

Denoising of Hyperspectral Images Using Group Low-Rank Representation

Mengdi Wang, Jing Yu, *Member, IEEE*, Jing-Hao Xue, and Weidong Sun, *Member, IEEE*

Abstract—Hyperspectral images (HSIs) have been used in a wide range of fields, such as agriculture, food safety, mineralogy and environment monitoring, but being corrupted by various kinds of noise limits its efficacy. Low-rank representation (LRR) has proved its effectiveness in the denoising of HSIs. However, it just employs local information for denoising, which results in ineffectiveness when local noise is heavy. In this paper, we propose an approach of group low-rank representation (GLRR) for the HSI denoising. In our GLRR, a corrupted HSI is divided into overlapping patches, the similar patches are combined into a group, and the group is reconstructed as a whole using LRR. The proposed method enables the exploitation of both the local similarity within a patch and the nonlocal similarity across the patches in a group simultaneously. The additional nonlocally-similar patches can bring in extra structural information to the corrupted patches, facilitating the detection of noise as outliers. LRR is applied to the group of patches, as the uncorrupted patches enjoy intrinsic low-rank structure. The effectiveness of the proposed GLRR method is demonstrated qualitatively and quantitatively by using both simulated and real-world data in experiments.

Index Terms—Denoising, hyperspectral image, low-rank representation, nonlocal similarity.

I. INTRODUCTION

IN recent years, hyperspectral images (HSIs) have been widely used in various fields, such as agriculture, environment monitoring, food safety, medicine and mineralogy. However, corruption of HSIs by various type of noise, including Gaussian noise, salt-and-pepper noise, stripe noise and deadlines, degrades the image quality greatly, leading to low accuracy in classification, object segmentation, unmixing and subpixel mapping, among others. Therefore, the denoising process is an essential preprocessing step before the following analysis and applications of HSIs.

Over the past several decades, many denoising methods have been proposed under different frameworks, including wavelet filtering, principal component analysis (PCA), total variation, tensor analysis, sparse representation (SR) and low rank representation (LRR). Othman and Qian [1] reduced the noise using a wavelet-shrinkage filter in the hybrid spatial-spectral derivative domain. Maggioni *et al.* [2] employed three-dimensional collaborative filtering in a nonlocal transform

domain. Tensor analysis was employed in [3] and [4] for HSI denoising, with the former using the Tucker decomposition and the latter using rank-1 decomposition. Zhang *et al.* [5] proposed a cubic total variation (CTV) regularization and Yuan *et al.* [6] extended CTV to be spectral-spatial adaptive. Sparse representation has been widely used in natural image denoising [7]–[9] and has been introduced into HSI denoising with different types of regularization. Qian *et al.* [10] proposed a 3-D nonlocal sparse (3D-NLS) denoising method with effective utilization of the nonlocal spatial similarity under the SR framework. Lu *et al.* [11] improved the performance by simultaneously employing the high spectral correlation redundancy. The spectral correlation property has also been regularized by a low-rank constraint in [12]. Furthermore, several algorithms have been proposed by using combinations of the methods mentioned above [13]–[15].

Low-rank representation has been explosively used in image processing [16]–[26], as well as in HSI analysis [21]–[26] in recent years. Lu *et al.* [22] introduced LRR for the removal of stripe noise in HSI based on the high correlation between different bands, and a graph regularization is constructed to consider the local geometrical structure. Zhang *et al.* [23] proposed an HSI denoising method based on low-rank matrix recovery (LRMR). LRMR achieves outstanding performance as the uncorrupted HSIs comply highly with the low-rank assumption. He *et al.* [26] improved LRMR by using total-variation regularization. However, the methods in [23], [26] just employ the local similarity within patches of HSIs, while within some corrupted regions the patches may have heavy noise, which makes the local image structure be corrupted heavily and leads to an ill-posed problem for the structure restoration without introducing extra information.

In this paper, we propose a novel method for HSI denoising using group low-rank representation (GLRR), in which the local and nonlocal similarity of HSIs can be simultaneously considered under a unified framework of LRR. In GLRR, similar patches are combined into a group, and the reconstructed unit by LRR is the group of patches instead of the individual patches. Incorporating the nonlocally similar patches into LRR will bring in extra structural information to help the reconstruction of the corrupted structure. Experiments are conducted on both simulated and real-world data, and results show that our GLRR method outperforms the state-of-art methods. Extensive analysis and discussion is also conducted for the parameters involved in GLRR.

The remainder of this paper is organized as follows. Section II gives the proposed method in detail. Experimental results and discussion are shown in Section III. Section IV

M. Wang and W. Sun are with the State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: wangmd12@mails.tsinghua.edu.cn; wdsun@tsinghua.edu.cn).

J. Yu is with the College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China (e-mail: jing.yu@bjut.edu.cn).

J.-H. Xue is with the Department of Statistical Science, University College London, London WC1E 6BT, U.K. (e-mail: jinghao.xue@ucl.ac.uk).

draws the conclusion.

II. PROPOSED METHOD: GLRR

A. Low-rank property of HSIs

Given a cube of HSIs $\mathcal{D} \in \mathbb{R}^{M \times N \times B}$, where M and N are the spatial dimensions of an image and B is the number of bands, we can reformulate it as an $MN \times B$ matrix $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_{MN}]^T$, in which $\mathbf{d}_i \in \mathbb{R}^B$ denotes the B -variate spectral vector representing the i th pixel.

