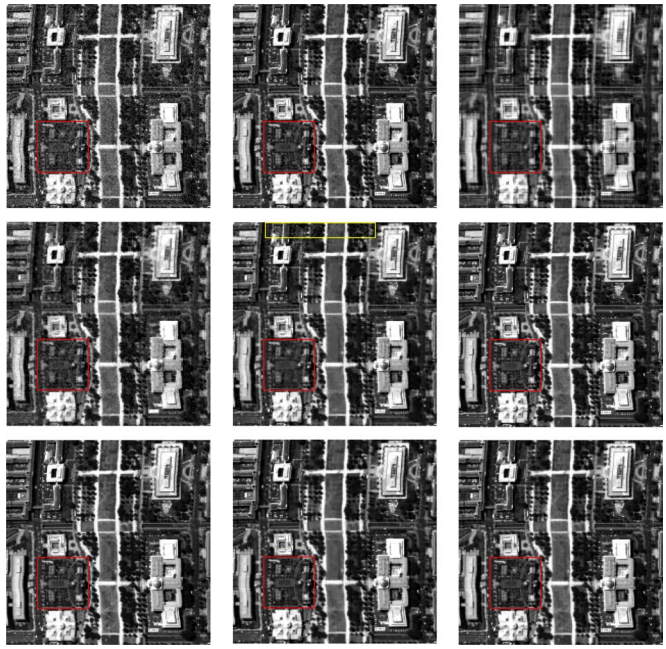


Fig. 6. Results of the Washington DC Mall image in Case 2. (Top row) Noisy band 35 with $\sigma_{35} = 8.3498$, HSSNR, and MWF. (Middle row) SSAHTV, K-SVD, and SRLR. (Bottom row) LRMR, K-SVD, the proposed method with one iteration, and the proposed method with multiple iterations.

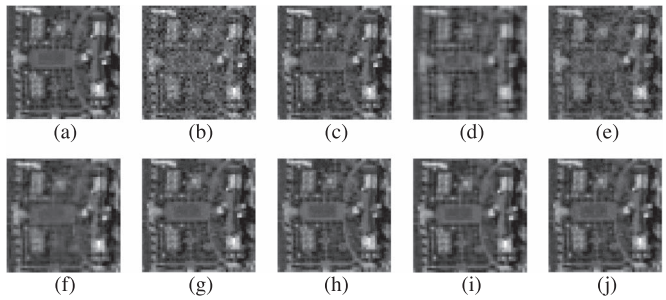


Fig. 7. $\times 4$ magnified results of the various approaches in the red block in Fig. 6. (a) Clean image. (b) Noisy image 35. (c) HSSNR. (d) MWF. (e) SSAHTV. (f) K-SVD. (g) SRLR. (h) LRMR. (i) Proposed method with one iteration. (j) Proposed method with multiple iterations.

cases with various noise levels are given in Table I. To give detailed contrasting results, $\sigma_b = 25$ and $\sigma_b = \mathrm{Gau}(200, 30)$ are selected to show the visual effect. Because of the large number of bands in the HSI, only a few bands are presented to give the visual results in each case. Fig. 4 shows the denoising results of the different methods in simulated Case 1 with the pseudocolor view of bands 17, 27, and 57 (see enlarged detail in Fig. 5), whereas Figs. 6–10 show the denoising results of the different methods in simulated Case 2. Figs. 7, 9, and 10 are enlarged details from Figs. 6 and 8. Each patch size from the noisy image in the proposed method was 8×8 with $n = 64$, and the bandwidth T was 8. For the one-iteration denoising, the parameters in the joint OMP were chosen manually as $\mathrm{Const} = 0.9$ and $\tau_b = \mathrm{Const} \cdot \sigma_b = 0.9 \cdot \sigma_b$. For the multiple-iteration denoising, $\mathrm{Const} = 0.5$, and $\tau_b = \mathrm{Const} \cdot \sigma_b = 0.5 \cdot \sigma_b$. The choice of parameter λ was dependent on the noise level [33], and we set it to $30/\sigma_b$.

Compared with the other methods in Table I, the proposed method achieves the highest MPSNR and MSSIM values and the lowest MSA values in most cases and shows a better visual quality in Figs. 4–10. Although the HSSNR method produces a lower spectral angle, as shown in Table I, which reflects that it suppresses the spectral noise better than the MWF, SSAHTV, and K-SVD methods, Figs. 4–10 show that it cannot effectively remove the noise in the spatial dimension. MWF applies the tensor decomposition technique to successfully remove the spectral noise, but it fails to preserve the structural information in Figs. 4–10. In the results of MWF, the vertical and horizontal directions import obvious fake stripes in the texture, and the edges are blurred. The fake stripes are more obvious when the standard variance of the noise increases. With the high noise intensity condition in Figs. 4 and 8, the SSAHTV method generates some fake artifacts and causes a disturbing sawtooth pattern in the edges.

