

Intracluster Structured Low-Rank Matrix Analysis Method for Hyperspectral Denoising

Wei Wei^{ID}, Senior Member, IEEE, Lei Zhang^{ID}, Member, IEEE, Yining Jiao, Chunna Tian, Member, IEEE, Cong Wang, and Yanning Zhang, Senior Member, IEEE

Abstract—Hyperspectral images (HSIs) denoising aims at eliminating the noise generated during the acquisition and transmission of HSIs. Since denoising is an ill-posed problem, utilizing proper knowledge of HSIs as regularization is essential for a good denoiser. Many HSI denoising methods have been proposed to leverage various prior knowledge, e.g., total variation, sparsity, and so on. Among those knowledge, a low-rank property has been shown to be effective for HSI denoising since it has the ability to deal with the missing values. However, most existing low-rank methods seldom consider mining the useful structures inside the low-rank matrix for a better denoising result. In addition, the rank number needs to be assigned manually. To address these problems, we propose an intracluster structured low-rank matrix analysis method for HSI denoising. First, we divide the original HSI into some clusters by taking advantages of both local similarity and nonlocal similarity structures, with which the resulted clusters are simpler and show more obvious low-rank property. Second, with singular value decomposition on the low-rank matrix in each cluster, the structured sparsity is modeled among the singular values to capture the structure of the low-rank matrix. Finally, an efficient optimization method is proposed to learn the structured sparsity adaptively from the data, as well as to inversely estimate the latent clean HSI from the noisy counterpart. The proposed method can not only obtain better denoising results compared with the-state-of-the-art methods but also automatically determine the rank number. Extensive experimental results demonstrate the effectiveness of the proposed method.

Index Terms—Hyperspectral denoising, low-rank analysis, singular value decomposition (SVD).

Manuscript received January 1, 2018; revised June 14, 2018; accepted July 27, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61671385 and Grant 61571354, in part by the Natural Science Basis Research Plan in Shaanxi Province of China under Grant 2017JM6021 and Grant 2017JM6001, and in part by the China Post-Doctoral Science Foundation under Grant 158201. (*Wei Wei and Lei Zhang contributed equally to this work.*) (*Corresponding author: Wei Wei.*)

W. Wei and Y. Zhang are with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: weiwei@nwpu.edu.cn; ynzhang@nwpu.edu.cn).

L. Zhang, Y. Jiao, and C. Wang are with the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: zhanglei211@mail.nwpu.edu.cn).

C. Tian is with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: chnatian@xidian.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.2018.2862384

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are 3-D data cube containing both spectral and spatial information [1], [2]. Compared with the traditional RGB image, the extra spectral information in HSI facilitates lots of remote sensing related applications such as classification [3], [4], anomaly detection [5]–[7], and so on. However, when they were acquired and transmitted, HSIs are unavoidably contaminated by noise [8]. Besides, the noise level varies across bands [9], which degrades the image quality and the completeness of HSIs [10]. Such noise enhances the difficulty of manual interpretation or machine understanding [11], [12], HSI denoising thus becomes a very important preprocessing step for further HSI processing.

HSI denoising methods can be roughly divided into noise-independent and noise-dependent methods according to the way that signal and noise combines [13]. These two kinds of methods are research focus of HSI denoising. We refer the readers to see [13] for more details on noise-dependent methods. In this paper, we mainly focus on the signal-independent noise corruption on HSIs.

For a signal-independent method, the relationship between the original clean/noise-free image $\mathcal{X} \in \mathbb{R}^{n_r \times n_c \times n_b}$ and noise $\mathcal{N} \in \mathbb{R}^{n_r \times n_c \times n_b}$ can be represented as $\mathcal{Y} = \mathcal{X} + \mathcal{N}$ [14], where $\mathcal{Y} \in \mathbb{R}^{n_r \times n_c \times n_b}$ denotes the noisy observation. n_r , n_c , and n_b represent the dimension of image height, width, and bands, respectively. The denoising task is to inversely estimate \mathcal{X} from \mathcal{Y} [15], [16]. However, without proper regularization/prior item, inferring \mathcal{X} from \mathcal{Y} is ill-posed. The inferred solution of \mathcal{X} will severely biased from the true \mathcal{X} . For HSI denoising, different prior knowledge (such as local smoothness) related with the image can be used to regularize the inferred image. Specifically, by representing \mathcal{X} and \mathcal{N} properly, an approximate solution of \mathcal{X} , denoted as $\hat{\mathcal{X}}$, can be inferred by implicitly or explicitly enforcing the prior on \mathcal{X} and with a specific designed optimization strategy [17]. For example, considering the local smoothness property, total variation (TV) regularization is often adopted for denoising. Yuan *et al.* [9] model a spectral and spatial adaptive TV regularizer to constrain X and use the Split Bregman optimization method to obtain \hat{X} . By transforming the original image into another representation space (such as wavelet domain) where the image signal lies in few significant coefficients (i.e., the signals are sparse) and the noise is

uniformly spread throughout all coefficients (i.e., the noise is nonsparse) [18], [19], typical sparse prior (e.g., ℓ_0 or ℓ_1 norms) is usually adopted to constrain \hat{X} [20], [21].

In recent years, the low-rank property has been shown as an effective regularization item on denoising since it has the advantage of being able to handle incomplete data with missing values [22]–[24]. For HSI, since adjacent bands in HSI exhibit strong correlations and nearby pixels are typically highly correlated, low-rank characteristic can be inherently considered an important prior knowledge for HSI [25], [26], which promotes some low-rank matrix recovery-based HSI denoising approaches [27], [28]. For example, Lu *et al.* [29] use the low-rank property to remove the stripe noise in HSI based on the high correlation among different bands. Considering HSI can be decomposed into some endmembers that are linearly combined with an abundance matrix, Zhang *et al.* [14] model HSI denoising as a low-rank matrix recovery problem and adopt an optimization algorithm named “go decomposition (GoDec)” [30] to obtain denoising results. To better regularize \mathcal{X} when solving the inverse problem, some methods are proposed using both the low-rank property and other regularization item. In [31], TV regularization is further introduced into the low-rank matrix analysis method to improve the HSI denoising performance. Zhao and Yang [28] combine low-rank constraint into sparse representation of HSI to remove Gaussian noise. Wang *et al.* [32] propose a group low-rank representation (LLR) denoising method employing the nonlocal similarity. In addition, the low-rank tensor-based method is further proposed by considering the 3-D data structure of HSI data [33], [34].

Although these methods function well on denoising, they consider the noise-free image as a low-rank data only without further mining its structure. How to exploit the structure inside the low-rank data to better regularize \mathcal{X} for denoising is seldomly studied, though structure has been demonstrated useful for denoising. In addition, the rank number needs to be assigned manually. However, an inappropriate assigned rank number will cause the denoising scheme to break down. To address these problems, an intracluster structured low-rank matrix analysis method is proposed in this paper. First, we divide the original HSI into some clusters taking advantages of both local and nonlocal similarity structures, with which the resulted clusters are simpler and show more obvious low-rank property. Second, in each cluster, we decompose low-rank matrix via singular value decomposition (SVD). The structured sparsity is then modeled among the singular values to capture the structure of low-rank matrix. Finally, an efficient optimization method is proposed to learn the structured sparsity adaptively from the data, as well as inversely estimate the denoising HSI from the noisy counterpart. The main ideas and contributions are summarized as follows.

- 1) We represent low-rank matrix with SVD. Based on this decomposition, a novel hierarchy sparsity prior is used to model the structure among singular values decomposed from low-rank matrix, which is equivalent to model the structure inside the low-rank matrix. Utilizing the structure inside the low-rank matrix not only helps to boost the denoising performance, but also has the merit

to automatically determine the rank number without manually assigning.

- 2) We divide the complicated content in HSI into some clusters, which are simpler and showing more obvious low-rank property, by utilizing the local and nonlocal spatial similarities inherently contained in the image. Dividing the whole image into small clusters guarantees that the HSI is better regularized with the low-rank property when it is inversely solved.
- 3) An efficient optimization method is proposed, in which the optimization of the denoising model is split into two subproblems, which are very easy to be optimized. More importantly, the optimization strategy enables the proposed method to be adaptive to the HSI data.
- 4) Compared with the state-of-the-art denoising methods, the proposed method obtains the best denoising effect.

The remainder of this paper is structured as follows. Section II describes the intracluster low-rank property. The proposed intracluster structured low-rank matrix analysis method for hyperspectral denoising is given in Section III. Experimental results and analysis are provided in Section V. Section VI concludes this paper.

II. INTRACLUSTER LOW-RANK CHARACTERISTICS

In this paper, 3-D HSI \mathcal{X} is rearranged into a 2-D matrix $X = [x_1, \dots, x_{n_p}] \in \mathbb{R}^{n_b \times n_p}$ for the convenient representation, where each row stacks a vectorized band image of the HSI. Each column x_i denotes the spectrum of one pixel. n_p is the number of pixels, i.e., $n_p = n_r \times n_c$. The same operation is applied on \mathcal{Y} and \mathcal{N} . Thus, X can be represented as $Y = X + N$ and the denoising task is to solve for X from Y .

From the perspective of linear mixing model, X is often assumed to be the multiplication of an endmember matrix (the assemble of pure materials) $M \in \mathbb{R}^{n_b \times n_e}$ and an abundance matrix $A \in \mathbb{R}^{n_e \times n_p}$, where n_e is the number of endmembers. Since the number of endmembers in an imaged scene is always limited, i.e., n_e is far smaller than n_b and n_p , X is thus considered as a low-rank matrix. The low-rank property can be exploited on the whole image X directly. However, HSI, especially remote sensing HSI, exhibits cluttered image content. Forcing the low-rank property directly on the whole image is too coarse to well regularize such content.

Although the content of HSIs may be complex, an HSI can be composed into some simple components which have obvious low-rank property. Thus, some methods were proposed by dividing the whole HSI image into collections of local patches and then enforcing the low-rank property on those patches separately. Compared with the whole image, local patches often contain simpler content and show more obvious low-rank property. Thus, they are easier to be regularized for recovery. However, due to the variety of image content (e.g., local patches also contains many dissimilar pixels) as well as the noise corruption, the amount of similar pixels within a local patch is rather limited, thus it is difficult to correctly remove the noise corruption, especially when the noise level is high. Therefore, it is necessary for an effective denoising approach to introduce extra information.