The low-rank property of HSIs can be described as

$$\text{rank}(\mathbf{D}) \ll \min\{MN, B\}, \quad (1)$$

where $\text{rank}(\cdot)$ denotes the rank of a matrix.

The property (1) can be illustrated by using the linear mixing model (LMM) [27]. Due to the limited spatial resolution of HSIs, within a pixel there is always multiple distinct materials (endmembers). Endmembers usually have relatively stable spectral characteristics termed signatures. Thus a spectrum of HSIs can be expressed as a mixture of a limited number of signatures. According to LMM, the systematic combination of signatures can be assumed to be linear:

$$\mathbf{d} = \sum_{j=1}^r a_j \mathbf{e}_j = \mathbf{E} \mathbf{a}, \quad (2)$$

where $\mathbf{e}_j \in \mathbb{R}^B$ is the signature of the j th endmember, r denotes the number of endmembers, all signatures together constitute a $B \times r$ signature matrix $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_r]$, a_j denotes the abundance of the j th endmember, and all abundances together constitute an abundance vector of r elements, $\mathbf{a} = [a_1, \dots, a_r]^T$, for the spectrum \mathbf{d} .

As \mathbf{E} is a $B \times r$ matrix, the rank of \mathbf{E} is smaller than r :

$$\text{rank}(\mathbf{E}) \leq r. \quad (3)$$

It follows that the whole HSI cube, as the matrix \mathbf{D} , can be represented as

$$\begin{aligned} \mathbf{D} &= [\mathbf{d}_1, \dots, \mathbf{d}_{MN}]^T = [\mathbf{E} \mathbf{a}_1, \dots, \mathbf{E} \mathbf{a}_{MN}]^T \\ &= [\mathbf{a}_1, \dots, \mathbf{a}_{MN}]^T \mathbf{E}^T = \mathbf{A} \mathbf{E}^T, \end{aligned} \quad (4)$$

where \mathbf{A} is an $MN \times r$ matrix of abundance. Therefore, the rank of \mathbf{D} has the following property:

$$\text{rank}(\mathbf{D}) \leq \text{rank}(\mathbf{E}) \leq r \ll \min\{MN, B\}, \quad (5)$$

which means that matrix \mathbf{D} is of low rank, or say the dimension of the spectral space spanned by the columns (i.e. the spectral bands) of \mathbf{D} is low, lower than the number of endmembers.

B. LRR-based denoising

The matrix \mathbf{D} is of low rank, thus the restoration of noisy HSIs can be modelled as a low-rank matrix representation and recovery problem, for which the LRR-based denoising approach has achieved excellent performance [23], [26].

In an observed HSI cube $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$, the $q \times q$ patch centred at (i, j) is denoted by $\mathcal{X}_{ij} \in \mathbb{R}^{q \times q \times B}$, and its

corresponding 2D matrix is denoted by $\mathbf{X}_{ij} \in \mathbb{R}^{q^2 \times B}$. This matrix can be decomposed as

$$\mathbf{X}_{ij} = \mathbf{D}_{ij} + \mathbf{E}_{ij}, \quad (6)$$

where \mathbf{D}_{ij} is the clean patch and \mathbf{E}_{ij} denotes the noises.

The noises can be divided into two classes according to the density of their distributions: sparse noise and dense noise, in which the sparse noise mainly contains stripe noise and salt-and-pepper noise, and the dense noise is typically Gaussian noise. Thus (6) can be further modelled as

$$\mathbf{X}_{ij} = \mathbf{D}_{ij} + \mathbf{S}_{ij} + \mathbf{N}_{ij}, \quad (7)$$

where \mathbf{S}_{ij} models the sparse noise and \mathbf{N}_{ij} is the dense noise. They are denoted by $\mathbf{X}, \mathbf{D}, \mathbf{S}, \mathbf{N}$ in the following for brevity.

As often only few bands were corrupted by stripe or salt-and-pepper noise, there should be just a limited number of nonzero elements in \mathbf{S} (the pseudo-norm $\|\cdot\|_0$ is used here with $\|\mathbf{S}\|_0$ denoting the number of nonzero elements in \mathbf{S}). Based on the low rank of \mathbf{D} and the sparsity of \mathbf{S} , in LRR the restoration of \mathbf{D} can be achieved by solving the following optimization problem:

$$\{\hat{\mathbf{D}}, \hat{\mathbf{S}}\} = \underset{\mathbf{D}, \mathbf{S}}{\text{argmin}} \|\mathbf{X} - \mathbf{D} - \mathbf{S}\|_F^2 + \lambda \text{rank}(\mathbf{D}) + \gamma \|\mathbf{S}\|_0, \quad (8)$$

where λ and γ are the parameters for the trade-off between sparsity and low rank. We can alter the optimization problem as

$$\begin{aligned} \{\hat{\mathbf{D}}, \hat{\mathbf{S}}\} &= \underset{\mathbf{D}, \mathbf{S}}{\text{argmin}} \|\mathbf{X} - \mathbf{D} - \mathbf{S}\|_F^2 \\ \text{s.t. } &\text{rank}(\mathbf{D}) \leq r, \quad \|\mathbf{S}\|_0 \leq k_s, \end{aligned} \quad (9)$$

in which r , the number of endmembers, is the rank constraint on \mathbf{D} , and k_s , an index of the amount of sparse noises, is the sparse constraint on \mathbf{S} . As r and k_s have natural physical interpretation, they can be determined based on the characteristics of the HSIs. Problem (9) can be solved using bilateral random projection (BRP) [28], [29].