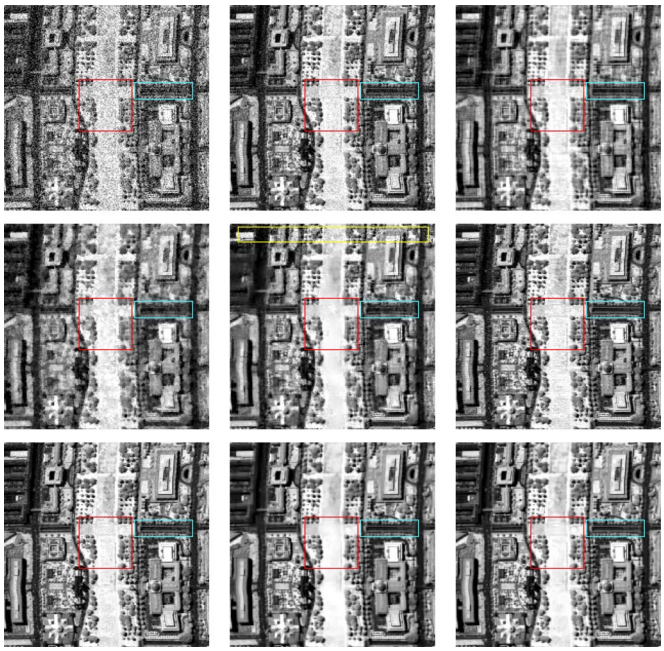


Fig. 8. Results of the Washington DC Mall image in Case 2. (Top row) Noisy band 97 with $\sigma_{97} = 23.0659$, HSSNR, and MWF. (Middle row) SSAHTV, K-SVD, and SRLR. (Bottom row) LRMR, K-SVD, the proposed method with one iteration, and the proposed method with multiple iterations.

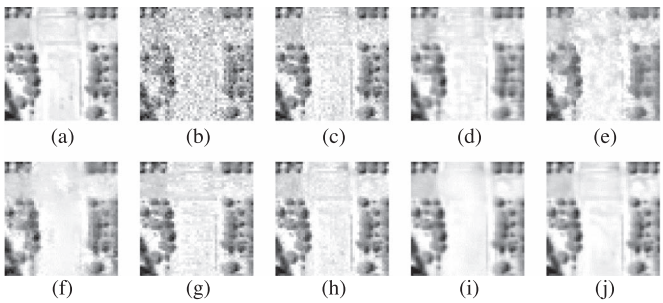


Fig. 9. $\times 4$ magnified results of the various approaches in the red block in Fig. 8. (a) Clean image. (b) Noisy image 97. (c) HSSNR. (d) MWF. (e) SSAHTV. (f) K-SVD. (g) SRLR. (h) LRMR. (i) Proposed method with one iteration. (j) Proposed method with multiple iterations.

The proposed method shows an obvious superiority when the noise intensity increases, which is shown in Table I and the visual results. From the magnified results, it can be seen that the proposed algorithm not only suppresses noise but also preserves more image details. Even when the noise intensity increases, the tiny objectives on the ground are still preserved well in the proposed method. In contrast, the K-SVD method removes some details such as small features, and its results are blurred in Figs. 8–10. The yellow block of the K-SVD results in Figs. 6 and 8 also shows obvious noise residuals in Case 2. By exploring the spatial–spectral information, SRLR and LRMR also provide better denoising results. However, there are still some noise residuals in the magnified regions, particularly in the high noise intensity condition.

The spectral reflectance is significant for HSI applications, and spectral distortion should be avoided in the denoising process. To verify the effectiveness of denoising in the spectral

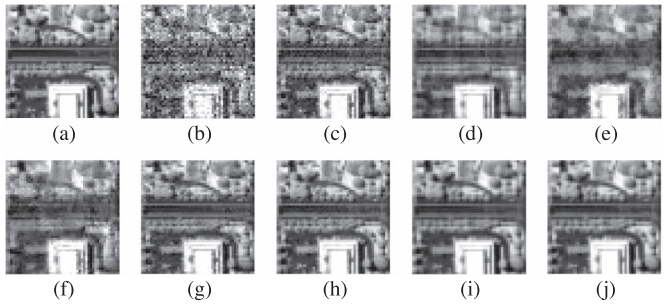


Fig. 10. $\times 4$ magnified results of the various approaches in the cyan block in Fig. 8. (a) Clean image. (b) Noisy image 97. (c) HSSNR. (d) MWF. (e) SSAHTV. (f) K-SVD. (g) SRLR. (h) LRMR. (i) Proposed method with one iteration. (j) Proposed method with multiple iterations.

dimension, Fig. 11 shows the spectra before and after denoising. The horizontal axis of the figures represents the value of all the pixels in each band, and the vertical axis shows the band number. Each row represents the spectral reflectance of an object. Each column shows the spatial information of a band. It is shown in Fig. 11 that the proposed method suppresses the spectral dimension noise better than HSSNR, SSAHTV, K-SVD, SRLR, and LRMR. SRLR produces a serious boundary effect. In the red region, the LRMR results retain more residual spectral noise than the results of the proposed method with multiple iterations. In the MWF denoising results, the noise along the spectral signatures is also effectively removed. However, referring to the original spectral signatures in Fig. 11(a), the proposed method obtains more abundant detailed spatial information than MWF as the spatial feature of the proposed method along each column is closer to the clean image.

To show the changes in the spectral reflectance after denoising, the spectral distortion of the different denoising methods on the Washington DC Mall data is also given in Fig. 12. The differences in the spectral signatures of pixel (111, 122) from the grass class, the pixel (180, 52) from the road class, and the pixel (67, 97) from the roof class are presented in Fig. 12(a)–(c), respectively. In Fig. 12, the vertical axis of the figures represents the digital number values, and the horizontal axis shows the spectral band number. Here, it can be observed that the curve of the proposed method is smoother than for the other methods, and the spectral residual is closer to zero, indicating that the proposed method is more effective at preserving the useful spectral information of the original HSI.

B. Real-Data Experiments

Two real-data experiments were also undertaken to further verify the proposed method in a real case. To adaptively suppress the noise for the different bands in the real-data experiments, the thresholding for the residual error τ_b was set in the different bands according to the noise level σ_b. Therefore, in a real-data case, the standard variance of the noise before the denoising processing needs to be estimated; thus, a patch-based noise level estimation algorithm [45] was used to estimate the noise standard variance. For a real-data case, there is no reference image to compute MPSNR, MSSIM, and MSA.

