Local Low-Rank and Sparse Representation for Hyperspectral Image Denoising

GUANQUN MA[ID][1], TING-ZHU HUANG[ID][1], JIE HUANG[ID][1], AND CHAO-CHAO ZHENG[2]

[1]Research Center for Image and Vision Computing, School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China

[2]Department of Mathematics, Hong Kong Baptist University, Hong Kong

Corresponding authors: Ting-Zhu Huang (tingzhuhuang@126.com) and Jie Huang (happyjie07mo@163.com)

ABSTRACT Hyperspectral image (HSI) denoising is a fundamental task in a plethora of HSI applications. Global low-rank property is widely adopted to exploit the spectral-spatial information of HSIs, providing satisfactory denoising results. In this paper, instead of adopting the global low-rank property, we propose to adopt a local low rankness for HSI denoising. We develop an HSI denoising method via local low-rank and sparse representation, under an alternative minimization framework. In addition, the weighted nuclear norm is used to enhance the sparsity on singular values. The experiments on widely used hyperspectral datasets demonstrate that the proposed method outperforms several state-of-the-art methods visually and quantitatively.

INDEX TERMS Hyperspectral image denoising, local low rankness, sparse representation, weighted nuclear norm.

I. INTRODUCTION

Hyperspectral images (HSIs) contain abundant spectral information of scenes, as they involve plenty of spectral bands. They have extensive applications in remote sensing, e.g, military, agriculture, medicine, and astrophysics [1]–[3]. From another perspective, they play an important role in target detection [4], hyperspectral unmixing [5]–[7], and image classification [8]–[10]. In a real scene, due to the limitation factors such as dark current, fluctuations in the power supply, and transmission errors, HSIs are unavoidably contaminated by noise during the acquisition process [11]. The contamination hinders the effectiveness of subsequent processing applications and analysis. Hence, HSI denoising is an essential issue and the denoising performance affects the accuracy of subsequent process.

In the past few years, many HSI denoising methods have been proposed [12]–[27]. One intuitive way is to utilize traditional 2-D denoising methods, such as K-SVD [28] and BM3D [29], to denoise HSIs band by band separately. Notice that an HSI contains lots of spectral bands and each band of the HSI can be presented as a 2-D gray image. However, the correlations between different spectral bands are

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

neglected in these methods, which causes spectral distortion [13].

To apply spectral and spatial correlations simultaneously, a serious of methods have been proposed. BM4D, an extension of BM3D, is based on nonlocal similarity and Wiener filtering [12]. Instead of utilizing a total variation (TV) term band by band, a spectral-spatial adaptive TV term is applied in [14] and [15]. Also, the TV regularization is utilized with the group sparsity constraint in [23] to effectively remove stripe noise in HSIs. Despite the good performance in HSI denoising, the TV regularization negatively affects the pixels which are not impulse corrupted [30]. To address this drawback of the TV regularization, a two-phase method is proposed in [24]. The first phase is identifying the impulse noise positions and the second phase is denoising by the patch-based model using noise position information.

Besides the TV term, the low-rank property of noise-free HSIs has been recently used in HSI denoising [20]. As a way to utilize the low-rank property, principal component analysis (PCA) is firstly applied to HSI denoising in [16]. In addition, an HSI denoising method based on superpixel-based subspace low-rank representation is proposed in [25]. In [21], a spectral difference-induced TV and low-rank approximation (SDTVLA) method is proposed for HSI denoising. Recently, the low-rank representation is utilized on spectral

difference space (LRRSDS) for HSI denoising [19]. Also, the low-rank matrix factorization is integrated with the TV regularization (LRTV) to further improve the denoising accuracy [17]. In LRTV, the low-rank model is applied to capture the spectral correlations and the TV regularization is used to capture the spatial piecewise smooth structure of HSIs. It is also worth noting that in the past few years, sparse representation has been successfully used in image denoising [22], [26]. One reason is that the statistical structure of noise-free HSIs allows for their efficient representation as a sparse linear combination of elementary features, but noisy HSIs lose the statistical structure [31]. In [18], a sparse representation and low-rank constraint (SRLR) method is proposed to remove HSI noise. It utilizes the local and the global redundancy and correlation of HSIs and effectively removes noise. We note that LRRSDS, LRTV, and SRLR all assume that noise-free HSIs lie in a low-rank subspace. Latterly, deep learning has been applied to HSI denoising. In [27], the authors propose a method by learning a nonlinear end-to-end mapping between noisy and noise-free HSIs with a combined spatial-spectral deep convolutional neural network (HSID-CNN). It simultaneously assigns both the spatial and the spectral information and gets good denoising performance.