C. GLRR-based denoising

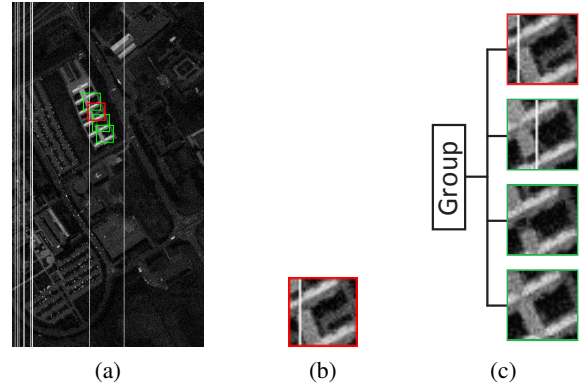


Fig. 1. Noise detection mechanisms in LRR and GLRR: (a) group patches in HSIs, (b) LRR-based denoising and (c) GLRR-based denoising.

1) *Grouping nonlocally similar patches:* The traditional LRR-based denoising method deals with the patches individually. However, for a heavily corrupted patch (which has too

much noise), its spatial structure information may have been lost, leading to difficulty in recovering \mathbf{D} from \mathbf{X} in (7) if no auxiliary information is introduced.

To solve this problem, we proposed a novel denoising method termed GLRR, in which nonlocally similar patches are assembled into a group and the group is collaboratively represented by using LRR. As shown in Fig.1(a) and (c), the patch represented by the red rectangle is the to-be-reconstructed patch, and the patches represented by the green rectangles are the most similar patches found in the image. The benefits from introducing the “group” concept here are twofold. Firstly, the addition of similar patches will bring in extra spatial information to help reconstruct the corrupted spatial structure in the to-be-reconstructed patch shown by the red rectangle. Secondly, the stripe noise will be more sparse over the whole group of patches than over the noisy patch itself (which is much smaller than the whole group), as shown by the comparison between Fig.1(c) and (b), and the improved sparsity can help \mathbf{S} more easily to be separated from \mathbf{X} in model (7).

In the grouping process of GLRR, the Euclidean distance is employed as the measure of similarity: for a total of K matrices (patches) $\mathbf{X}_k \in \mathbb{R}^{m \times n}$, for $k = 1, \dots, K$, the Euclidean distance between matrices \mathbf{X}_i and \mathbf{X}_j is defined as

$$d_{ij} = d(\mathbf{X}_i, \mathbf{X}_j) = \left(\sum_{u=1}^m \sum_{v=1}^n |x_{uv}^{(i)} - x_{uv}^{(j)}|^2 \right)^{1/2}, \quad (10)$$

where i and j are the indices of the matrices, and $x_{uv}^{(i)}$ and $x_{uv}^{(j)}$ are the entries of \mathbf{X}_i and \mathbf{X}_j located at (u, v) , respectively. Then the group $\mathcal{G}(\mathbf{X}_k)$ for patch \mathbf{X}_k is formed using its l nearest patches.

Algorithm 1 GLRR-based HSI Denoising

Input:

Corrupted HSI $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$

Output:

Reconstructed HSI $\mathcal{D} \in \mathbb{R}^{M \times N \times B}$

Step 1: Divide \mathcal{X} into patches $\{\mathbf{X}_k\}_{k=1}^K$ and, for each patch \mathbf{X}_k , form a group matrix $\mathbf{X}_k^{(\mathcal{G})}$ for the group $\mathcal{G}(\mathbf{X}_k)$ constructed by using (10);

Step 2: Restore the low-rank structure $\mathbf{D}_k^{(\mathcal{G})}$ from $\mathbf{X}_k^{(\mathcal{G})}$ using the LRR framework described in (9);

Step 3: Reconstruct the HSI as $\mathcal{D} \in \mathbb{R}^{M \times N \times B}$ by weighted averages using (11) and (12).

2) *The GLRR framework of denoising:* The framework of the proposed GLRR-based denoising method is described in Algorithm 1. The method can be decomposed into three steps: data arrangement, LRR of the group, and HSI reconstruction.

Firstly, an HSI $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$ is divided into K overlapping patches $\{\mathbf{X}_k \in \mathbb{R}^{q^2 \times B}\}_{k=1}^K$ (reshaped from the 3D cube) with spatial size $q \times q$ and sliding step s (set as $s = q/3$ in this paper), thus we can get that $K = \lfloor \frac{M-q}{q/3} \rfloor \lfloor \frac{N-q}{q/3} \rfloor$, where $\lfloor \cdot \rfloor$ is the downward rounding operator. Each patch $\{\mathbf{X}_k\}$ is assembled with its l most similar patches, according to the distance defined in (10), into a matrix group $\mathcal{G}(\mathbf{X}_k)$, which

is then rearranged into a bigger matrix $\mathbf{X}_k^{(\mathcal{G})} \in \mathbb{R}^{m \times B}$ with $m = (l+1) \times q^2$.