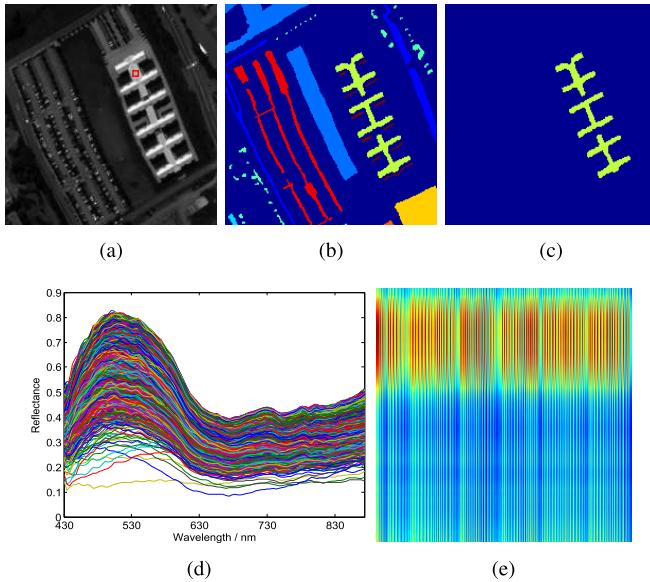


Fig. 1. Intracluster low-rank characteristics in an HSI. (a) Original band image with a selected pixel centered at the marked window. (b) Ground truth of clustering map. (c) Cluster map of the selected pixel. (d) Spectra in the selected cluster. (e) LLR of spectra in the selected cluster (e.g., each column denotes a spectrum). It can be seen that similar spectra from locally smooth areas and recurrent patterns gather into one cluster. Moreover, the intracluster similarity and correlation across spectrum result in obvious LLR.

To address this problem, considering that similar structures may repeat within the spatial domain of an HSI (i.e., nonlocal spatial similarity) [35], we introduce nonlocal spatial similarity as the desired extra information, and divide HSI into some components by utilizing both the local and nonlocal spatial similarities in this paper. To effectively generate those components without exhaustive searching, we employ the clustering method. It takes advantages of collecting both local and nonlocal spectral similar pixels effectively. Specifically, provided that the imaging scene contains K homogeneous areas, pixels in X thus can be grouped into K spatial clusters based on the spectral similarity. Let $X_k \in \mathbb{R}^{n_b \times n_k}$ denote the pixels in the k th cluster, where $k = 1, \dots, K$ and n_k is the number of pixels in this cluster. Given all X_k s, the HSI X can be represented as $X = [X_1, \dots, X_K]$ with a permutation on columns.

To show the low-rank property within the cluster, we select a homogeneous cluster from an HSI as Fig. 1(b) shows. The selected cluster and the intracluster spectra are shown in Fig. 1(c) and (d), respectively. Without loss of generality, the selected cluster is denoted as X_k . It can be seen that these continuous spectra not only exhibit obvious similarity (i.e., similarity among columns of X_k) but also show high correlation across each spectrum (i.e., correlation within each column of X_k), namely, X_k contains an obvious low-rank structure shown as Fig. 1(e).

Since the clustering-based method can generate more obvious low-rank structure, we propose to exploit the low-rank property within each cluster for HSI denoising, which is termed as an intracluster low-rank characteristic in this paper. Specifically, we divide the HSI into some clusters first. Then, based on the relationship in the following equation,

we estimate X_k from Y_k using the intracluster low-rank characteristic of X_k :

$$Y_k = X_k + N_k \quad (1)$$

where Y_k denotes the noisy observation, which combines the low-rank matrix X_k and noise N_k . Finally, we assemble the estimated X_k from different clusters into X , as the final denoising result.

It is noted that when the cluster number equals to one, the proposed method degenerates into the method that enforces the low-rank property on the whole HSI. The denoising experiments with different cluster numbers are given in Section V, from which we can see that the low-rank property can constrain X better when X is divided into some clusters.

III. INTRACLUSTER STRUCTURED LOW-RANK MATRIX ANALYSIS METHOD FOR HYPERSPECTRAL DENOISING

With the above-introduced intracluster low-rank characteristics, a novel low-rank matrix analysis-based HSI denoising method is proposed within each cluster in this section.

Similar to other denoising methods, we assume N comes from a zero-mean matrix normal distribution as $\mathcal{N}(\mathbf{0}, \Sigma_n, \mathbf{I})$, where $\Sigma_n = \lambda \mathbf{I}$. λ denotes the noise level and \mathbf{I} is an identity matrix with a proper size. According to the observation model, we have the following likelihood:

$$\begin{aligned} p(Y|X, \lambda) &\propto \exp\left(-\frac{1}{2\lambda}\|Y - X\|_F^2\right) \\ &= \prod_k \exp\left(-\frac{1}{2\lambda}\|Y_k - X_k\|_F^2\right) \end{aligned} \quad (2)$$

where $\|\cdot\|_F$ indicates the Frobenius norm.

Given the noise λ and proper priors, $p(X_k|\Theta_k)$, X can be estimated with the maximum *a posteriori* (MAP) estimation as

$$\max_X p(X|Y) \propto p(Y|X, \lambda) \prod_k p(X_k|\Theta_k) \quad (3)$$

where $p(X_k|\Theta_k)$ represents any kind of appropriate prior probability parameterized by Θ_k .

Applying the low-rank prior on X_k as well as the $-\log(\cdot)$ operation, (3) amounts to

$$\min_{\{X_k\}} \sum_k \frac{1}{2\lambda} \|Y_k - X_k\|_F^2 - \log p(X_k|\Theta_{lk}) \quad (4)$$

where $p(X_k|\Theta_{lk})$ denotes the low-rank prior on X_k and Θ_{lk} represents all parameters related to this prior.

Although the nuclear norm can represent the low-rank property of X_k , it is not easy to further capture the structure among the low-rank matrix. In addition, rank number needs to be assigned manually when nuclear norm is directly used to constraint the low-rank matrix. Thus, instead of using nuclear norm in this paper, we decompose the low-rank matrix via SVD and turn to explore the sparsity on singular values of X_k in this paper.

Provided that SVD on X_k is $X_k = U_k S_k V_k^T$, (4) can be reformulated as

$$\min_{\{S_k\}} \sum_k \frac{1}{2\lambda} \|Y_k - U_k S_k V_k^T\|_F^2 - \log p(S_k|\Theta_{lk}) \quad (5)$$

where $s_k \in \mathbb{R}^{R_k}$ denotes the singular value vector from the diagonal of S_k and $R_k = \min(n_b, n_k)$. $p(s_k | \Theta_{lk})$ denotes an appropriate prior for the singular vector s_k (i.e., consisting of singular values). It can be seen that solving (5) requires knowing U_k and V_k in advance. However, U_k and V_k are unknown in reality. Thus, we solve the minimization problem in (5) iteratively and adopt the following strategy to obtain U_k and V_k . Specifically, in the first iteration, we only have the observation image Y_k . Thus, we apply SVD on Y_k to obtain U_k and V_k , with which we can estimate S_k as well as the latent \tilde{X}_k as $\tilde{X}_k = U_k S_k V_k^T$. In the following iterations, we generate U_k and V_k based on the SVD of \tilde{X}_k obtained in the previous iteration.

With the SVD of X_k , the low-rank recovery problem in (4) is reduced to find a proper sparse singular vector s_k from (5) [36]. Thus, $p(s_k | \Theta_{lk})$ denotes an appropriate sparsity prior. However, it is difficult to solve (5) directly, especially for a complicated sparsity prior $p(s_k | \Theta_{lk})$ (e.g., a Gaussian mixture prior). In this paper, we adopt the half-quadratic splitting scheme to reduce (5) into two simple subproblems.

First, an auxiliary variable t_k is introduced into (5) as

$$\min_{\{s_k, t_k\}} \sum_k \frac{1}{2\lambda} \|Y_k - U_k S_k V_k^T\|_F^2 + \frac{\beta}{2} \|s_k - t_k\|_2^2 - \sum_k \log p(t_k | \Theta_{lk}) \quad (6)$$

where s_k is the 1-D representation of S_k and $p(t_k | \Theta_{lk})$ is a proper sparsity prior for t_k . β is a predefined scalar.

Then, $\{s_k\}$ can be obtained by alternatively solving the following two subproblems.

A. Sparse Learning for Rank Determination

Given $\{s_k\}$, we can estimate the sparse $\{t_k\}$ by solving the following problem:

$$\min_{\{t_k\}} \sum_k \frac{\beta}{2} \|s_k - t_k\|_2^2 - \log p(t_k | \Theta_{lk}). \quad (7)$$

Since the determined sparseness of t_k equals to the rank of the candidate matrix X_k , viz., exploiting the sparsity in t_k amounts to determining the rank of X_k , we term this procedure as the sparse learning for rank determination.

It can be proved that (7) amounts to a MAP inference for denoising each $\{s_k\}$ with a sparsity prior of $\{t_k\}$ independently [37]. The corresponding noisy observation can be represented as $s_k = t_k + \epsilon_k$, where ϵ_k is the noisy corruption. Provided that ϵ_k comes from a zero-mean Gaussian distribution, we have the following likelihood:

$$p(s_k | t_k, \epsilon_k) \propto \exp \left[-\frac{1}{2} (s_k - t_k)^T \Sigma_{\epsilon_k}^{-1} (s_k - t_k) \right] \quad (8)$$

where $\epsilon_k = [\epsilon_{1k}, \dots, \epsilon_{R_k k}]^T$ and $\Sigma_{\epsilon_k} = \text{diag}(\epsilon_k)$.¹

Considering the accurate estimation of $\{t_k\}$ given $\{s_k\}$ depends on an appropriate sparsity prior of $\{t_k\}$, we then focus on modeling such a sparsity prior. Different sparsity priors such as the Laplace prior can be used to capture the sparsity of t_k . However, the Laplace prior treats coefficients in

¹ $\text{diag}()$ denotes the “`diag()`” function in MATLAB.

Algorithm 1: Intracluster Low-Rank Matrix Analysis-Based HSIs Denoising

Input: Noisy observation Y , Cluster number K , noise level, and HSI $X = Y$.

Outer loop: For $n = 1, 2, \dots, N$ do

1. Divide pixels of X into K clusters with K -means method based on spectral similarity;
2. Low rank matrix X estimation:

Inner loop: for $t = 1, 2, \dots, T$ do

- (a) Sparse learning for rank determination by solving (13);
- (b) Low rank recovery as (26);

End for

End for

Output: Denoised HSI X .

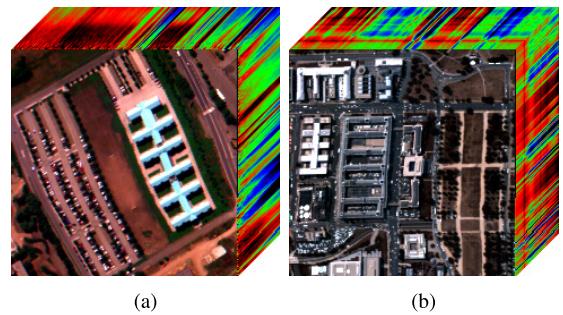


Fig. 2. 3-D cubes of (a) PaviaU and (b) Wdc in simulated noisy HSI experiments.

a sparse vector independently without modeling the underlying correlation among those coefficients. Recent progress in sparsity learning demonstrates that coefficients in a sparse signal are not independent but correlated, thus exhibiting a specific structure [38], [39]. For example, the sparse wavelet coefficients for an image shows a tree structure [40]. Moreover, exploiting such structure can greatly improve the accuracy of sparse learning [38]. Inspired by this, we propose to exploit the underlying structure in the sparse singular vectors $\{t_k\}$ to better model the low-rank characteristics in the matrices $\{X_k\}$. To this end, we employ the reweighted Laplace prior in [39] and [41] to model t_k , which has been proven to be able to capture well the underlying structure among those coefficients in a sparse signal.