Clearly, many of the above-mentioned methods utilize the low-rank property in a global way [17]–[21]. However, in many local patches of a noise-free HSI, neighboring pixels have high correlations in both spectral and spatial domains [32], [33]. Therefore, it is reasonable to assume that local patches of noise-free HSIs are low-rank [34], [35]. Thus, instead of adopting the global low-rank property, we propose to use the local low-rank representation to utilize the local spectral-spatial information and obtain a local low-rank and sparse representation model for HSI denoising. This model simultaneously captures the statistical structure and local spectral-spatial information of HSIs. Particularly, we use the statistical structure of HSIs in the framework of sparse representation. And the local spectral-spatial information is utilized by local low-rank representation. In addition, we use the weighted nuclear norm to enhance the low-rank property [36]. It is shown in [37] that compared to the standard nuclear norm, the weighted nuclear norm preserves major data components better during the denoising process. Thus, details of the HSI can be preserved better. We use the alternative minimization algorithm to solve the proposed model and obtain the local low-rank and sparse representation (LLRSR) method.

The main idea and contributions of this paper are summarized as follows:

1) A local low-rank and sparse representation model is proposed for HSI denoising. Instead of the global low-rank property of HSIs, the local low-rank property is adopted to utilize the local spectral-spatial information.

2) We adopt the weighted nuclear norm to enhance the sparsity on singular values.

3) The alternative minimization framework is designed to solve the proposed model, and we obtain the LLRSR

method. We numerically show the convergence stability of LLRSR. Also, our experiments demonstrate that LLRSR outperforms several popular methods for HSI denoising.

The rest of the paper is organized as follows. Section II introduces the sparse representation and the weighted nuclear norm as preliminary work. In section III, we describe the LLRSR method which is based on alternative minimization framework. Experiments in Section IV demonstrate the effectiveness of LLRSR, compared with other HSI denoising methods. Finally, section V gives concluding remarks.

II. PRELIMINARY WORK
A. SPARSE REPRESENTATION
Suppose that a noisy image $Y \in \mathbb{R}^{m \times a}$ holds

$$Y = X + N, \qquad (1)$$

where $X \in \mathbb{R}^{m \times a}$ is a noise-free image, $N \in \mathbb{R}^{m \times a}$ is a zero-mean Gaussian noise with a standard variance σ, and m and a are the numbers of rows and columns of the image, respectively. We note that in Section III, Y is a 2-D image reshaped from the original 3-D version HSI, similarly as in [18]. For an over-complete dictionary matrix D, its columns are regarded as prototype signal atoms. So a signal X can be represented as a sparse linear combination of the atoms, i.e., $X = D\alpha$, where most coefficients in α are close to zero [38]. Thus, the representation of X over D can be estimated from Y by:

$$\min_{D,\alpha} \|Y - D\alpha\|_F^2 + \eta \|\alpha\|_0, \qquad (2)$$

where η is a positive parameter and $\|\alpha\|_0$ is the number of non-zero elements in α. Once D and α are obtained, X can be estimated by $X = D\alpha$. In (2), getting D is the process of dictionary learning and getting α is called sparse coding. We remark that the K-means singular value decomposition (K-SVD) is a widely used algorithm to solve (2) [28].