Secondly, low-rank structure $\mathbf{D}_k^{(\mathcal{G})}$ of $\mathbf{X}_k^{(\mathcal{G})}$ is reconstructed by using LRR as with (9). Then the low-rank matrix group $\mathcal{G}(\mathbf{D}_k)$ corresponding to $\mathcal{G}(\mathbf{X}_k)$ can be extracted from $\mathbf{D}_k^{(\mathcal{G})}$.

Finally, the denoised HSI is reconstructed by a two-stage average of the restored patches. At the first stage, each patch is reconstructed by the weighted average of multiple reconstruction results. As each patch may be included in different groups, *e.g.* \mathbf{X}_k may be covered by C groups $\{\mathcal{G}(\mathbf{X}_{k_1}), \dots, \mathcal{G}(\mathbf{X}_{k_C})\}$, there will be multiple reconstructed patches for \mathbf{X}_k , denoted as $\{\mathbf{D}_k^{(k_1)}, \dots, \mathbf{D}_k^{(k_C)}\}$, where k_i ($i = 1, \dots, C$) is the group index. The reconstructed matrix for the k th patch \mathbf{X}_k is therefore estimated by the following weighted average:

$$\mathbf{D}_k = \sum_{i=1}^C \omega_i \mathbf{D}_k^{(k_i)} / \sum_{i=1}^C \omega_i, \quad (11)$$

$$\omega_i = 1 / (1 + \tau d_{kk_i}), \quad (12)$$

where τ is an inverse parameter from distance measurement to similarity measurement and d_{kk_i} is the distance between patches \mathbf{X}_k and \mathbf{X}_{k_i} . At the second stage, the denoised HSI \mathcal{D} is obtained as the average of the overlapped patches.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Both simulated data and real-world data are employed in our experiments. To evaluate the effectiveness of our GLRR method, BM4D [2] and LRMR [23] are used as benchmark methods in the experiments. BM3D [30] has reached the state-of-art performance in the natural image denoising to our knowledge, and BM4D is an extension of BM3D to volumetric data. LRMR is actually an LRR-based denoising method, with only the local similarity of HSIs considered. In addition, further experiments and discussion are conducted on the impact and determination of the parameters involved in our method.

A. Experimental results on simulated data

Reflective optics system imaging spectrometer (ROSIS) images of the Pavia University, Italy is used for our experiments on the simulated data. The spatial size of the data is 610×340 , and there are 103 bands in the spectral range of $0.43 \mu\text{m}$ to $0.86 \mu\text{m}$.

In the simulation experiment, the original data is added with Gaussian noise, salt-and-pepper noise and stripe noise, of which the last one is typical in the HSIs acquired by pushbroom sensors. Gaussian noise is added to all the bands with $\sigma = 5\%$; salt-and-pepper noise and stripe noise are randomly added to 10 bands; the percentage of salt-and-pepper noise is 20%; and the number of stripe lines is 10 in each band. Additionally, there are 2 bands that are added with all the three kinds of noises, one of which is shown Fig.2(b), with Fig.2(a) denoting the original clean band.

Experimental results of band 83 in the Pavia University using BM4D, LRMR and our GLRR method are shown in