First, a Gaussian distribution is imposed on t_k as

$$p(t_k | \gamma_k) \propto \exp \left(-\frac{1}{2} t_k^T \Gamma_k^{-1} t_k \right) \quad (9)$$

where $\gamma_k = [\gamma_{1k}, \dots, \gamma_{R_k k}]^T$ and $\Gamma_k = \text{diag}(\gamma_k)$.

Then, a Gamma distribution is adopted for γ_k as

$$p(\gamma_k | \kappa_k) \propto \prod_j \frac{\kappa_{jk}}{2} \exp \left(-\frac{\kappa_{jk} \gamma_{jk}}{2} \right). \quad (10)$$

In previous sparse learning works, the noise variance vector ϵ_k and the sparsity priors γ_k, κ_k are given beforehand,

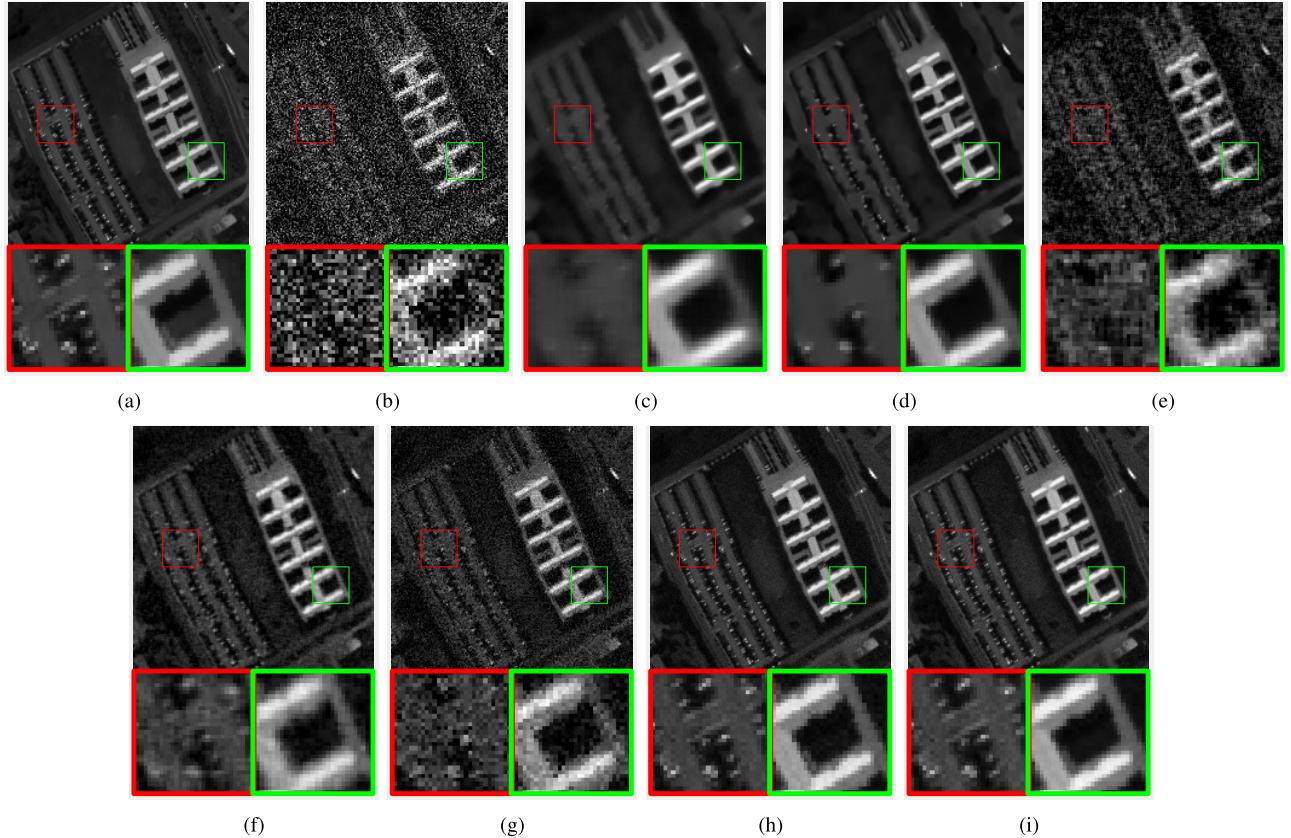


Fig. 3. Visual comparison on the denoised 17th band of PaviaU by different denoising methods (e.g., noise level $\sigma = 0.22$). (a) Original band. (b) Noisy band. (c) NLM3D. (d) BM4D. (e) PARAFAAC. (f) LRTA. (g) TensorDL. (h) CMESSC. (i) Ours.

which impose a fixed sparsity prior on the signal. In contrast, to make the prior data adaptive, we try to learn a data-specific sparse prior of $\{t_k\}$ from the intermediate $\{s_k\}$ in this paper. Specifically, we turn to learn those parameters from the given s_k with a latent variable-based Bayesian learning approach to adaptively learn a data-specific sparsity prior as

$$\begin{aligned} & \max_{\epsilon_k, \gamma_k, \kappa_k} \int p(s_k | t_k, \epsilon_k) p(t_k | \gamma_k) p(\gamma_k | \kappa_k) dt_k \\ &= \min_{\epsilon_k, \gamma_k, \kappa_k} \|s_k\|_{\Sigma_{bk}}^2 + n_k \log |\Sigma_{bk}| + \sum_j (\kappa_{jk} \gamma_{jk} - 2 \log \kappa_{jk}) \end{aligned} \quad (11)$$

where the latent variable t_k is integrated out. $\Sigma_{bk} = \Sigma_{\epsilon_k} + \Gamma_k$. To optimize (11), the following relation:

$$\|s_k\|_{\Sigma_{bk}}^2 = \min_{t_k} (s_k - t_k)^T \Sigma_{\epsilon_k}^{-1} (s_k - t_k) + t_k^T \Gamma_k^{-1} t_k \quad (12)$$

is utilized to obtain a unified variational framework

$$\begin{aligned} & \min_{t_k, \epsilon_k, \gamma_k, \kappa_k} (s_k - t_k)^T \Sigma_{\epsilon_k}^{-1} (s_k - t_k) + t_k^T \Gamma_k^{-1} t_k \\ &+ n_k \log |\Sigma_{bk}| + \sum_j (\kappa_{jk} \gamma_{jk} - 2 \log \kappa_{jk}) \end{aligned} \quad (13)$$

where the desired t_k and model parameters can be directly learned from the given s_k .

Equation (13) can be effectively solved by the algorithm proposed in [39] with an alternating optimization strategy.

Specifically, the problem in (13) is reduced to four subproblems and each subproblem only involves a group of unknown variables. During the optimization, all subproblems are iteratively optimized in an alternative scheme until convergence. In the following, we will introduce these four subproblems one by one.

- 1) *Sparse Signal Recovery*: Optimization for t_k . Through removing terms irrelevant to t_k in (13), the subproblem for t_k can be formulated as

$$\min_{t_k} (s_k - t_k)^T \Sigma_{\epsilon_k}^{-1} (s_k - t_k) + t_k^T \Gamma_k^{-1} t_k. \quad (14)$$

This is a standard ridge regression problem. The solution is reached when the gradient of the objective function over t_k is zero. Thus, the solution can be formulated as

$$t_k = (\Sigma_{\epsilon_k}^{-1} + \Gamma_k^{-1})^{-1} \Sigma_{\epsilon_k}^{-1} s_k. \quad (15)$$

In the light of the Woodbury matrix identity [42], the solution can be reformulated as

$$t_k = \Gamma_k (\Sigma_{\epsilon_k} + \Gamma_k)^{-1} s_k = \Gamma_k \Sigma_{bk}^{-1} s_k. \quad (16)$$

- 2) *Prior Learning*: Optimization for γ_k . Similarly, the subproblem for γ_k can be formulated as

$$\min_{\gamma_k} t_k^T \Gamma_k^{-1} t_k + n_k \log |\Sigma_{bk}| + \sum_j \kappa_{jk} \gamma_{jk}. \quad (17)$$

According to $\log |\Sigma_{bk}| = \log |\Sigma_{\epsilon_k} + \Gamma_k| = \log |\Sigma_{\epsilon_k}^{-1} + \Gamma_k^{-1}| + \log |\Sigma_{\epsilon_k}| + \log |\Gamma_k|$, the problem in (17) can be

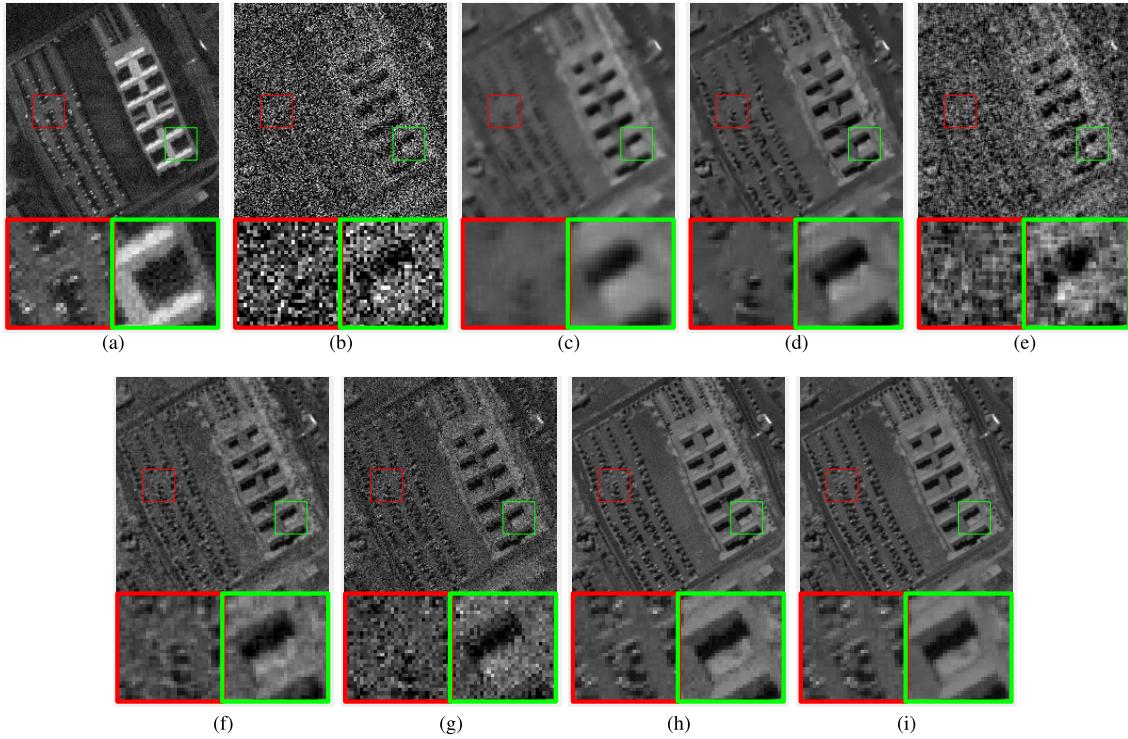


Fig. 4. Visual comparison on the denoised 76th band of PaviaU by different denoising methods (e.g., noise level $\sigma = 0.25$). (a) Original band. (b) Noisy band. (c) NLM3D. (d) BM4D. (e) PARAFAC. (f) LRTA. (g) TensorDL. (h) CMESSC. (i) Ours.