B. WEIGHTED NUCLEAR NORM
The weighted nuclear norm of X is defined as:

$$\|X\|_{\phi,*} = \sum_{g=1}^{n} \phi_g \sigma_g(X), \qquad (3)$$

where n is the number of the singular values, $\phi = [\phi_1, \phi_2, \cdots, \phi_n] \geq 0$ is the weight vector, $\sigma_g(X)$ is the gth singular value of X, for $g = 1, \cdots, n$, and ϕ_g is the weight to the singular value $\sigma_g(X)$ [36]. Clearly, the weighted nuclear norm improves the flexibility of the standard nuclear norm with $\phi_g = 1$ for all g [37]. In order to utilize the weighted nuclear norm in HSI denoising, we generally solve the following problem:

$$\min_{X} \lambda \|X\|_{\phi,*} + \frac{1}{2} \|Y - X\|_F^2, \qquad (4)$$

where $\lambda > 0$ is a regularization parameter.

It is well-known that (4) has a closed-form solution when the elements of the weighted vector is set to be nonnegative and non-descending [35]. In order to express this closed-form solution, we first give two definitions: SHR and SVT. First, the soft-thresholding operator on the matrix Σ is defined as

$$\text{SHR}_\Delta(\Sigma) = \text{sign}(\Sigma) \max(0, |\Sigma| - \Delta), \tag{5}$$

where Δ is the matrix that contains thresholding parameters [39]–[41]. The operator in (5) performs in an element-wise manner. It is worth noting that when the soft-thresholding operator is applied on a diagonal matrix, we only shrink the diagonal elements. Thus, these diagonal elements can be shrinked by thresholding parameters contained in a vector. Then, we define the singular value thresholding operation on matrix Y as

$$\text{SVT}_\delta(Y) = U \, \text{SHR}_\delta(\Sigma) \, V^T, \tag{6}$$

where $Y = U\Sigma V^T$ is the singular value decomposition (SVD) of Y and δ is a vector whose elements are the thresholding parameters shrinking the diagonal elements of Σ. Then, from [35], [42], the solution of (4) can be expressed as:

$$X = \text{SVT}_{\lambda\phi}(Y). \tag{7}$$

The weighted nuclear norm has been widely used in many applications, such as image denoising [37], compressive sensing [43], hyperspectral unmixing [35], and so on.

III. PROPOSED METHOD
A. THE PROPOSED MODEL
The classic sparse representation HSI denoising model is:

$$\min_{X,D,\{\alpha_i\}} \gamma \|X - Y\|_F^2 + \sum_{i=1}^{J} \left(\|R_i X - D\alpha_i\|_F^2 + \eta \|\alpha_i\|_0 \right), \tag{8}$$

where γ is a positive parameter, X is the denoised HSI and Y is the noisy HSI, R_i is the operator which extracts ith overlapping local patch $R_i X \in \mathbb{R}^{t \times t}$ from X, J is the number of patches, D is the trained dictionary, and α_i is the representation coefficient of $R_i X$ over D.

It is clear that the objective function in (8) has two terms: the data fidelity term and the summation term. The fidelity term restricts the proximity between the denoised HSI X and the noisy HSI Y, where the weight γ depends on the noise level [26]. It helps to remove visible artifacts on patch boundaries [18]. The summation term indicates that every $R_i X$ has a sparse representation with small error, in other words, every $R_i X$ can be represented linearly by a few atoms in D with a small error.

To improve the denoising performance, a low-rank regularizer is combined with the sparse representation for the HSI denoising denoising problem in [18]. Mathematically, the resulting denoising model is:

$$\min_{X,D,\{\alpha_i\}} \gamma \|X - Y\|_F^2 + \sum_{i=1}^{J} \left(\|R_i X - D\alpha_i\|_F^2 + \eta \|\alpha_i\|_0 \right)$$
$$+ \mu \|X\|_*, \tag{9}$$

where μ and η are regularization parameters. This model jointly utilizes the local and the global redundancy and correlation of HSIs and achieves a good noise removal performance.

It is assumed in [44] that the rank of a noise-free HSI is far smaller than its size and a noisy HSI is hard to keep this property. We note that the low-rank constraint on the global HSI utilizes spectral correlation [45]. Due to piecewise smoothness and spectral correlation of HSIs, we use the local low-rank property to utilize the spatial-spectral correlation in local patches.