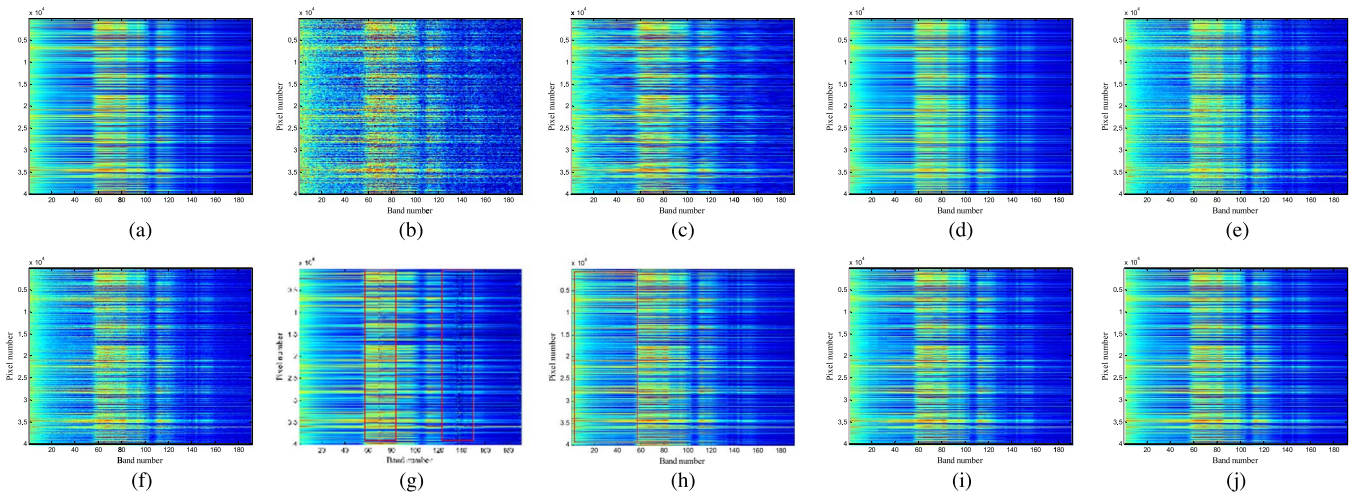


Fig. 11. Denoising results of the spectral dimension with $\sigma_b = 25$ in Case 1. (a) Clean image. (b) Noisy image. (c) HSSNR. (d) MWF. (Bottom row) (e) SSAHTV. (f) K-SVD. (g) SRLR. (h) LRMR. (i) Proposed method with one iteration. (j) Proposed method with multiple iterations.

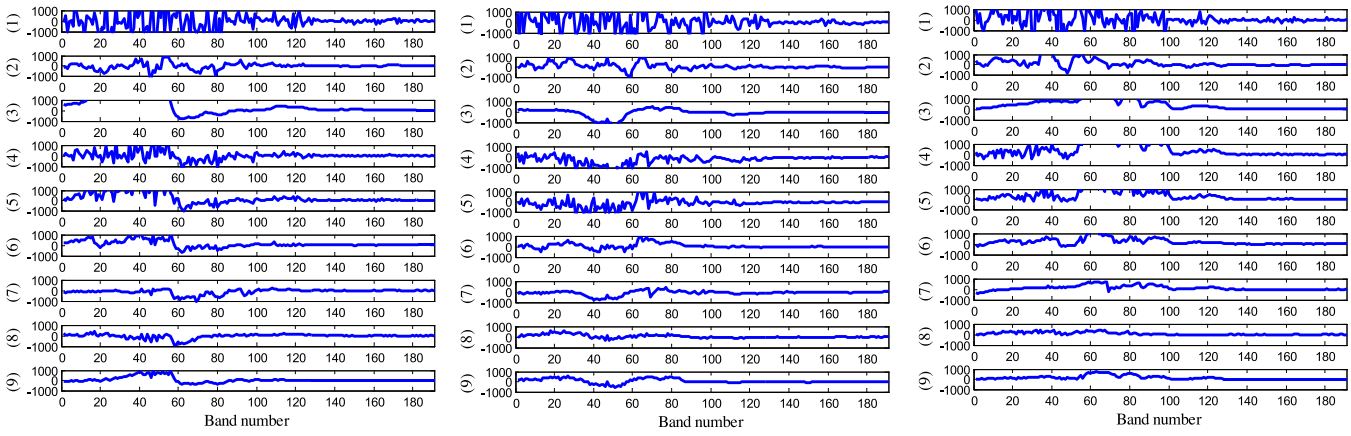


Fig. 12. Difference between the noise-free spectrum and the restoration results of (a) pixel (111, 122), which belongs to the grass class, (b) pixel (180, 52), which belongs to the road class, and (c) pixel (67, 97), which belongs to the roof class. Curves (1)–(9) in the different classes denote the results of the noisy image, HSSNR, MWF, SSAHTV, K-SVD, the proposed method with one iteration, SRLR, LRMR, and the proposed method with multiple iterations, respectively.

Fig. 13. AVIRIS Indian Pines data set used in real-data experiment 1. (a) False-color hyperspectral 3-D cube. (b) Reference map containing 16 mutually exclusive land cover classes.

Hence, the classification accuracy improvement after denoising was used to evaluate the performance. The reference maps of the two real data sets are shown in Figs. 13(b) and 18(b). Support vector machine (SVM) [25], [46] was utilized as the classifier. The overall accuracy (OA) and the kappa coefficient were used as evaluation indices.