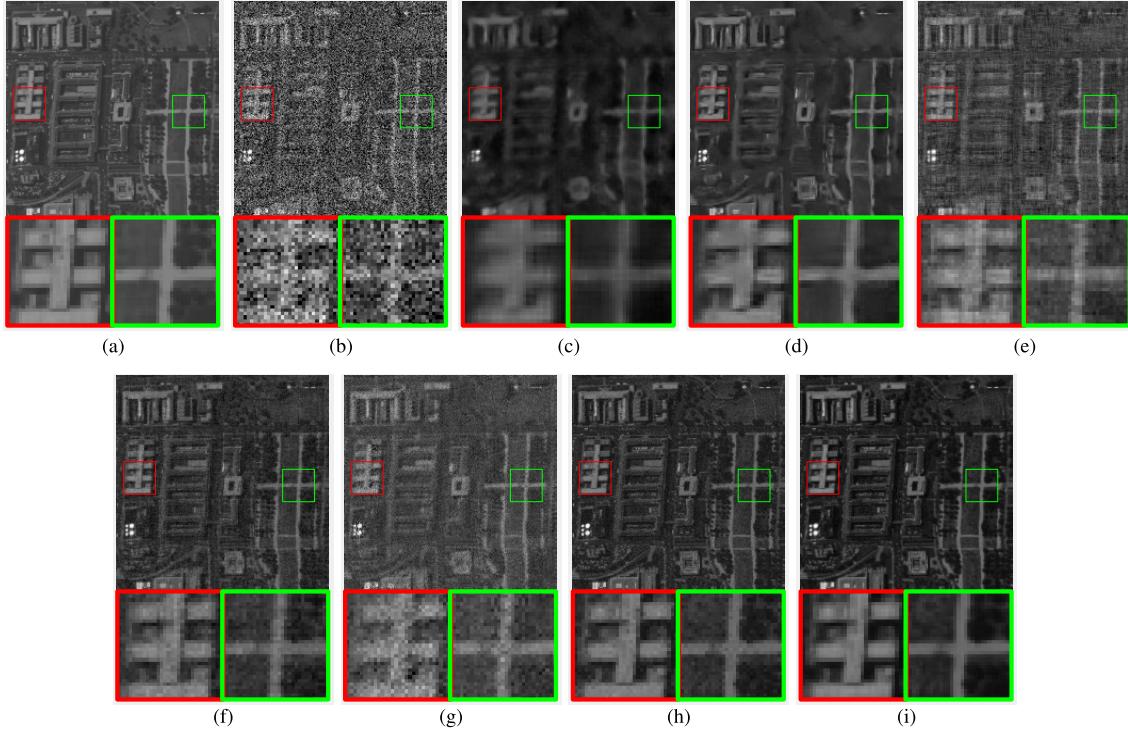


Fig. 5. Visual comparison on the denoised 35th band of Wdc by different denoising methods (e.g., noise level $\sigma = 0.14$). (a) Original band. (b) Noisy band. (c) NLM3D. (d) BM4D. (e) PARAFAC. (f) LRTA. (g) TensorDL. (h) CMESSC. (i) Ours.

reformulated as

$$\begin{aligned} \min_{\gamma_k} & t_k^T \Gamma_k^{-1} t_k + n_k \log |\Sigma_{\epsilon_k}^{-1} + \Gamma_k^{-1}| \\ & + n_k \log |\Gamma_k| + \sum_j \kappa_{jk} \gamma_{jk}. \quad (18) \end{aligned}$$

Although this problem is nonconvex, according to [39], we can obtain the following solution for each entry γ_{jk}

through leveraging the property of concave conjugate function for approximation:

$$\gamma_{jk} = \frac{\sqrt{4\kappa_{jk} n_k (\bar{t}_{jk}^2 + z_j) + n_k^2} - n_k}{2\kappa_{jk}} \quad (19)$$

where z_j denotes the j th entry of an intermediate vector $z = \text{diag}(\Gamma_k - \Gamma_k \Sigma_{bk}^{-1} \Gamma_k)$.

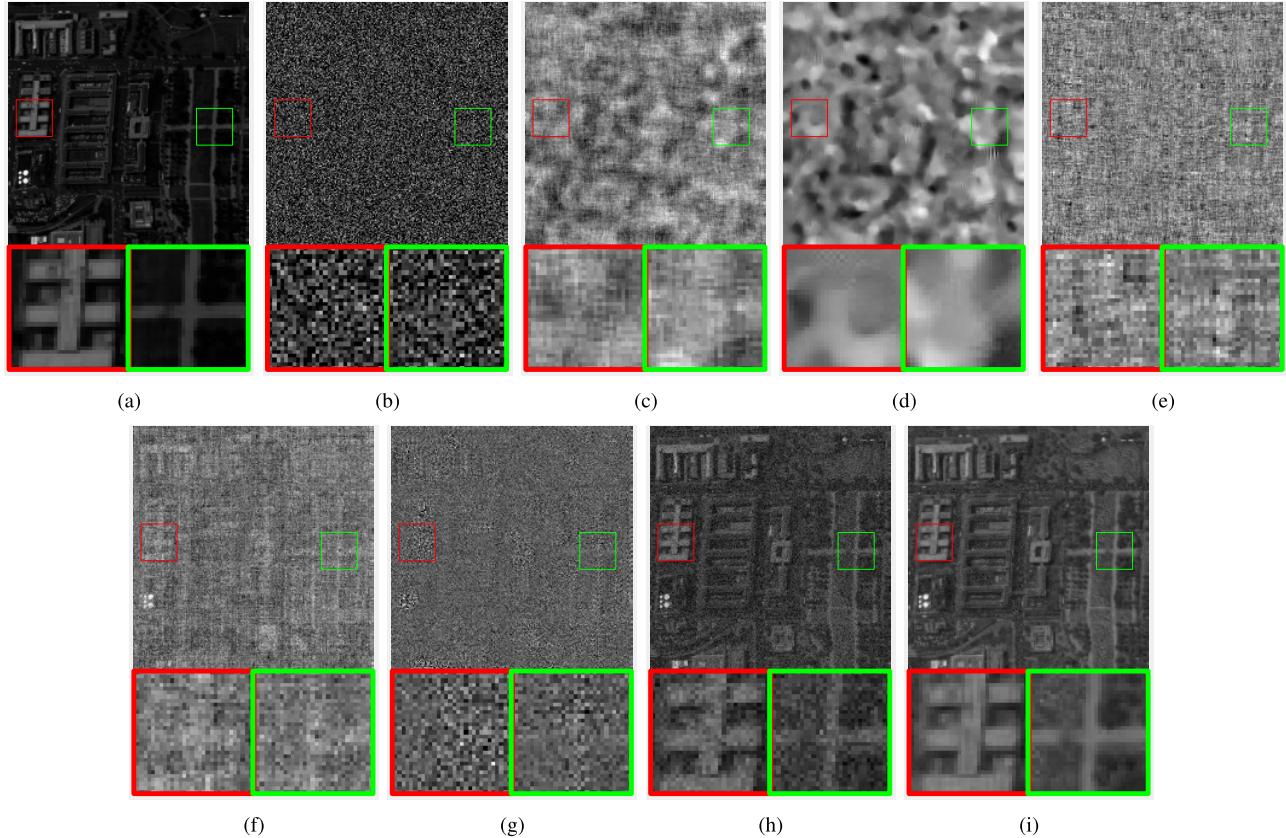


Fig. 6. Visual comparison on the denoised 150th band of Wdc by different denoising methods (e.g., noise level $\sigma = 0.29$). (a) Original band. (b) Noisy band. (c) NLM3D. (d) BM4D. (e) PARAFAC. (f) LRTA. (g) TensorDL. (h) CMESSC. (i) Ours.

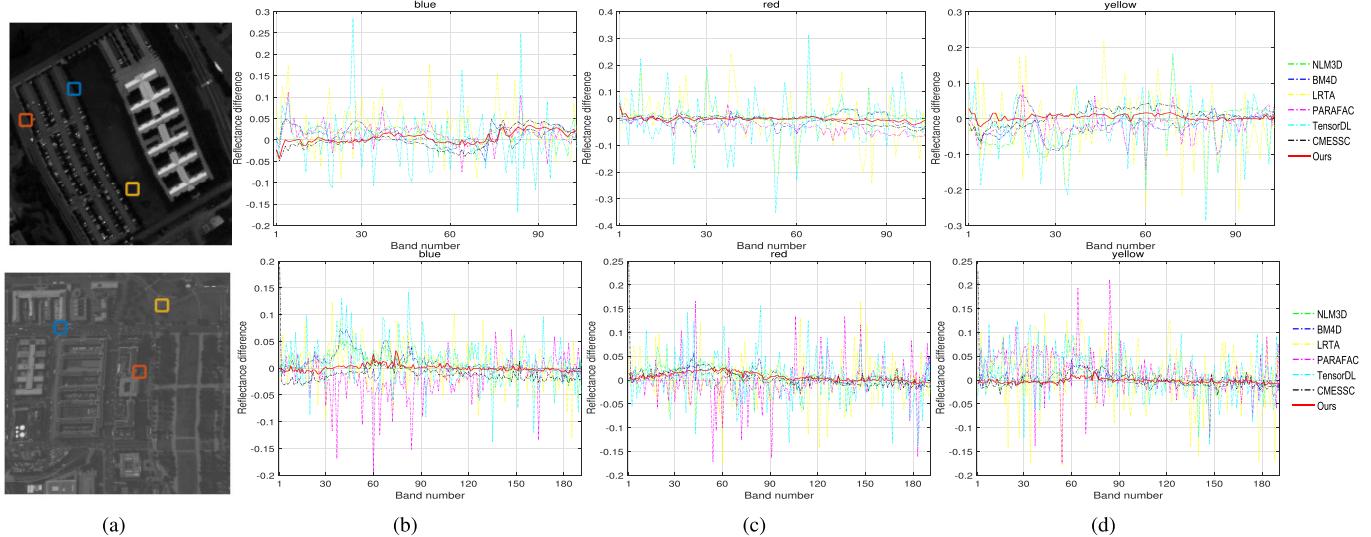


Fig. 7. Spectral reflectance difference curves of all methods on two HSIs. (a) HSI with three marked positions. (b)–(d) Curves corresponding to those three marked positions in each HSI. (First row) Curves on PaviaU. (Second row) Curves on Wdc.