To compare the local and the global low-rank properties of HSIs, we give a simple example on the Washington DC Mall dataset which is used in Section IV-A and many other studies [44], [46]. First of all, in order to value the low-rank property, we introduce the accumulation energy ratio of the first h singular values as

$$\text{AccEgyR} = \frac{\sum_{p=1}^{h} \sigma_p}{\sum_{p=1}^{q} \sigma_p}, \tag{10}$$

where σ_p is the pth singular value of a matrix and q is the number of the singular values. The AccEgyR shows the proportion of the top h singular values in all singular values [47], [48]. So when the AccEgyR is higher, the proportion is larger. Then, in order to show the low-rank property of the local patches, we randomly select 5000 different patches with size 8×8 from $\widetilde{Y} \in \mathbb{R}^{65536 \times 68}$ whose columns are vectorized bands of the Washington DC Mall dataset and calculate the mean singular values of these patches.
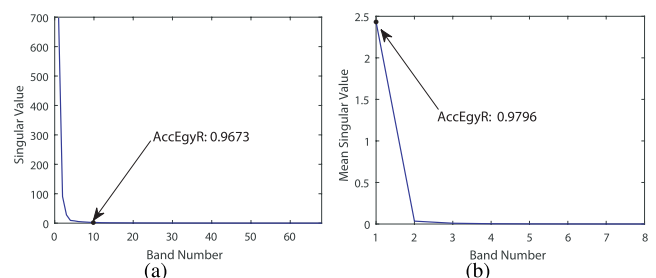
FIGURE 1. Comparision of the global low-rank property and the local low-rank property of the Washington DC Mall dataset. (a) The singular values of $\widetilde{Y}$. (b) The mean singular values of 5000 different patches from $\widetilde{Y}$.

Fig. 1(a) shows the singular value curve of the matrix $\widetilde{Y}$. We note that the AccEgyR attains to 0.9673 when $h = 10$. Fig. 1(b) shows the curve of the mean singular values of 5000 different local patches. And the AccEgyR attains to 0.9796 when $h = 1$. It says that the proportion of the first one singular value in all singular values of the local patches is larger than the proportion of the first ten singular values in all singular values of the matrix $\widetilde{Y}$. Therefore, we expect that the low-rank constraint on the local patches is more reasonable than on the global HSI and thus improves the denoising performance.

Based on the above-mentioned observations, we propose to utilize the low-rank constraint on the local patches to make better use of the spectral-spatial correlation in HSIs. In particular, we use a nuclear norm regularization on each local patch and the resulting denoising model is:

$$\min_{X,D,\{\alpha_i\}} \gamma \|X - Y\|_F^2$$
$$+ \sum_{i=1}^{J} \left(\|R_iX - D\alpha_i\|_F^2 + \eta\|\alpha_i\|_0 + \mu\|R_iX\|_* \right). \quad (11)$$

We note that the difference between (9) and (11) is the use of the nuclear norm. We also note that singular values contain important physical information [49]. Large singular values contribute to the main physical characteristics of the HSI and vice versa. To further enhance the low-rank property of the local patches, we use the weighted nuclear norm as a low-rank regularizer. Thus, the proposed HSI denoising model becomes:

$$\min_{X,D,\{\alpha_i\}} \gamma \|X - Y\|_F^2$$
$$+ \sum_{i=1}^{J} \left(\|R_iX - D\alpha_i\|_F^2 + \eta\|\alpha_i\|_0 + \mu\|R_iX\|_{w_i,*} \right), \quad (12)$$

where each $w_i \geq 0$ is a weight vector.