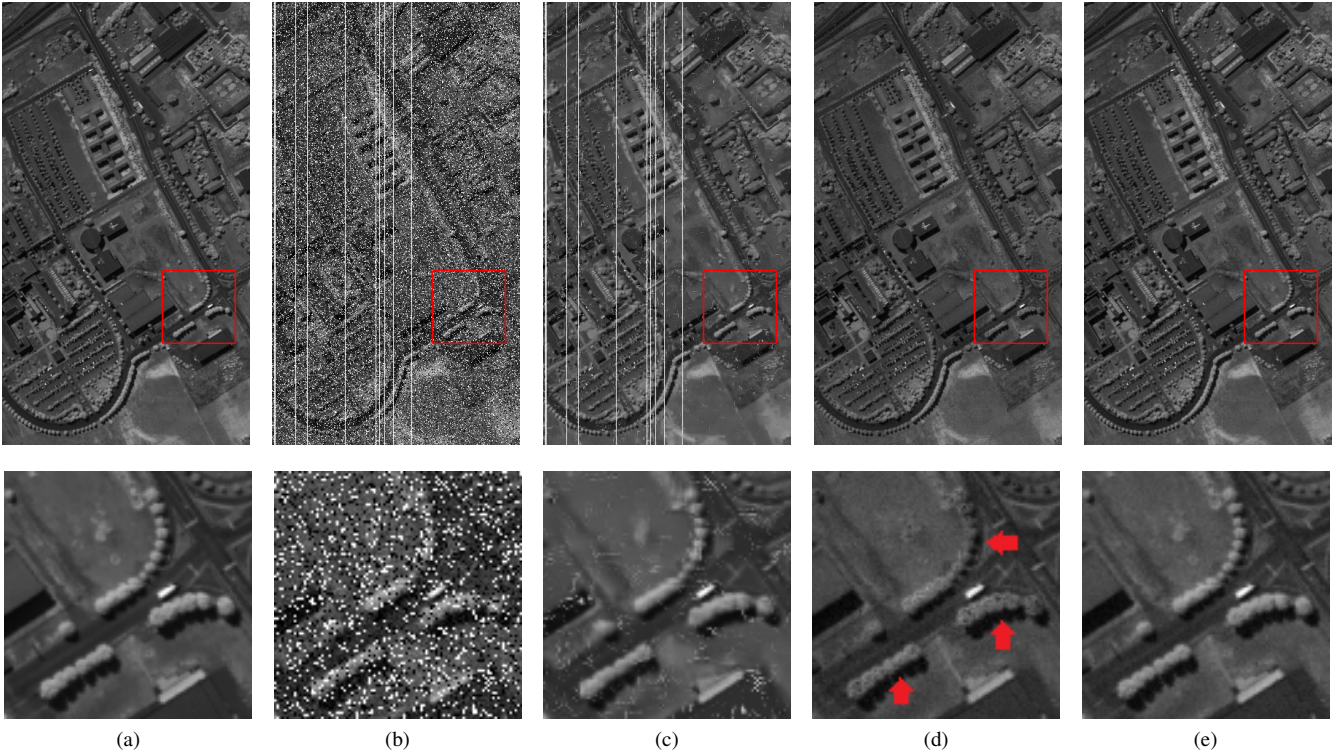


Fig. 2. Results of band 83 in the Pavia University data: (a) original band; (b) noisy band (PSNR=11.02 dB, SSIM=0.1006); (c) BM4D result (PSNR=22.49 dB, SSIM=0.7234); (d) LRMR result (PSNR=27.27 dB, SSIM=0.8460); (e) GLRR result (PSNR=35.68 dB, SSIM=0.9458). Upper row: whole band; lower row: closeup.

Fig.2(c)-(e), respectively. In GLRR, the parameters are set as: blocksize $q = 30$, rank constraint $r = 20$ and sparsity $k_s = 12000$ (See section III-C for analysis of the determination of these values). The source code of BM4D can be obtained from [2] which uses adaptive estimated noise variation σ , and the parameters of LRMR are set according to the discussion in [23].

It can be observed in Fig.2(c) that BM4D cannot remove the stripe noise and the salt-and-pepper noise thoroughly from the images, as this method was intuitively proposed for the removal of Gaussian noise and Rician noise. Moreover, it tends to over-smooth the denoised area due to the collaborative Wiener filtering procedure, which can be shown in the closeup in Fig.2(c). LRMR performs much better than BM4D, however it cannot remove the Gaussian noise and the salt-and-pepper noise totally, leading to failure in the reconstruction of some fine structures, as indicated by the red arrows in Fig.2(d). This is because that within the local area the noise is so much as to heavily corrupt the local spatial structure, resulting in difficulty in restoration of matrix \mathbf{D} from \mathbf{X} in (7) if there is no auxiliary information to be exploited. Our GLRR method combines similar patches into a group and the nonlocal similar patches can bring in extra spatial information to help reconstruct better the corrupted structure in the noisy patch. Fig.2(e) shows the denoised result by GLRR. We can see that GLRR outperforms BM4D and LRMR, removing all the noises and reconstructing the fine spatial structures simultaneously.

To further and quantitatively evaluate the effectiveness of our method, the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) [31] are adopted, with higher

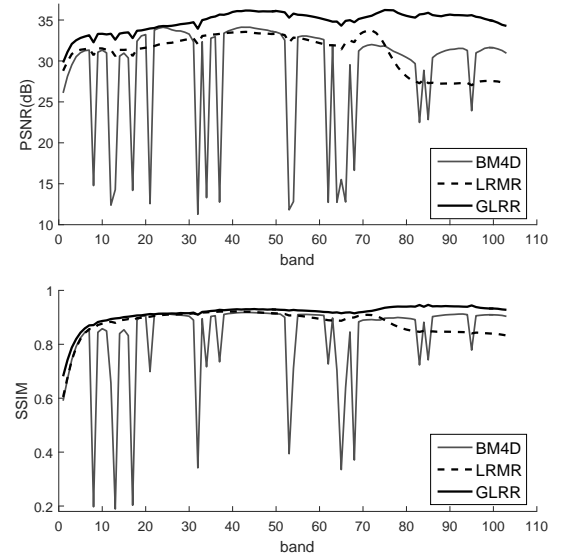


Fig. 3. Quantitative evaluation results of the three denoising methods (BM4D, LRMR and GLRR).

PSNR and SSIM indicating better denoising and reconstruction performance. Both PSNR and SSIM indices are calculated band-by-band between the reconstructed HSIs and the original clean data. Fig.3 shows the evaluation results, from which it can be observed that: our method achieves superior improvement to the other two methods; BM4D shows very poor performance on some bands, as it fails to remove salt-and-

pepper noise and stripe noise; and LRMR shows instability on bands 75-103. The average improvement of our method upon BM4D and LRMR is 5.81dB and 3.77dB in PSNR, and 0.0861 and 0.0361 in SSIM, respectively.