3) *Hyperparameter Update*: Optimization for κ_k . The subproblem for κ_k can be given as

$$\min_{\kappa_k} \sum_j (\kappa_{jk} \gamma_{jk} - 2 \log \kappa_{jk}). \quad (20)$$

For each entry κ_{jk} , the problem in (20) is a convex optimization problem and the solution is reached when the gradient of objective function over κ_{jk} is zero. Thus,

the solution for κ_{jk} can be given as

$$\kappa_{jk} = 2 \gamma_{jk}^{-1}. \quad (21)$$

4) *Noise Estimation*: Optimization for ϵ_k . The subproblem for ϵ_k can be given as

$$\min_{\epsilon_k} (s_k - t_k)^T \Sigma_{\epsilon_k}^{-1} (s_k - t_k) + n_k \log |\Sigma_{bk}|. \quad (22)$$

Similar to [39], this problem can produce a closed-form solution through leveraging the property of concave

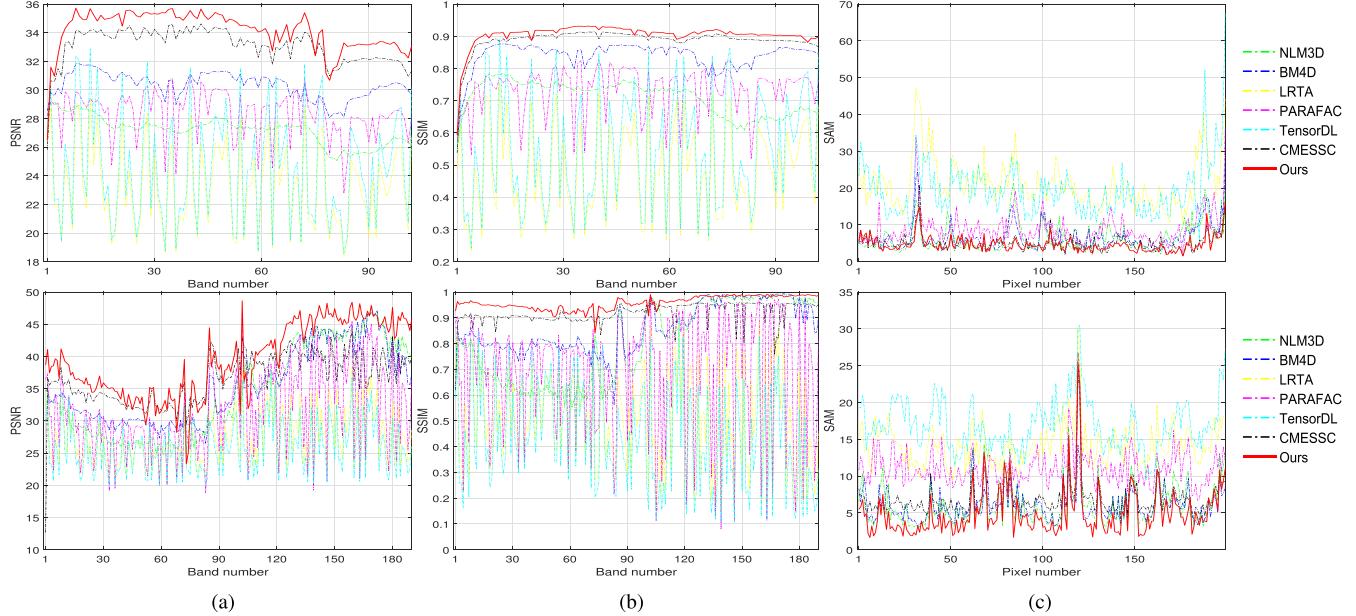


Fig. 8. PSNR and SSIM values of all denoising methods in each band of two HSIs as well as SAM values of the first 200 pixels. (First row) Results on PaviaU. (Second row) Results on Wdc. (a) PSNR. (b) SSIM. (c) SAM.

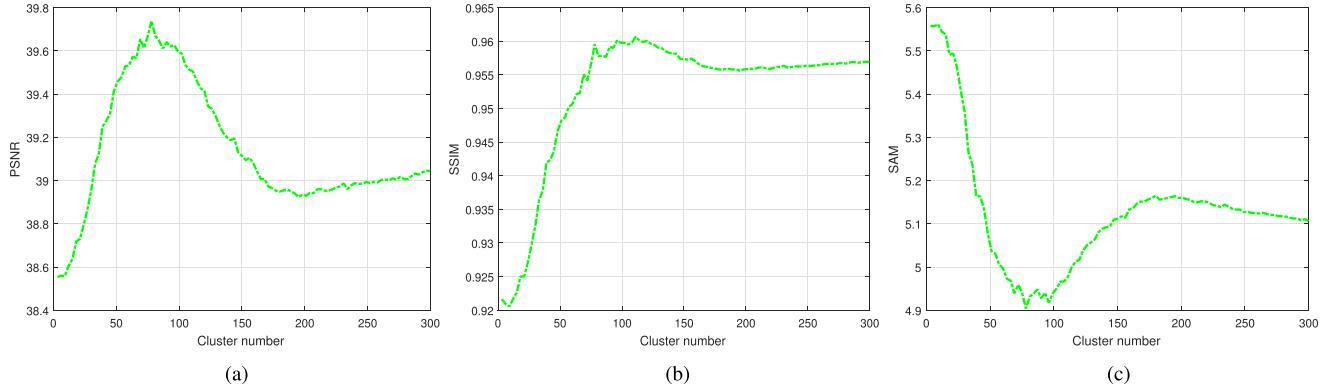


Fig. 9. Relationship between cluster number and (a) PSNR, (b) SSIM, and (c) SAM values.

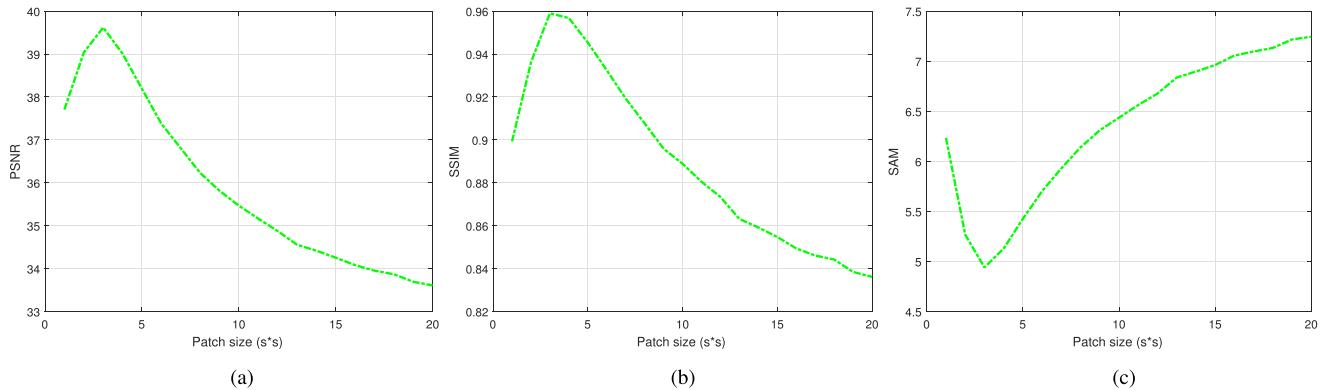


Fig. 10. Relationship between patch size and (a) PSNR, (b) SSIM, and (c) SAM values.

conjugate function for approximation, as

$$\epsilon_k = \sqrt{\bar{q}_j \alpha_j^{-1}}, \alpha = \text{diag}(\Sigma_{bk}^{-1}) \quad (23)$$

where α_j is the j th entry of vector α , \bar{q}_j is the j th entry of an intermediate vector $q = \text{diag}(QQ^T)$, and $Q = t_k - s_k$.

B. Low-Rank Recovery

Given $\{t_k\}$, $\{s_k\}$ can be obtained by solving the following problem:

$$\min_{\{s_k\}} \sum_k \frac{1}{2\lambda} \|Y_k - U_k s_k V_k^T\|_F^2 + \frac{\beta}{2} \|s_k - t_k\|_2^2. \quad (24)$$

TABLE I
AVERAGE PSNR, SSIM, AND SAM FOR DIFFERENT DENOISING METHODS ON PAVIAU DATA SET

Evaluation measure \ Method	NLM3D	BM4D	LRTA	PARAFAC	TensorDL	CMESSC	Ours
PSNR	27.2015	30.3406	23.2764	28.3625	24.5816	33.0632	34.1326
SSIM	0.7105	0.8403	0.4790	0.7338	0.5483	0.8872	0.9051
SAM	6.5797	6.6929	21.3776	8.8856	19.3914	5.0338	4.6821

TABLE II
AVERAGE PSNR, SSIM, AND SAM FOR DIFFERENT DENOISING METHODS ON WDC DATA SET

Evaluation measure \ Method	NLM3D	BM4D	LRTA	PARAFAC	TensorDL	CMESSC	Ours
PSNR	34.5779	35.6555	27.2191	31.5035	25.8538	36.5782	39.6198
SSIM	0.8056	0.8634	0.5028	0.6876	0.4310	0.9181	0.9590
SAM	6.2890	6.0168	14.2780	11.2130	16.9933	6.8627	4.9405

This is a quadratic optimization problem on s_k , which can be effectively solved by the conjugate gradient method. For simplicity, in our implementation, we turn to solve the following equivalent problem:

$$\min_X \sum_k \frac{1}{2\lambda} \|Y_k - X_k\|_F^2 + \frac{\beta}{2} \|X_k - U_k T_k V_k^T\|_2^2 \quad (25)$$

where T_k is the rectangular diagonal matrix of t_k . This quadratic optimization problem results in a closed-form solution on each X_k as

$$X_k = U_k T_k V_k^T + \tilde{\beta} Y_k \quad (26)$$

where $\tilde{\beta} \propto \beta\lambda$.

C. Algorithm

The whole procedure of the proposed intracluster low-rank matrix analysis method is summarized in Algorithm 1.