B. THE LLRSR METHOD

We now solve (12) under the alternative minimization framework. It is clear that there is more than one variable to be solved in (12), so variable splitting is utilized. Let $P_i = R_iX$, for $i = 1, \cdots, J$, and then we rewrite (12) to the following constrained problem:

$$\min_{X,\{P_i\},D,\{\alpha_i\}} \gamma \|X - Y\|_F^2$$
$$+ \sum_{i=1}^{J} \left(\|R_iX - D\alpha_i\|_F^2 + \eta\|\alpha_i\|_0 + \mu\|P_i\|_{w_i,*} \right),$$
$$\text{s.t.} \qquad P_i = R_iX(i = 1, 2, \cdots, J). \quad (13)$$

By utilizing a proper positive parameter ρ for the quadratic penalty term, we obtain the following unconstrained problem:

$$\min_{X,\{P_i\},D,\{\alpha_i\}} \gamma \|X - Y\|_F^2 + \sum_{i=1}^{J}(\|R_iX - D\alpha_i\|_F^2 + \eta\|\alpha_i\|_0$$
$$+ \mu\|P_i\|_{w_i,*} + \rho\|R_iX - P_i\|_F^2). \quad (14)$$

We now solve the unconstrained problem (14) by alternative minimization. In the process of alternative minimization, each variable in (14) is optimized while fixing others. So we obtain the following sub-problems:

1) *D*- and $\{\alpha_i\}$-subproblem:

$$\min_{D,\{\alpha_i\}} \sum_{i=1}^{J} \left(\|R_iX - D\alpha_i\|_F^2 + \eta\|\alpha_i\|_0 \right). \quad (15)$$

This step is dictionary learning and sparse coding, which can be solved by K-SVD [28].

2) P_i-subproblem, for $i = 1, \cdots, J$:

$$\min_{P_i} \frac{\mu}{2\rho}\|P_i\|_{w_i,*} + \frac{1}{2}\|P_i - R_iX\|_F^2,$$
$$i = 1, \cdots, J. \quad (16)$$

Let the weight coefficient vector w_i be

$$w_i = \left(\frac{1}{\sigma_1(R_iX) + \epsilon}, \cdots, \frac{1}{\sigma_t(R_iX) + \epsilon} \right), \quad (17)$$

similarly as in [35]. Here, $\sigma_1(R_iX), \cdots, \sigma_t(R_iX)$ are the non-ascending singular values of R_iX and $\epsilon = 10^{-16}$ is set in our experiments to avoid singularity. Therefore, the elements in w_i are nonnegative and non-descending. It follows from Section II-B that every P_i-subproblem in (16) has a closed-form solution:

$$P_i = \text{SVT}_{\frac{\mu}{2\rho}w_i}(R_iX), \quad i = 1, \cdots, J. \quad (18)$$

3) *X*-subproblem:

$$\min_{X} \gamma \|X - Y\|_F^2 + \sum_{i=1}^{J}\|R_iX - D\alpha_i\|_F^2$$
$$+ \rho\sum_{i=1}^{J}\|R_iX - P_i\|_F^2. \quad (19)$$

It is a least squares problem. Thus we have:

$$X = \left((1 + \rho)\sum_{i=1}^{J}R_i^T R_i + \gamma \right)^{-1}$$
$$\left(\sum_{i=1}^{J}R_i^T (D\alpha_i + \rho P_i) + \gamma Y \right). \quad (20)$$

We stop the optimization procedure when a pre-defined maximum number of iterations is reached.

We now summarize the proposed algorithm, named LLRSR, in Algorithm 1. It is worth noting that the convergence of Algorithm 1 is hard to obtain theoretically due to the non-convexity of the objective function in (14). However, we will numerically show the convergence of Algorithm 1 in Section IV-A.

C. COMPUTATIONAL COMPLEXITY

We now show the computational complexity of Algorithm 1. Consider the HSI $\mathcal{Y} \in \mathbb{R}^{M \times N \times b}$, where M is the number of rows, N is the number of columns, and b is the number of spectral bands. The computational complexity of Algorithm 1 mainly consists of three parts as follows.

1) Dictionary learning and sparse coding. Dictionary can be learned offline to accelerate the algorithm. So we only consider the computational complexity of sparse coding. If the step size of extracting patch is d and the patch number is $\lceil \frac{MN}{d} \rceil \cdot \lceil \frac{b}{d} \rceil$, then the dictionary size used in Algorithm 1 is t^2c. So the computational complexity of sparse coding is $\mathcal{O}(t^2c \cdot \lceil \frac{MN}{d} \rceil \cdot \lceil \frac{b}{d} \rceil)$.