TABLE I
CLASSIFICATION RESULTS (%) FOR THE THREE DENOISING METHODS
(BM4D, LRMR AND GLRR) USING SVM

class	training / test sample size	BM4D	LRMR	GLRR
asphalt	50 / 6631	59.69	74.91	86.01
meadows	50 / 18649	67.60	87.85	90.92
gravel	50 / 2099	64.70	77.10	92.52
trees	50 / 3064	88.67	95.16	94.50
metal sheets	50 / 1345	99.52	99.72	98.30
bare soil	50 / 5029	63.38	89.92	93.38
bitumen	50 / 1330	69.15	91.89	98.01
bricks	50 / 3682	58.03	78.15	90.50
shadows	50 / 947	99.80	76.02	89.96
OA	-	68.19	85.48	91.77
AA	-	74.51	85.63	93.23
κ	-	59.90	81.15	89.34

Table I lists the classification results of the Pavia University data after denoising by BM4D, LRMR and GLRR, respectively. In our experiments the support vector machine (SVM) is used as the classification method and 50 samples of each class are randomly selected as the training samples. The classification experiments are repeated for ten times and the mean results of the overall accuracy (OA), average accuracy (AA), the kappa coefficient (κ) and the classification accuracies for each class are shown in Table I, in which the best result among the three methods are shown in bold. We can observe that the AA, OA and κ of our GLRR method achieve great improvement upon both BM4D and LRMR. This is also true for the classification accuracies of most of the classes. We note that the accuracies of LRMR and GLRR for the class of shadows are both worse than that of BM4D. This is mainly because that the light to the shadow area is from the light scattered by other materials rather than from the sunlight directly, which means that the spectral vector of shadows is not ideally consistent with the LMM in 2) and therefore leads to not as high accuracy in reconstruction of shadows as that of the other land-covers.

In summary, the experiments on the simulated data show that GLRR performs much better than BM4D and LRMR both qualitatively and quantitatively.

B. Experimental results on real-world data

Experiments on real-world data are conducted in a hyperspectral digital collection experiment (HYDICE) dataset,

Urban of Copperas Cove, Texas [32], denoted by “Urban” in the following. The spatial size of Urban is 307×307 ; the spatial resolution is $4m/\text{pixel}$. There are 210 bands in the original dataset; however, considering the pollution by atmosphere and water absorption, bands 104-108, 139-151 and 207-210 are removed from the data. Therefore a dataset with size of $307 \times 307 \times 188$ is actually used in our experiments. Several bands of the data is polluted by heavy Gaussian noise, as well as by stripe noise due to the detector-to-detector difference in pushbroom sensors. Fig.4(a) shows an example of the noisy bands.

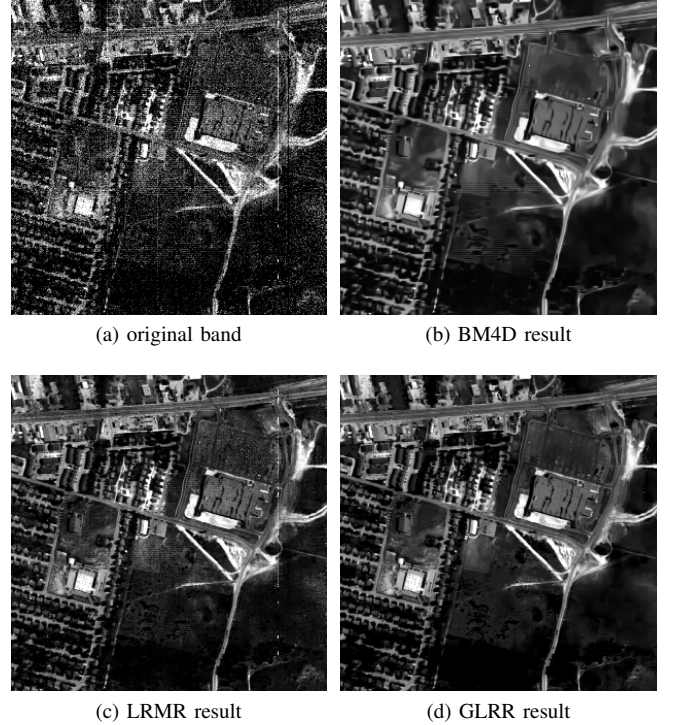


Fig. 4. Experimental results of band 188 in Urban.

Reconstruction results by BM4D, LRMR and GLRR are shown in Fig.4(b)-(d), respectively. It can be observed that, as with the simulated data in section III-A, BM4D oversmooths the denoised area and LRMR fails in the thorough removal of stripe noise and Gaussian noise. Fig.4(d) shows that our GLRR method outperforms the other two methods greatly, eliminating most of the noises and reconstructing the spatial details.

C. Analysis of the parameters

There are five parameters involved in GLRR: blocksize q , inverse parameter τ in (12), number of similar patches l , rank constraint r and sparsity constraint k_s in (9). To evaluate and discuss the impact and optimal values of these parameters, we conduct more experiments and use the average PSNR (APSNR) as the evaluation measure, which is defined as

$$\text{APSNR} = \frac{1}{B} \sum_{i=1}^B \text{PSNR}_i, \quad (13)$$

where PSNR_i is the PSNR of band i . The data used for the analysis and discussion in this section is the same as the simulated data used in Section III-A.

TABLE II
APSNR (IN DB) AT DIFFERENT INVERSE PARAMETERS τ IN (12)

τ	10	10^2	10^3	10^4
APSNR	35.040	35.035	35.030	35.030

TABLE III
APSNR (IN DB) AT DIFFERENT NUMBERS l OF SIMILAR PATCHES

l	0	1	2	3	4	5	6	7	8
APSNR	31.25	32.92	34.96	34.98	35.02	34.93	34.94	34.91	34.02

The best way to determine the optimal values for these five parameters is to search for a global optimum in a five-dimensional parameter space of the values of all the parameters simultaneously. However, this will inevitably require vast if not infeasible computation for HSIs. To avoid this, our strategy is to adopt a greedy method to decide the values of the parameters one by one and iteratively. Although this strategy may lead to a local optimum, it is more computationally efficient and applicable, and it can produce acceptable denoising performance, as shown in sections III-A and III-B.