IV. EXPERIMENT

In this section, both the simulated and real noisy HSIs are adopted to demonstrate the superiority of the proposed method. Six state-of-the-art HSI denoising methods, including NLM3D [43], BM4D [44], LRTA [33], PARAFAC [45], TensorDL [8], and CMESSC [21], are used to compare with the proposed method. All these comparison methods are implemented with the codes published by the authors and are tuned to their best performance.

For the proposed method, we employ the K -means algorithm [46] to divide all pixels in each HSI into K homogeneous clusters, where the Euclidean distance between the spectra of two pixels is adopted as the distance metric. Considering that an individual pixel is easily affected by noise corruption, thus influencing the clustering accuracy, as well as the ultimate denoising result, we represent each pixel with a local cube (i.e., spectra lie in a fixed-size spatial local patch) centered at this pixel. In this paper, we use the patch size to denote the spatial size of the local cubes and set the patch size as 3×3 . We then conduct the K -means algorithm based on these local cubes, where the cluster number K is

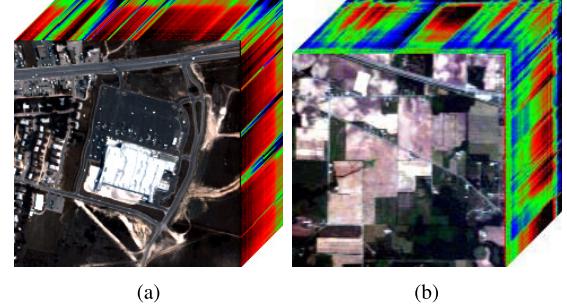


Fig. 11. 3-D cubes of urban and Indian Pine used in the experiments.
(a) PaviaU. (b) Wdc.

set as 93 and each local cube is reshaped into a vector to compute the Euclidean distance. In addition, through several trials of denoising the simulated data, we fix the parameter $\tilde{\beta}$ at 0.001 to give the best performance. In Algorithm 1, the noise level ϵ_k is initialized as a vector with all ones. The inner loop is terminated when either the maximum iteration number $N_{\max} = 100$ or the minimum update difference $\eta_{\min} = 10^{-3}$ is reached. The update difference is defined as $\|X_k^{\text{new}} - X_k\|_F / \|X_k\|_F$, where X_k^{new} and X_k denote the updated k th cluster in the current and last iterations. We only run one round of the outer loop in the experiment.

A. Simulated Data Experiments

1) *Data Set*: To evaluate the proposed method both quantitatively and visually, two simulated data including Pavia University data set and Washington DC Mall data set are used in this section. For simplicity, a subimage of size $200 \times 200 \times 191$ is cropped from Pavia University data set and a subimage of size $200 \times 200 \times 103$ is cropped from Washington DC Mall data set in this experiment. We denote these two subimages as PaviaU and Wdc, which can be seen in Fig. 2. Before the simulation process, each HSI image X is normalized into the range of $[0, 1]$. Then, we add zero-mean Gaussian noise to all bands of HSI to obtain the simulated noisy HSI Y . For different bands, different levels of noise are added and the corresponding noise standard deviation σ varies from 0.05 to 0.3. Given the noisy HSI Y , all denoising methods are then conducted to reconstruct the clean X .

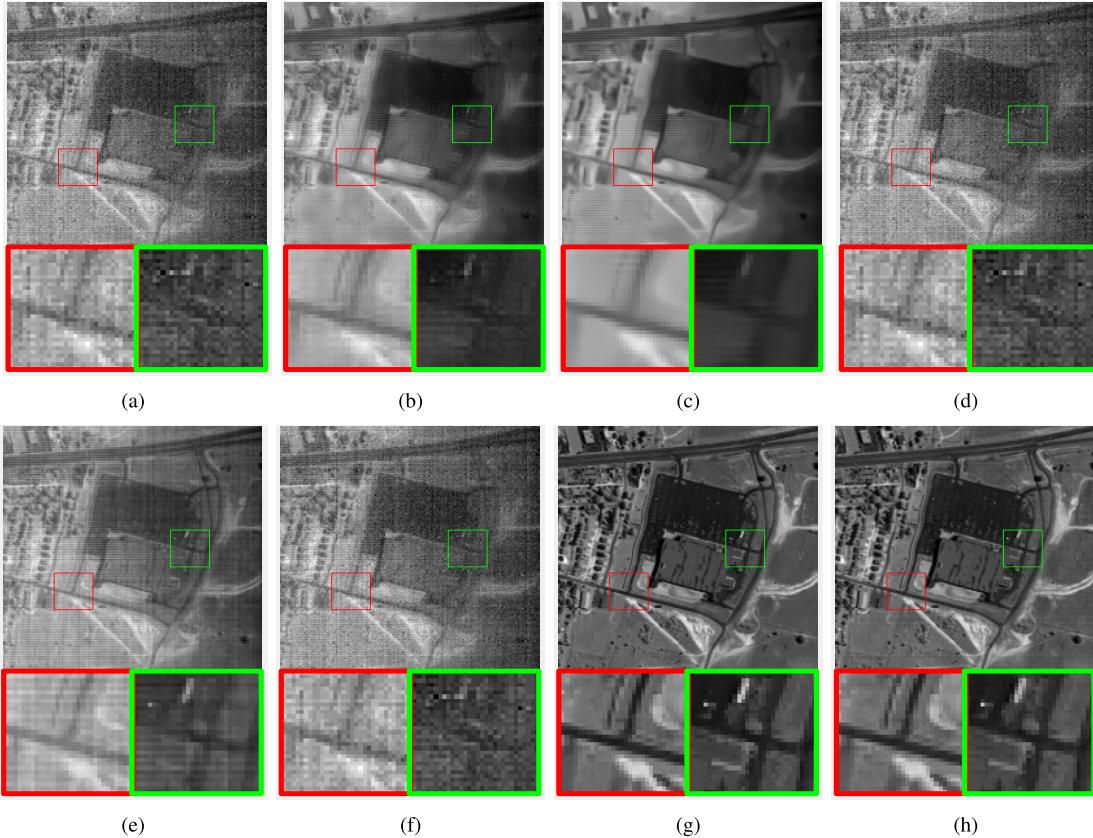


Fig. 12. Visual comparison on the denoised 104th band of urban by different denoising methods. (a) Original band. (b) NLM3D. (c) BM4D. (d) PARAFAAC. (e) LRTA. (f) TensorDL. (g) CMSSC. (h) Ours.

2) Evaluation Measures: Three HSI quality evaluation metrics, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM), and spectral angle mapper (SAM) [8], are adopted to evaluate the performance of different denoising methods in this paper. PSNR and SSIM are used to measure the similarity between two considered images according to their mean squared error and spatial structure discrepancy, respectively. SAM is proposed to measure the spectrum similarity of HSIs by calculating the average angle between the corresponding spectrum vectors at the same position from the two considered HSIs. Larger PSNR and SSIM, and smaller SAM indicate better denoising performance.

3) Superiority to Other Competitors: To clarify the superiority of the proposed low-rank method (denoted as “ours”), we illustrate the denoising results from all methods on typical bands of PaviaU and Wdc data sets, as shown in Figs. 3–6. Figs. 3 and 4 provide the denoising results on the 17th and 76th band images of PaviaU data set, where two different levels of noise are added and two interested areas are zoomed for detail comparison. In Fig. 3(a) and (b), the original noiseless band image and simulated noisy band image are shown, while the denoising results of NLM3D, BM4-D, LRTA, PARAFAAC, TensorDL, CMESSC, and ours are shown in Fig. 3(c)–(i), respectively. By comparing the original noiseless band image and the denoising visual results, we can see that the denoising results of the proposed method approximate the original noiseless band images. The proposed method can not only suppress

the noise but also preserves the details of the original HSI (e.g., parking zone enclosed in the red rectangle). Although NLM3D and BM4D function well on suppressing the noise, they fail to preserve the details of the original HSI to some extent, e.g., some details in the parking zone are missing (it can be seen more obviously in the zoomed-in area). On the contrary, LRTA and PARAFAAC preserve both the details and obvious noise. Although the newly proposed CMESSC method has a comparable visual result with the proposed method in Fig. 3, the proposed method has a clear superiority when handling with more severe noise corruption, as shown in Fig. 4.

Similarly, the denoised results of the 35th and 150th band images of Wdc data set are illustrated in Figs. 5 and 6. We can see that the results on these two noisy bands are similar to that in PaviaU data set and the proposed method still clearly outperforms other competing methods. In addition, the superiority is more obvious when noise corruption is heavy.

In addition, we further demonstrate the reconstruction results of the proposed method from the spectral perspective. In Fig. 7, the reconstruction error between the ground-truth spectrum and the reconstruction results is used. Smaller difference denotes better spectrum recovery results. Since thousands of pixels are contained in one HSI, we cannot visually illustrate all of them (the average reconstruction result calculated from all pixels can be seen from Tables I and II). Thus, we select three representative pixels to show the reconstruction results visually. To obtain a stable result, we calculate the

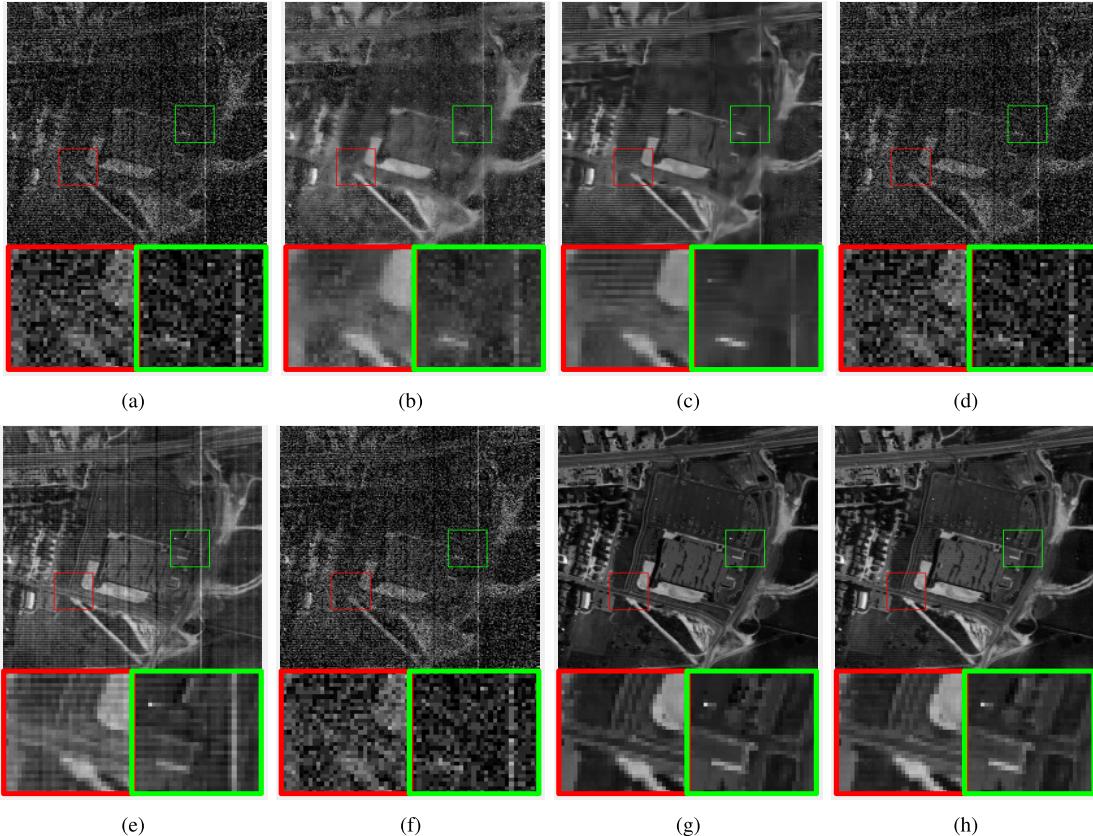


Fig. 13. Visual comparison on the denoised 207th band of urban by different denoising methods. (a) Original band. (b) NLM3D. (c) BM4D. (d) PARAFAC. (e) LRTA. (f) TensorDL. (g) CMEMSC. (h) Ours.

average spectral difference with a 3×3 neighboring area centered at that pixel instead of using the selected pixel individually. The selected three 3×3 neighboring areas are marked by rectangles in blue, red, and yellow, respectively, shown in Fig. 7(a). Fig. 7(b)–(d) denotes the average spectral difference within these three colored rectangles, respectively.