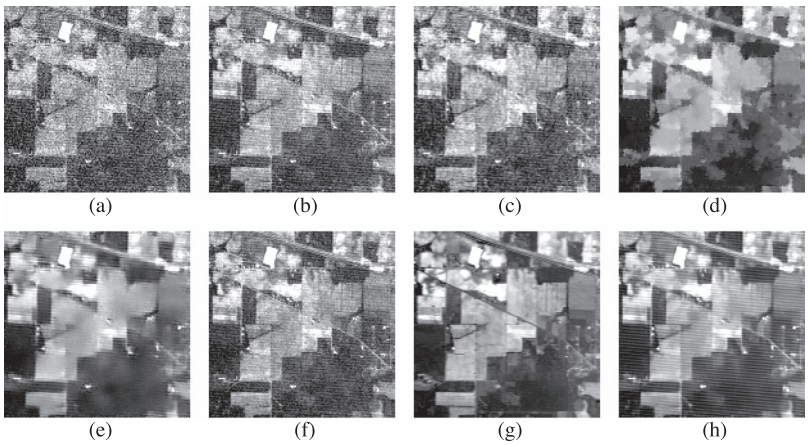


Fig. 14. Results of the Indian Pines image. (a) Noisy band 2. (b) HSSNR. (c) MWF. (d) SSAHTV. (e) K-SVD. (f) SRLR. (g) LRMR. (h) Proposed method.

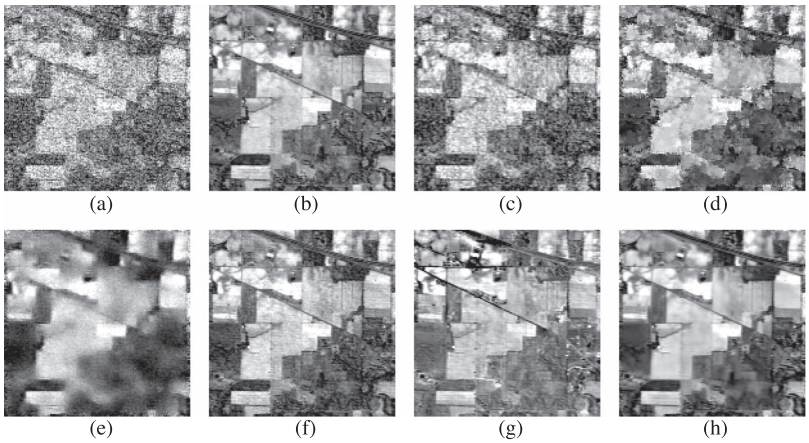


Fig. 15. Results of the Indian Pines image. (a) Noisy band 103. (b) HSSNR. (c) MWF. (d) SSAHTV. (e) K-SVD. (f) SRLR. (g) LRMR. (h) Proposed method.

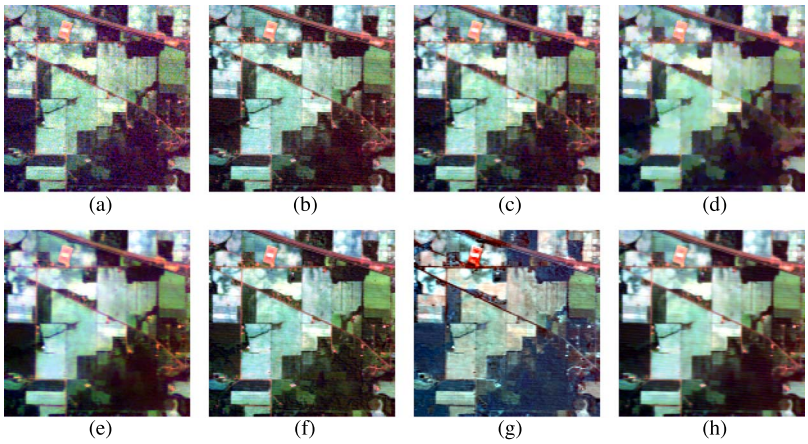


Fig. 16. Results of the Indian Pines image. (a) Noisy band (3, 110, 204). (b) HSSNR. (c) MWF. (d) SSAHTV. (e) K-SVD. (f) SRLR. (g) LRMR. (h) Proposed method.

In the first real-data experiment with the Indian Pines data, the results of the one iteration mode with a patch size of 8×8 and a bandwidth of 8 were given. The thresholds were chosen manually as $\text{Const} = 1.15$ and $\tau_b = \text{Const} \cdot \sigma_b = 1.15 \cdot \sigma_b$. Parameter λ was set to $30/\sigma_b$. The visual results of bands 2 and 103 with a high noise intensity are presented in Figs. 14 and 15, respectively, and the pseudocolor views of bands 3, 110, and 204 are shown in Fig. 16. It can be clearly observed that HSSNR and MWF produce heavy residual noise in the results, such as the red particles in the smooth region in Fig. 16. This is because the hard thresholding method cannot reduce the noise completely in HSSNR, whereas MWF cannot distinguish the signal and noise subspaces well. Although SSAHTV can get rid of the noise and sharpen the edges, the results show a disturbing sawtooth pattern and lose some textures. The K-SVD method produces some artifacts in Fig. 14, and the results appear oversmoothed in Fig. 15. SRLR can preserve more textures but leave the obvious noise. Although LRMR can remove stripe noise well in Fig. 14, it has a poor performance for the Poisson–Gaussian mixed noise in Fig. 15 and causes the

TABLE II
CLASSIFICATION ACCURACY RESULTS OF THE INDIAN PINES DATA SET

	The original	HSSNR	MWF	SSAHTV	K-SVD	SRLR	LRMR	The proposed method
OA	71.73%	72.22%	74.54%	83.48%	77.93%	74.55%	73.32%	83.97%
Kappa	0.6738	0.6798	0.7069	0.8107	0.7468	0.7074	0.6924	0.8171