1) l and τ : Table II and Table III describe the relationship between APSNR and the inverse parameter τ and the number of similar patches l , respectively, with the other parameters are fixed at the optimal levels. From Table II, it can be observed that APSNR is rather stable with different values of τ , with the standard deviation being only 0.005dB. From Table III, we can observe that APSNR achieves remarkable improvement when l is increased from 0 to 2. (Note that when $l = 0$, GLRR actually degenerates to LRMR.) This improvement can be attribute to the introduction of nonlocal similar patches into a group, which helps the reconstruction of patches with corrupted structures. We can also observe that APSNR encounters a decline when l is increased from 7 to 8. This is mainly because that the similarity within a group cannot be guaranteed high when l is big. However, it can be observed that APSNR is quite stable when l is between 2 and 7, with the standard deviation just 0.03dB. Therefore we can see that GLRR is robust with τ and a reasonable range of l . That is, suboptimal τ and l would not decrease the denoising result much. In this paper l and τ are fixed at $l = 4$ and $\tau = 10$ for all the experiments.

2) r : Fig.5 plots the experimental results of APSNR versus the low-rank constraint r , under different values of blocksize q and with other parameters fixed. Different curves show results under different blocksize q . It can be observed that the optimal low-rank constraint is often approximately at $r = 20$, even with different spatial sizes of the patches ($q \times q$). This result complies with the intuition that the low-rank constraint is related to the number of endmembers and thus has little to do with the spatial structure. In practice, we

can determine r as follows: if the number of endmembers in our target data is known, then r can be chosen approximately the number (typically slightly bigger with consideration of the precision of sensors in reality, the absorption and scattering by the atmosphere, and the quantization error); otherwise r is chosen between 0-20 as the endmember number within an HSI is typically less than 20, and then we can use subjective evaluation to determine the optimal value.

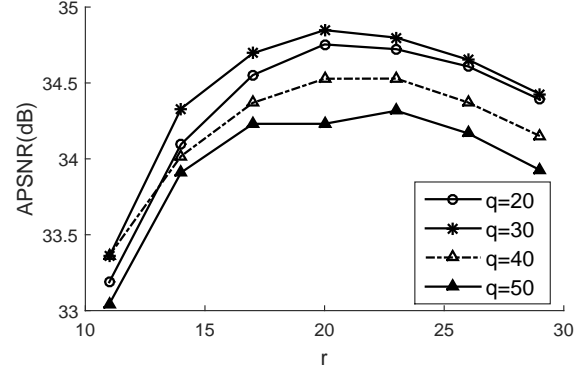


Fig. 5. The relationship between APSNR and the low-rank constraint r .

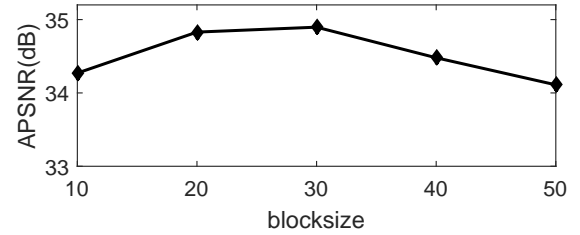


Fig. 6. The relationship between APSNR and the blocksize q .

3) q : The experimental results of APSNR versus the blocksize q are depicted in Fig.6, with the other parameters fixed. It can be observed that when $q = 30$ the proposed method achieves the best APSNR. The variation of APSNR is slight (between 34dB and 35dB), indicating that with a suboptimal choice of q the denoising result can still be reasonably good.

4) k_s : The regression of the optimal sparsity constraint k_s on the blocksize q is depicted in Fig.7, with other parameters fixed. In Fig.7, the horizontal axis is q^2 , the number of pixels in a patch. It can be observed that the optimal sparse constraint k_s is approximate linear to q^2 . That is to say, the optimal k_s is proportional to the number of pixels in a patch. This is in accordance with the principle that \mathbf{S} models the sparse noise in HSIs and k_s is the constraint on nonzero elements in \mathbf{S} , as in (9), so the amount of sparse noise should be approximately proportional to the number of pixels.

In short, from all the evaluation results above, we can suggest that: 1) the blocksize q can be set at a moderate level (see Fig.6); 2) the proposed method is insensitive to τ and l in a reasonable range (see Table II and Table III); 3) the optimal rank constraint r and sparsity constraint k_s can be estimated based on the characteristics of the HSIs to be denoised, while r can be chosen based on the information of the endmember

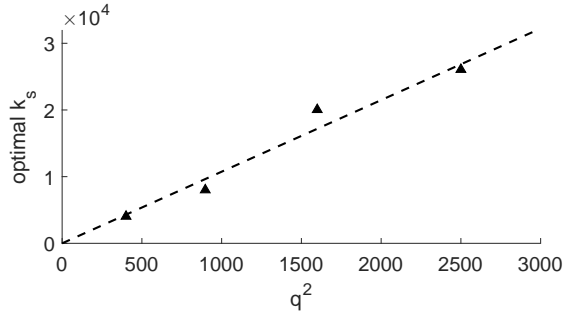


Fig. 7. The relationship between the optimal sparse constraint k_s and the blocksize q .

number and k_s can be determined based the noise amount in the HSIs.