From the average spectral difference, we can see that the proposed method has the smallest average spectral difference, which is the closest to zero and indicates that the proposed method has the best spectral recovery results among all methods.

Besides, Fig. 8 illustrates the PSNR, SSIM, and SAM values on each band of two experimental HSI data sets where the horizontal axis represents the band number and the vertical axis represents the PSNR/SSIM/SAM values. Those values are obtained using the original noise-free image and the image obtained from different denoising methods. From Fig. 8, we can find that the proposed method has the highest PSNR, the highest SSIM, and the lowest SAM values among all methods in most bands of both PaviaU and Wdc data sets. In addition, compared with other methods such as NLM3D, BM4D, and PARAFAC, and so on, the proposed method performs stably on all measures for PaviaU data set. Although the proposed method only performs stably in terms of SSIM on Wdc, it still outperforms other methods in most bands for PSNR and SAM. To further clarify that the proposed method performs better than other competitors,

average numerical results on PSNR, SSIM, and SAM are given in Tables I and II. It can be seen that the proposed method obtains the best denoising performance among all methods. For example, compared with CMEMSC, the proposed method improves the PSNR by 0.99 db on PaviaU and 3.57 db on Wdc, and decreases the SAM values by 0.59° on PaviaU and 2.32° on Wdc.

From the above-mentioned results, we find that the proposed method performs better than others in denoising HSIs. Especially, it can be seen that compared with the traditional low-rank method without mining the structure inside the low-rank data (e.g., LRTA), the proposed structured low-rank matrix has better denoising performance both visually and quantitatively, which demonstrates the effectiveness of the modeling structure inside the low-rank matrix. In addition, compared with the method that models structured sparsity directly on the HSI data (e.g., CMEMSC), the modeling structure via the low-rank property in the proposed method can further boost the denoising performance.

4) Effect of Cluster Number and Patch Size: In the proposed method, cluster number and patch size are two important parameters. To testify the influence of these two parameters on denoising results, we conduct the following two experiments.

In the first experiment, we fix the patch size but set different cluster numbers uniformly distributed from 1 to 300. The experimental results on Wdc data set can be seen in Fig. 9. From Fig. 9, we can see that the average PSNR, SSIM, and SAM values are varied with different cluster numbers. When

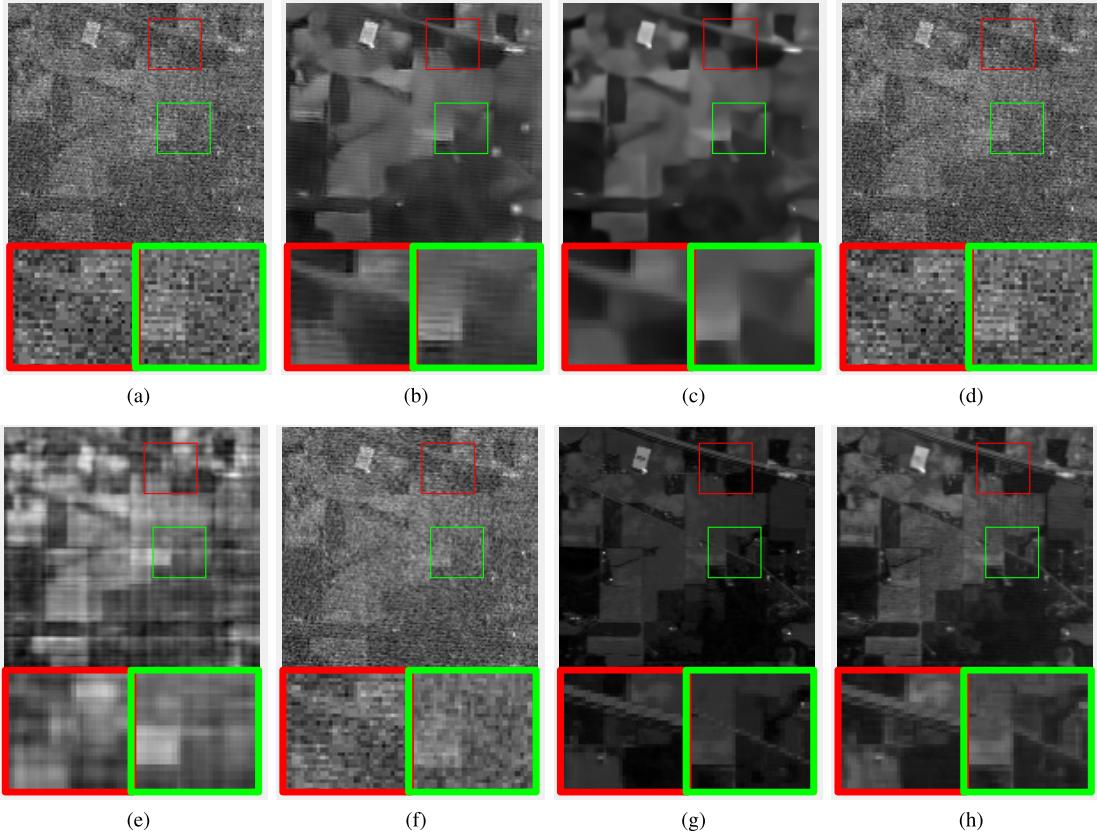


Fig. 14. Visual comparison on the denoised first band of Indiana Pine by different denoising methods. (a) Original band. (b) NLM3D. (c) BM4D. (d) PARAFAC. (e) LRTA. (f) TensorDL. (g) CMSSC. (h) Ours.

the cluster number is small, the denoising effect is relatively low, i.e., it has lower PSNR and SSIM, and larger SAM values. It is noted that when the cluster number equals to 1, the proposed method becomes the whole image-based low-rank method. With the increase in the cluster number, the performance of denoising can be improved, which suggests that clustering HSI into some clusters is effective for denoising. For Wdc data set, when the cluster number is around 93, all evaluation measures including PSNR, SSIM, and SAM have the best performance. For simplicity, in the following experiments, we set the cluster number as 93.

In the second experiment, we fix the cluster number but make the patch size to vary from 1×1 , 2×2 , ..., to 20×20 . The experimental result can be seen in Fig. 10. Compared with the pixel-level-based method (i.e., patch size is 1×1), the patch size varies from 3×3 to 6×6 results in better denoising results, which indicates the spatial information can benefit the denoising algorithm. However, when the patch size becomes larger, the denoising performance decreases. This is because it is hard to find very similar patches from HSI given a large patch, which will influence the low-rank property within the cluster. Considering both the performance and the computation cost, we set the patch size as 3×3 in the experiments.

5) Computational Cost Analysis: Taking PaviaU data set as an example, we report the runtime of all methods consumed in denoising in Table III. All methods are conducted in the MATLAB software on a computer with 8-GB RAM and 4 Intel

i5-2320 CPUs. It can be seen that the runtime of the proposed method is moderate among all competing methods. However, it is noticeable that our code is implemented in MATLAB and has not been carefully optimized. Moreover, since each cluster in the proposed method is processed independently, the proposed method can be further speeded up through exploiting the structured low-rank property within each cluster in a totally parallel manner.

B. Performance Evaluation on Real Data Set

To further testify the effectiveness of the proposed method on real data set, part of HYDICE urban data set of $200 \times 200 \times 210$ and the whole AVIRIS Indian Pines data set of $145 \times 145 \times 220$ are employed in the denoising experiments. They can be seen in Fig. 11.

First, we denoise the HYDICE urban data set on all bands with six kinds of contrast methods listed earlier to verify the effectiveness of our model. The experimental results on some typical bands including band 104 and band 207, which are corrupted with different levels of noise (stripe noise is also included in the band 207), are shown in Figs. 12 and 13. The actual band image is given in Figs. 12(a) and 13(a). The denoising results of NLM3D, BM4D, LRTA, PARAFAC, TensorDL, CMSSC, and the proposed method are shown in Figs. 12(b)–(e) and 13(b)–(e), respectively. From the visual results, we can see that the proposed method and CMSSC perform better than other methods. They can not only remove