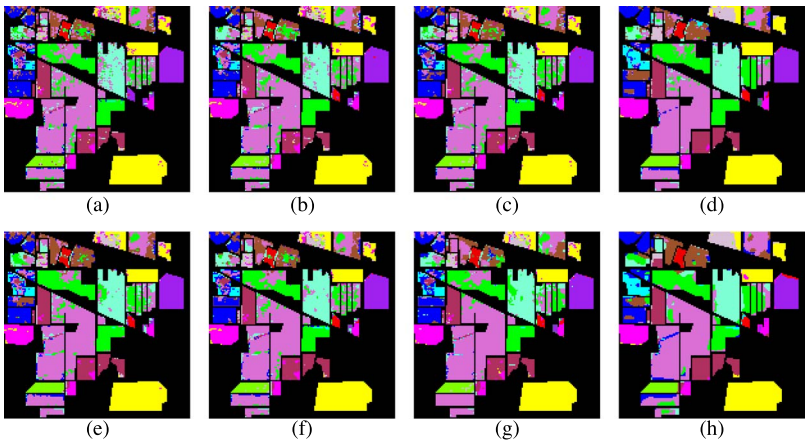


Fig. 17. Classification results of the Indian Pines image using SVM before and after denoising. (a) Noisy band. (b) HSSNR. (c) MWF. (d) SSAHTV. (e) K-SVD. (f) SRLR. (g) LRMR. (h) Proposed method.

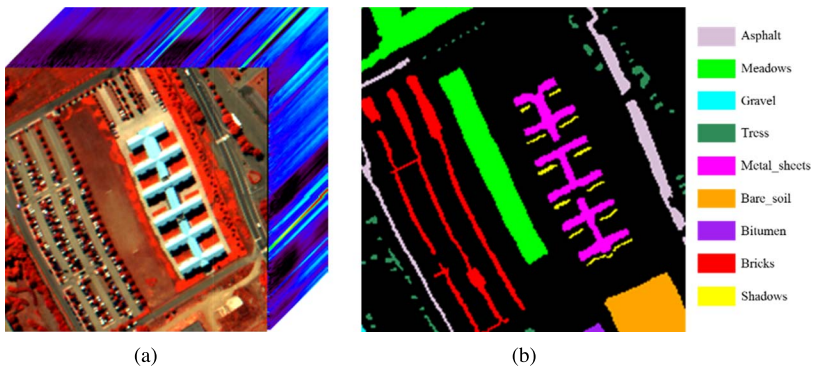


Fig. 18. ROSIS Pavia University data set used in real-data experiment 2. (a) False-color hyperspectral 3-D cube. (b) Reference map containing nine mutually exclusive land cover classes.

heavy spectral distortion, which comes with low classification accuracy in Table II. The visual results show that the proposed method can suppress the noise well and effectively preserve the edge and detail information.

In the classification experiment, 16 ground-truth classes, with a total of 10 366 samples, were considered in the experiment. The training sets consisted of 10% of the number of test samples randomly drawn from each class. The classification results of the Indian Pines image before and after denoising are given in Fig. 17. In the classification result before denoising, the visual results appear fragmentary, and the OA and the kappa coefficient are only 71.73% and 0.6738. After denoising, the OA and the kappa coefficient show different levels of improvement. However, the classification results from HSSNR, MWF, K-SVD, SRLR, and LRMR still show a serious fragmentary effect. SSAHTV and the proposed method reduce the fragmentary effect in most regions of the image, but the proposed method provides a better classification result in the upper left corner of the image, including the Oats, Corn, and Soybeans-no till classes. From Table II, the proposed method obtains the highest OA and kappa coefficient values of 83.97% and 0.8171.

The second real HSI data set was of an urban area of Pavia, Italy, which is centered at the University of Pavia. The results with three iterations were selected to show the superiority of the proposed method. The patch size was 8×8, and the bandwidth was 8 in this experiment. The thresholds were chosen manually as Const $= 0.5$ and $\tau_b = 0.5 \cdot \sigma_b$. Parameter λ was set to $30/\sigma_b$. To assess the results, nine ground-truth classes, with a total of 9310 samples, were used to compute the OA and the kappa coefficient. The training sets consisted of 10% of the number of each class in the ground truth randomly drawn from the test sets. The false-color composite and the ground-truth map are shown in Fig. 18(b).

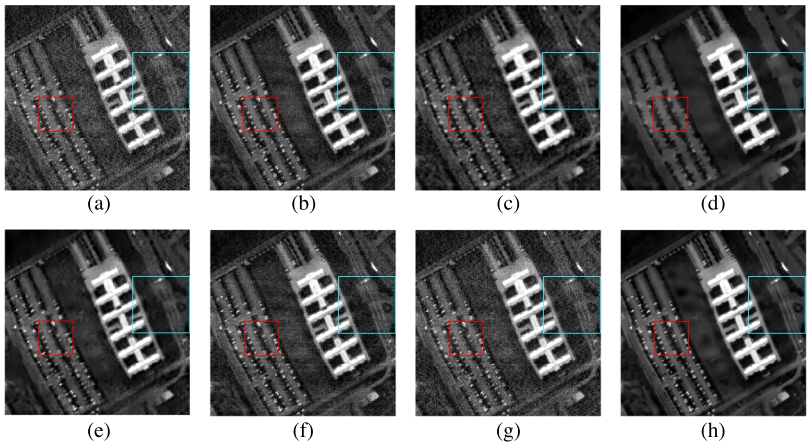


Fig. 19. Results of the Pavia University image. (a) Noisy band 1. (b) HSSNR. (c) MWF. (d) SSAHTV. (e) K-SVD. (f) SRLR. (g) LRMR. (h) Proposed method.