D. Computational complexity

According to the analysis in [29], for a matrix with size $(l+1)q^2 \times B$, the computational cost of the BRP-based low-rank matrix recovery is $O(T(r^2(2B+r) + 4(l+1)q^2Br))$, where T (typically set as 50 in our method) is the iteration number within BRP. Considering the fact that $r \ll \min\{(l+1)q^2, B\}$, the complexity is reformed to $O(T(l+1)q^2Br)$. Therefore the computational complexity for the whole K patches is $O(KT(l+1)q^2Br)$.

IV. CONCLUSION

In this paper, we have proposed a group low-rank representation (GLRR) denoising method for the reconstruction of corrupted HSIs. The grouping strategy of GLRR takes the nonlocal similarity into consideration during the reconstruction. The nonlocal similar patches can introduce extra spatial structure information to help reconstruct the spatial structure in the corrupted patches. The use of LRR by GLRR enjoys the intrinsic low-rank structure of HSIs. Both subjective qualitative and objective quantitative evaluations, including PSNR, SSIM and classification accuracies, of the experiments have demonstrated that the proposed GLRR method can detect and remove the corrupting noise effectively, as well as retaining fine features of the HSIs.

An extensive discussion has also been conducted on the impact of the parameters involved in our method and its computational complexity, and the experimental results show that GLRR performs relatively stably within a reasonable range of the parameters.

ACKNOWLEDGEMENT

The authors are grateful to the associate editor and the reviewers for their constructive comments, which have led to a substantial improvement of our manuscript, in particular for Section III. This work was partly supported by the National Nature Science Foundation (No.61171117) and the Key Project of the Science & Technology Development Program of BEC (No.kz201310028035) of China. M. Wang is grateful to Tsinghua University for a Tsinghua Scholarship for

Overseas Graduate Studies and to the Department of Statistical Science at University College London (UCL) for supporting and hosting her visit at UCL.

REFERENCES

- [1] 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.
- [2] 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, Jan 2013.
- [3] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr 2008.
- [4] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm. and Remote Sens.*, vol. 83, no. 0, pp. 50–63, 2013.
- [5] H. Zhang, "Hyperspectral image denoising with cubic total variation Model," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, pp. 95–98, Jul. 2012.
- [6] 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.
- [7] 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.
- [8] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27–35, Jan 2009.
- [9] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun 2010.
- [10] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr 2013.
- [11] T. Lu, S. Li, L. Fang, Y. Ma, and J. Benediktsson, "Spectral-spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan 2016.
- [12] 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.
- [13] G. Chen and S. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar 2011.
- [14] T. Lin and S. Bourennane, "Hyperspectral image processing by jointly filtering wavelet component tensor," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3529–3541, Jun 2013.
- [15] D. Letexier, S. Bourennane, and J. Blanc-Talon, "Nonorthogonal tensor matricization for hyperspectral image filtering," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 3–7, Jan 2008.
- [16] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan 2013.
- [17] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. 20th ICML*, vol. 3, 2003, pp. 720–727.
- [18] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, 2013.
- [19] Z. Liu, G. and Lin and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th ICML*, 2010, pp. 663–670.
- [20] X. Zhou, C. Yang, H. Zhao, and W. Yu, "Low-rank modeling and its applications in image analysis," *ACM Comput. Surv.*, vol. 47, no. 2, pp. 36:1–36:33, 2015.
- [21] H. Huang, A. Christodoulou, and W. Sun, "Super-resolution hyperspectral imaging with unknown blurring by low-rank and group-sparse modeling," in *Proc. ICIP*, 2014, pp. 2155–2159.
- [22] 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, July 2013.
- [23] 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.

- [24] A. Sumarsono and Q. Du, "Low-rank subspace representation for estimating the number of signal subspaces in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6286–6292, Nov 2015.
- [25] R. Zhu, M. Dong, and J.-H. Xue, "Spectral nonlocal restoration of hyperspectral images with low-rank property," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 3062–3067, Jun 2015.
- [26] 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.
- [27] M. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun 2011.
- [28] T. Zhou and D. Tao, "Bilateral random projections," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, Jul 2012, pp. 1286–1290.
- [29] —, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc ICML*, 2011, pp. 33–40.
- [30] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.
- [31] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr 2004.
- [32] "Us army topographic engineering center, hypercube, urban," <http://www.tec.army.mil/Hypercube/>.



Weidong Sun (M'14) received the Ph.D. degree in information engineering from the University of Tokyo, Tokyo, Japan, in 1988. He is currently a Professor with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include image processing, pattern recognition, and spatial information processing and application.



Mengdi Wang received the B.Eng. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2012. She is currently pursuing the Ph.D. degree in the Department of Electronic Engineering, Tsinghua University, Beijing, China. Her research interests include hyperspectral image denoising and inpainting.



Jing Yu (M'15) received the Ph.D. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China in 2011. She is now a Lecturer in the College of Computer Science and Technology, Beijing University of Technology, Beijing, China. Her current research interests include image processing and pattern recognition.



Jing-Hao Xue received the Dr.Eng. degree in signal and information processing from Tsinghua University in 1998 and the Ph.D. degree in statistics from the University of Glasgow in 2008. Since 2008 he has worked in the Department of Statistical Science at University College London as a Lecturer and Senior Lecturer. His current research interests include statistical classification, high-dimensional data analysis, computer vision and pattern recognition.