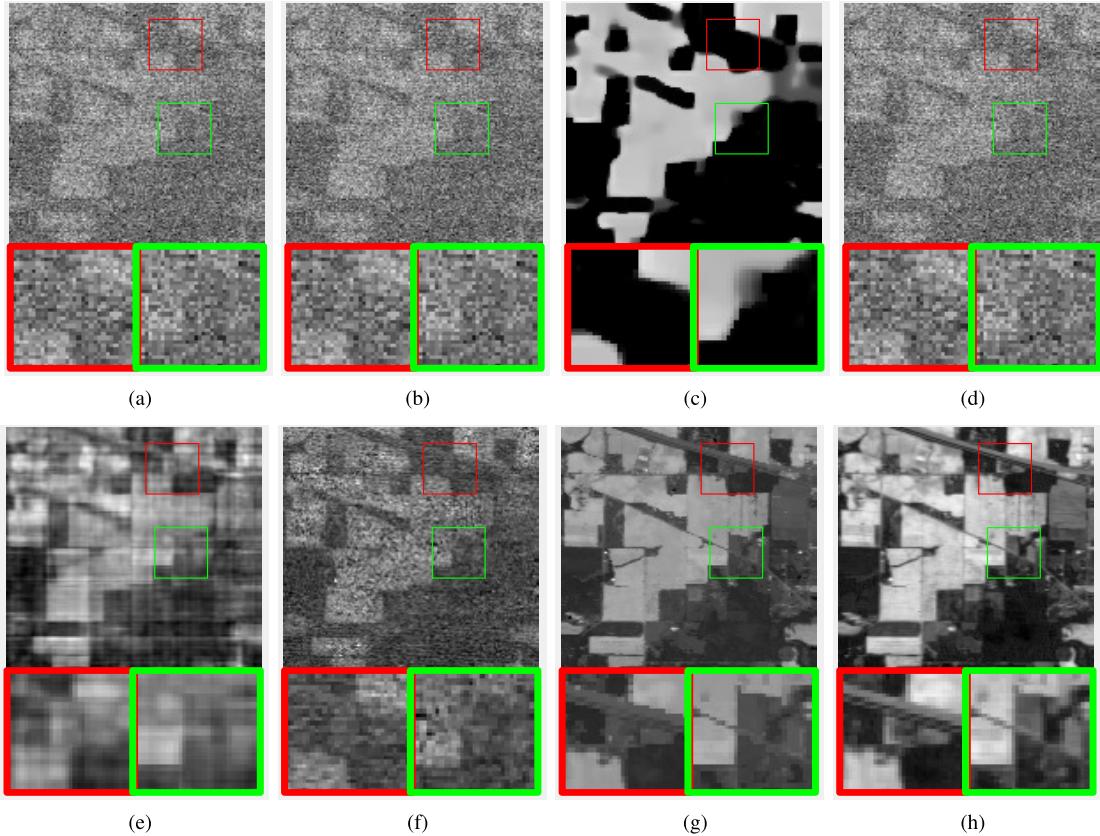


Fig. 15. Visual comparison on the denoised 152th band of Indiana Pine by different denoising methods. (a) Original band. (b) NLM3D. (c) BM4D. (d) PARAFAC. (e) LRTA. (f) TensorDL. (g) CMSSC. (h) Ours.

TABLE III
RUNTIME OF DIFFERENT METHODS IN DENOISING PAVIAU DATA SET

Method	NLM3D	BM4D	LATA	PARAFAC	TensorDL	CMSSC	OURS
Time(s)	1293.01	184.09	19.09	711.53	28.88	390.91	196.46

the Gaussian-like noise but also address the stripe noise very well.

Then, we apply different denoising methods on all bands for AVIRIS Indian Pines data set. Considering band 1 and band 152 have different kinds of noise, especially band 152 is severely corrupted, we select band 1 and band 152 to show the performance of different denoising methods. The results are illustrated in Figs. 14 and 15. From the visual results, we can see that the proposed method has the best performance. The proposed method not only suppresses the noise but also preserves the details of the original HSI. Similar to the experiments in the simulated data, the superiority of the proposed method is more obvious when confronting with more severe noise corruption (e.g., band 152). The reason can be partly attributed to the fact that the low rank-based method is inherently capable of reconstructing the spatial details. From previous experiments, we can conclude that the proposed method is appropriate for denoising the real noisy HSIs.

V. CONCLUSION

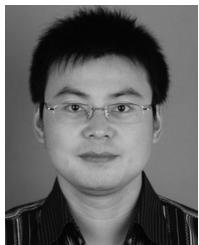
In this paper, we present a novel intracluster structured low-rank matrix analysis method to enhance the HSI denoising

performance. Local and nonlocal similarities together with structures inherently contained in the low-rank matrix are fully used to better regularize the HSI for denoising. By utilizing the local and nonlocal similarities of HSI, we divide HSI into groups of clusters, which are simpler and show more obvious low-rank property. Then, in each cluster, we attempt to capture the structure inside the low-rank matrix via SVD decomposition and structured sparsity. We further propose an efficient optimization method to obtain the denoised HSI adaptively from the noisy counterpart. Our experiments using both the simulated and real images reveal that the proposed method is comparable or better than several state-of-the art denoising methods. In addition, the proposed method can automatically determine the rank. In the future, we will seek better methods to capture the structure inside the low-rank matrix.

REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

- [2] C.-I Chang, "Hyperspectral target detection," in *Real-Time Progressive Hyperspectral Image Processing*. New York, NY, USA: Springer, 2016, pp. 131–172.
- [3] R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2693–2705, May 2017.
- [4] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Semisupervised hyperspectral classification using task-driven dictionary learning with laplacian regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1161–1173, Mar. 2015.
- [5] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [6] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Kernel sparse representation for hyperspectral target detection," *Proc. SPIE*, vol. 8390, p. 839005, Jun. 2012.
- [7] L. Zhang, Y. Zhang, H. Yan, Y. Gao, and W. Wei, "Salient object detection in hyperspectral imagery using multi-scale spectral-spatial gradient," *Neurocomputing*, vol. 291, pp. 215–225, 2018.
- [8] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2949–2956.
- [9] 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.
- [10] 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.
- [11] L. Zhang, W. Wei, Y. Zhang, H. Yan, F. Li, and C. Tian, "Locally similar sparsity-based hyperspectral compressive sensing using unmixing," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 86–100, Jun. 2016.
- [12] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.
- [13] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2957–2971, Aug. 2011.
- [14] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
- [15] 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.
- [16] M. Rizkinia, T. Baba, K. Shirai, and M. Okuda, "Local spectral component decomposition for multi-channel image denoising," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3208–3218, Jul. 2016.
- [17] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.
- [18] H. Yu, L. Zhao, and H. Wang, "Image denoising using trivariate shrinkage filter in the wavelet domain and joint bilateral filter in the spatial domain," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2364–2369, Oct. 2009.
- [19] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2458–2467, Jun. 2014.
- [20] W. Li, Q. Du, F. Zhang, and W. Hu, "Hyperspectral image classification by fusing collaborative and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4178–4187, Sep. 2016.
- [21] W. Wei, L. Zhang, C. Tian, A. Plaza, and Y. Zhang, "Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6860–6876, Dec. 2017.
- [22] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
- [23] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2009, Art. no. 11.
- [24] L. Zhang, W. Wei, Q. Shi, C. Shen, A. van den Hengel, and Y. Zhang, (2017). "Beyond low rank: A data-adaptive tensor completion method." [Online]. Available: <https://arxiv.org/abs/1708.01008>
- [25] S. Mei, Q. Bi, J. Ji, J. Hou, and Q. Du, "Hyperspectral image classification by exploring low-rank property in spectral or/and spatial domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2910–2921, Jun. 2017.
- [26] L. He, J. Li, A. Plaza, and Y. Li, "Discriminative low-rank Gabor filtering for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1381–1395, Mar. 2017.
- [27] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6688–6698, Oct. 2014.
- [28] Y.-Q. 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.
- [29] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, 2013.
- [30] T. Zhou and D. Tao, "GoDec: Randomized low-rank—sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Bellevue, WA, USA, Jun./Jul. 2011, pp. 33–40.
- [31] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.
- [32] M. Wang, J. Yu, J.-H. Xue, and W. Sun, "Denoising of hyperspectral images using group low-rank representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4420–4427, Sep. 2016.
- [33] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
- [34] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.
- [35] L. Zhang, W. Wei, Y. Zhang, C. Shen, A. van den Hengel, and Q. Shi, "Cluster sparsity field: An internal hyperspectral imagery prior for reconstruction," *Int. J. Comput. Vis.*, vol. 126, no. 8, pp. 797–821, 2018.
- [36] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, Jan. 2017.
- [37] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6236–6255, Sep. 2011.
- [38] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, Jan. 2011.
- [39] L. Zhang, W. Wei, C. Tian, F. Li, and Y. Zhang, "Exploring structured sparsity by a reweighted Laplace prior for hyperspectral compressive sensing," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4974–4988, Oct. 2016.
- [40] C. Chen and J. Huang, "Compressive sensing MRI with wavelet tree sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1115–1123.
- [41] L. Zhang, W. Wei, Y. Zhang, C. Shen, A. van den Hengel, and Q. Shi, "Dictionary learning for promoting structured sparsity in hyperspectral compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7223–7235, Dec. 2016.
- [42] K. B. Petersen *et al.*, "The matrix cookbook," *Tech. Univ. Denmark*, vol. 7, no. 15, p. 510, 2008.
- [43] Y. Qian, Y. Shen, M. Ye, and Q. Wang, "3-D nonlocal means filter with noise estimation for hyperspectral imagery denoising," in *Proc. Geosci. Remote Sens. Symp.*, 2012, pp. 1345–1348.
- [44] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Apr. 2013.
- [45] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.
- [46] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.



Wei Wei (SM'18) received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2012.

He is currently an Associate Professor with the School of Computer Science and Engineering, Northwestern Polytechnical University. He has authored more than 40 papers including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, *Pattern Recognition*, CVPR, ICCV, ECCV, AAAI, IJCAI, and so on. His research interests include image processing, machine learning, and pattern recognition.

Dr. Wei is a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE REMOTE SENSING LETTERS, and the IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEM, and so on. He served as the PC Member for around 10 major international conference including CVPR, ICME, and so on.



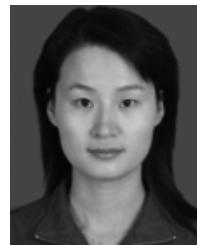
Lei Zhang (M'18) received the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2018.

He is currently a Research Staff with the School of Computer Science, The University of Adelaide, Adelaide, SA, Australia. His research interests include image processing and machine learning.



Yining Jiao received the B.S. degree in electronic science and technology from Honors College, Northwestern Polytechnical University, Xi'an, China, in 2017. She is currently pursuing the master's degree with the Medical Image Computing Laboratory, School of Biomedical Engineering, Shanghai Jiaotong University, Shanghai, China.

Her research interests include data-mining and machine learning in computer-assisted detection/diagnosis.



Chunna Tian (M'07) received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 2002, 2005, and 2008, respectively.

From 2006 to 2007, she was a Visiting Student with the Visual Computing and Image Processing Laboratory, Oklahoma State University, Stillwater, OK, USA. She is currently an Associate Professor with the School of Electronic and Engineering, Xidian University. Her research interests include image processing, machine learning, and pattern recognition.



Cong Wang received the M.Sc. degree from Northwestern Polytechnical University, Xi'an, China, in 2016, where she is currently pursuing the Ph.D. degree.



Yanning Zhang (SM'12) received the B.S. degree from the Dalian University of Technology, Dalian, China, in 1988, the M.S. degree from the School of Electronic Engineering, Northwestern Polytechnical University, Xi'an, China, in 1993, and the Ph.D. degree from the School of Marine Engineering, Northwestern Polytechnical University, in 1996.

She is currently a Professor with the School of Computer Science, Northwestern Polytechnical University. She is also a Cheung Kong Professor of Ministry of Education, China. She has authored more than 200 papers. Her research interests include remote sensing image analysis, computer vision and pattern recognition, and so on.

Dr. Zhang is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.