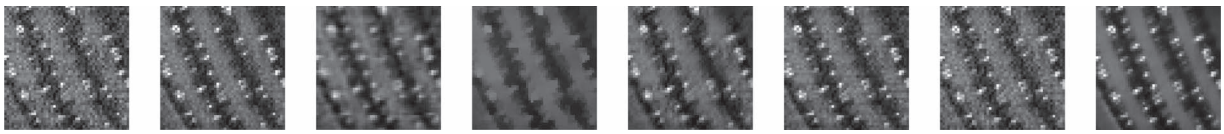


Fig. 20. ×4 magnified results of the various approaches in the red block of the Pavia University image. (Left to right) Noisy band 1, HSSNR, MWF, SSAHTV, K-SVD, SRLR, LRMR, and the proposed method.

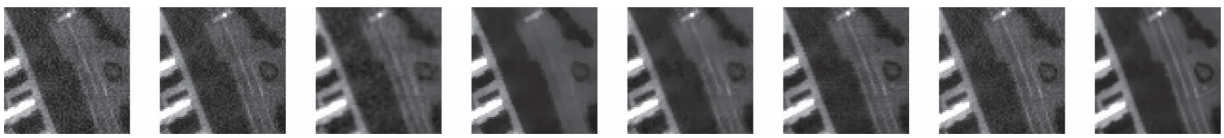


Fig. 21. ×3 magnified results of the various approaches in the cyan block of the Pavia University image. (Left to right) Noisy band 3, HSSNR, MWF, SSAHTV, K-SVD, SRLR, LRMR, and the proposed method.

TABLE III
CLASSIFICATION ACCURACY RESULTS OF THE PAVIA UNIVERSITY IMAGE

	The Original	HSSNR	MWF	SSAHTV	K-SVD	SRLR	LRMR	The proposed method
OA	75.40%	75.44%	76.92%	83.90%	77.11%	89.98%	90.44%	90.57%
Kappa	0.6819	0.6825	0.7016	0.7927	0.7044	0.8727	0.8785	0.8802

In the Pavia University image, as the noise is mainly concentrated in the front bands of the image, the first 30 bands with a high noise intensity are selected to show the classification improvement. Fig. 19 shows the results of denoising using the different methods in band 1. In order to show the visual superiority of the proposed method, we also present some detailed regions cropped from the denoised images. The red block and the cyan block in Fig. 19 label the position of the enlarged regions. Figs. 20 and 21 show the enlarged details for bands 1 and 3, respectively. From these enlarged regions, it can be observed that the MWF and K-SVD methods smooth the detailed information. Some noise is retained, and fake artifacts are generated in the results of the methods HSSNR, MWF, SSAHTV, K-SVD, SRLR, and LRMR. In the result of the SSAHTV method, although the main edges are preserved, the detail and the texture information are oversmoothed. As shown in Figs. 20 and 21, tiny points and thin edges are removed in the SSAHTV result. However, the proposed method not only suppresses the noise but also preserves the image details and texture information.

The classification results in Table III also reflect the effectiveness of the proposed approach, which produces the highest OA and kappa coefficient values of 90.57% and 0.8802. In Fig. 22, it is shown that SSAHTV, SRLR, LRMR, and the proposed method can both suppress the fragmentary effect better than the other methods. However, the proposed method not only obtains the most reasonable classification result but also gives a better visual performance for edges and textures in Figs. 19–21.

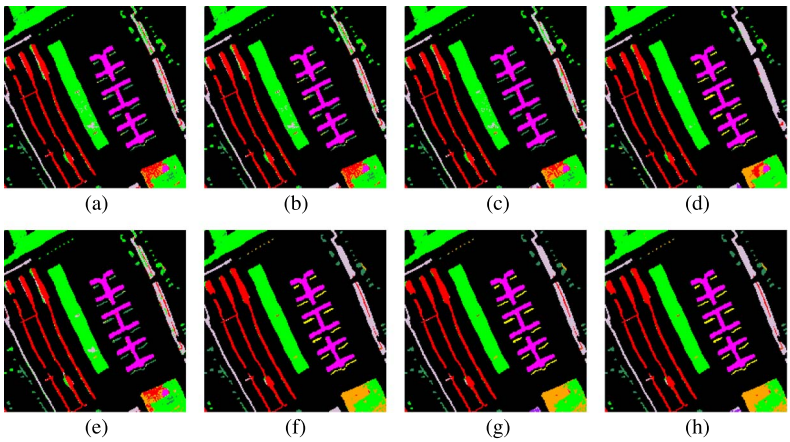


Fig. 22. Classification results of the Pavia University image using SVM before and after denoising. (a) Noisy band. (b) HSSNR. (c) MWF. (d) SSAHTV. (e) K-SVD. (f) SRLR. (g) LRMR. (h) Proposed method.

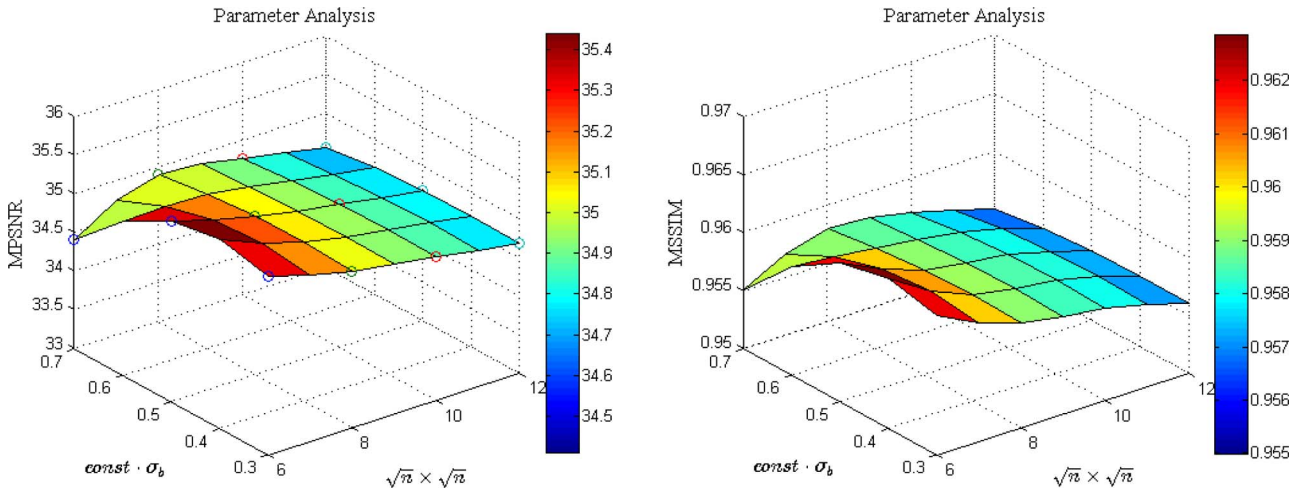


Fig. 23. Restoration results under different values of the threshold Const · σ_b and the patch size $\sqrt{n} \times \sqrt{n}$.

C. Discussion

In all the simulated and real-data experiments, the number of atoms in the dictionary K was manually set to $2nT$, where n is the size of one patch, and T is the number of correlated patches, such as $\boldsymbol{x} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]^T \in \mathbb{R}^{n \times T}$. The parameters λ and μ were adaptively obtained based on the noise intensity: λ was set to $30/\sigma_b$, and μ was dependent on the threshold of the representation error $\varepsilon \leq \text{each}(\text{Const} \cdot \sigma_b)$ $(b = 1, \ldots, T)$, which was explained in Section II-C. The restoration results are mainly influenced by the threshold Const · σ_b and the size of patch $\sqrt{n} \times \sqrt{n}$. In the simulated experiments, Fig. 23 presents the quantitative evaluation results of different thresholds and patch sizes under the multiple-iteration mode. It can be clearly seen that the results of the proposed method are quite robust with regard to the values of the threshold when the patch size has been already determined. When the size of patch is between 6 and 8 and the value of the threshold is between 0.4 and 0.5, the MPSNR and MSSIM values of the results produced by the proposed method show much improvement after multiple iterations.

Finally, the computational complexity of the proposed method is discussed. From the optimization algorithm of joint OMP, it can be observed that the computational complexity of each iteration per pixel in a vector $\boldsymbol{x} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]^T$ is $O(nTK)$, where nT is the dimension of vector $\boldsymbol{x}$, and K is the number of atoms in the dictionary. As shown in [33], with $S = S_c + \sum_{t=1}^{T} S_t$ number of nonzero elements in each coefficient vector, the computational complexity of OMP for a vector is $O(nTKS)$ when these stages are iterated S times. However, it should be noted that joint OMP needs only S_c iterations to find the sparse approximations for T signals. Hence, the computation of joint OMP with $O(nTKS_c)$ is less than the computation of OMP, and the aforementioned experiments also confirm that joint OMP can obtain better results than OMP. Finally, the dictionary update with J iterations then requires $O(nTKS_cJ)$ operations per pixel from image $\boldsymbol{X}$.

IV. Conclusion

In this paper, we have proposed an HSI denoising method by jointly utilizing spatial and spectral redundancy and correlation. A joint sparse representation framework is established to effectively use the spatial structure similarity and the spectral correlation. The sparse coding consists of a common component reflecting the interband correlation and a specific component depicting the intraband structure. The sparse coefficients of the common component are obtained from the common dictionary and represent the similar information of all the bands and the continuity of the spectra. The specific components are coded from the specific atoms of the dictionary and record the unique structural features for each band. By setting different threshold values for the representation errors of different noise intensity bands, the proposed method can effectively achieve spectrally adaptive image denoising.

The experimental results demonstrate that the proposed denoising method can achieve a better performance than the traditional HSI denoising methods. The joint spectral–spatial HSI denoising method with distributed sparse representation can recover a clean structure and preserve the spectral consistency in the original image. In the future, we will concentrate on removing noise for HSIs along the spectral dimension. In addition, accelerating the speed of the algorithm will be another interesting research direction.

Acknowledgment

The authors would like to thank Prof. Y.-Q. Zhao of the School of Automation, Northwestern Polytechnical University, Xi'an, China, for sharing the SRLR code. They would also like to thank the handling editors and anonymous reviewers for their helpful comments.

References

[1] J. M. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jul. 2013.

[2] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, Jan. 2014.

[3] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[4] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[5] D. Tao, X. Lin, L. Jin, and X. Li, "Principal component 2-dimensional long short-term memory for font recognition on single Chinese characters," *IEEE Trans. Cybernetics*, vol. 46, no. 3 pp. 756–765, Mar. 2016.

[6] A. Zare, J. Bolton, P. Gader, and M. Schatten, "Vegetation mapping for landmine detection using long wave hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 172–178, Jan. 2008.

[7] J. P. Kerekes and J. E. Baum, "Hyperspectral imaging system modeling," *MIT Linc. Lab. J.*, vol. 14, no. 1, pp. 117–130, 2003.

[8] I. Atkinson, F. Kamalabadi, and D. Jones, "Wavelet-based hyperspectral image estimation," in *Proc. IEEE IGARSS*, 2003, vol. 2, pp. 743–745.

[9] G. Chen and S. 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, Oct. 2008.

[10] G. Chen, S. Qian, and W. Xie, "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.

[11] H. Othman and S. 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.

[12] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[13] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[14] P. Scheunders and S. D. Backer, "Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1865–1872, Jul. 2007.

[15] D. Muti and S. Bourennane, "Multidimensional filtering based on a tensor approach," *Signal Process.*, vol. 85, no. 12, pp. 2338–2353, Dec. 2005.

[16] 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.

[17] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[18] A. Karami, M. Yazdi, and A. Zolghadre 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.

[19] X. Guo, X. Huang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm.*, vol. 83, pp. 50–63, Sep. 2013.

[20] Q. Wang, L. Zhang, Q. Tong, and F. Zhang, "Hyperspectral imagery denoising based on oblique subspace projection," *J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2468–2480, Jun. 2014.

[21] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[22] D. Tao, J. Cheng, M. Song, and X. Lin, "Manifold ranking-based matrix factorization for saliency detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1122–1134, Jun. 2016.

[23] Y. Wang, R. Niu, and X. Yu, "Anisotropic diffusion for hyperspectral imagery enhancement," *IEEE Sensors J.*, vol. 10, no. 3, pp. 469–477, Mar. 2010.

[24] S. Chen, X. Hu, and S. Peng, "Hyperspectral imagery denoising using a spatial–spectral domain mixing prior," *J. Comput. Sci. Technol.*, vol. 27, no. 4, pp. 851–861, Jul. 2012.

[25] 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.

[26] P. Liu, F. Huang, G. Li, and Z. Liu, "Remote-sensing image denoising using partial differential equations and auxiliary images as priors," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 358–362, May 2012.

[27] 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.

[28] M. Ye, Y. Qian, and J. Zhou, "Multitask sparse nonnegative matrix factorization for joint spectral–spatial hyperspectral imagery denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2621–2639, May 2015.

[29] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral–spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan. 2016.

[30] Y. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[31] H. Zhang, W. He, L. Zhang, and H. Shen, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[32] Y. Yuan, X. Zheng, and X. Lu, "Spectral–spatial kernel regularized for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3815–3832, Jul. 2015.

[33] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[34] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

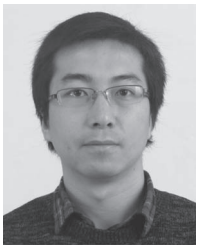
[35] X. Zeng, W. Bian, W. Liu, J. Shen, and D. Tao, "Dictionary pair learning on Grassmann manifolds for image denoising," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4556–4569, Nov. 2015.

[36] G. Martin, J. M. Bioucas-Dias, and A. Plaza, "HYCA: A new technique for hyperspectral compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2819–2831, May 2015.

[37] S. Mallat and Z. Zhang, "Matching pursuit in a time–frequency dictionary," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
[38] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–59, 2001.
[39] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27th Asilomar Conf. Signals, Syst., Comput.*, 1993, pp. 40–44.
[40] L. Zelnik-Manor, K. Rosenblum, and Y. Eldar, "Dictionary optimization for block-sparse representations," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2386–2395, May 2012.
[41] W. Chen, I. J. Wassell, and M. R. D. Rodrigues, "Dictionary Design for Distributed Compressive Sensing," *IEEE Signal Process. Lett.*, vol. 22, no. 1, pp. 95–99, Jan. 2015.
[42] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2015.
[43] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[44] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[45] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Signal Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.
[46] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.

Jie Li (S'13) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in June 2011. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University.

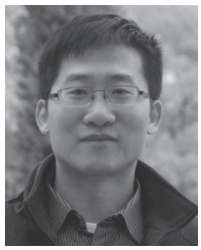
His research interests include image quality improvement, image superresolution reconstruction, data fusion, remote sensing image processing, and sparse representation.



Qiangqiang Yuan (M'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is currently an Associate Professor. He has authored or coauthored over 30 research papers, including over 20 peer-reviewed articles in international journals such as the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His current research interests include image reconstruction, remote sensing image processing and application, and data fusion.

Dr. Yuan has frequently served as a referee for more than ten international journals for remote sensing and image processing. He was a recipient of the Top Ten Academic Star of Wuhan University in 2011 and the Hong Kong Scholar Award from The Society of Hong Kong Scholars and the China National Postdoctoral Council in 2014.



Huanfeng Shen (M'11–SM'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In July 2007, he joined the School of Resource and Environmental Science, Wuhan University, where he is currently a Luojia Distinguished Professor. He has been supported by several talent programs, such as the New Century Excellent Talents by the Ministry of Education of China (2011), the China National Science Fund for Excellent Young Scholars (2014), and the Youth Talent Support Program of China (2015). He has authored or coauthored over 100 research papers. His research interests include image quality improvement, remote sensing mapping and application, data fusion and assimilation, and regional and global environmental change.

Prof. Shen is currently a member of the editorial board of the *Journal of Applied Remote Sensing*.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has authored or coauthored over 450 research papers and five books. He is the holder of 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He was the General Chair for the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). He is a Fellow of The Institution of Engineering and Technology, an Executive Member (board of governor) of the China National Committee for the International Geosphere-Biosphere Programme, and an Executive Member of the China Society of Image and Graphics. He regularly serves as a Cochair of the series SPIE conferences on multispectral image processing and pattern recognition, conference on Asia remote sensing, and many other conferences. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geospatial Information Science*, and the *Journal of Remote Sensing* and a Guest Editor of the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He edits several conference proceedings, issues, and geoinformatics symposiums. He was a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE Sensing (JSTARS). His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or the finalists of the IEEE International Geoscience and Remote Sensing Symposium student paper contest in recent years. He was a recipient of the best reviewer awards from IEEE GRSS for his service to IEEE JSTARS in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2014. He was also a recipient of the 2010 best paper Boeing award and the 2013 best paper ERDAS award from the American Society of Photogrammetry and Remote Sensing.