2) Updating $\{P_i\}$. For a $t \times t$ patch, the computational cost of SVD is $\mathcal{O}(t^3)$. In addition, it costs $\mathcal{O}(2t)$ to calculate

Algorithm 1 Local Low-Rank and Sparse Representation (LLRSR) for HSI Denoising

Input: Noisy HSI Y
Output: Denoised HSI X
Initialize: $X^0 = Y$, $k = 1$, parameters γ, μ, ρ,
the maximum number of iterations K,
and the size of $R_i X$ $t \times t$
While: $k \le K$
  1. Dictionary learning and sparse coding.
  $\{D^k, \{\alpha_i^k\}\}$
  $= \arg\min_{D, \{\alpha_i\}} \sum_{i=1}^{J} (\|R_i X^{k-1} - D\alpha_i\|_2^2 + \eta\|\alpha_i\|_0),$
  t which is solved by the K-SVD method.
  2. Updating auxiliary variables P_i, $i = 1, \cdots, J$.
  $P_i^k = \text{SVT}_{\frac{\mu}{2\rho} w_i}\left(R_i X^{k-1}\right),$
  where $w_i = \left(\frac{1}{\sigma_1(R_i X^{k-1})+\epsilon}, \cdots, \frac{1}{\sigma_t(R_i X^{k-1})+\epsilon}\right)$.
  3. Updating denoised HSI X.
  $X^k = \left((1+\rho)\sum_{i=1}^{J} R_i^T R_i + \gamma\right)^{-1}$
  $\left(\sum_{i=1}^{J} R_i^T \left(D^k \alpha_i^k + \rho P_i^k\right) + \gamma Y\right).$
  4. $k = k + 1$.
EndWhile.

the weighting coefficient and solve the element-wise soft shrinkage. Thus, the overall computational complexity to update $\{P_i\}$ is $\mathcal{O}(\lceil\frac{MN}{d}\rceil \cdot \lceil\frac{b}{d}\rceil \cdot t^3)$.

3) Updating X. The computational complexity of this step is $\mathcal{O}(MNb)$.

In conclusion, the overall computational complexity of Algorithm 1 per iteration is $\mathcal{O}(\lceil\frac{MN}{d}\rceil \cdot \lceil\frac{b}{d}\rceil \cdot t^2(t + c))$.

IV. EXPERIMENT RESULTS AND DISCUSSION

To demonstrate the effectiveness of LLRSR, the denoising results are compared with the results of four well-known methods, including: 1) the low-rank representation method in the spectral difference space (LRRSDS) [19]; 2) the block-matching and four-dimensional filtering algorithm (BM4D) [12]; 3) the sparse representation and low-rank constraint method (SRLR) [18]; 4) the TV-regularized low-rank method (LRTV) [17].

Three quantitative picture quality indices (PQIs) are employed for evaluation, namely, peak signal-to-noise ratio (PSNR) [40], [41], structure similarity (SSIM) [50], and spectral angle distance (SAD) [51]. These PQIs are adopted as mean PSNR (MPSNR), mean SSIM (MSSIM), and mean SAD (MSAD) for HSI evaluation. The MSAD is defined as

$$\text{MSAD} = \frac{1}{b}\sum_{i=1}^{b} \frac{180}{\pi} \cdot \arccos\frac{(\zeta^s)^T \cdot (\hat{\zeta}^s)}{\|\zeta^s\| \cdot \|\hat{\zeta}^s\|}, \quad (21)$$

where ζ^s and $\hat{\zeta}^s$ are the sth spectral signatures of the noise-free and denoised HSIs, respectively, and b is the number of bands. We note that BM4D is a parameter-free method. For other four methods, we try different parameters to reach the maximum MPSNR value.

TABLE 1. Parameters of LLRSR under different noise levels.

Noise Level	γ	μ	ρ
$\sigma = 0.025$	0.04	0.0025	3
$\sigma = 0.05$	0.02	0.01	3
$\sigma = 0.075$	0.015	0.02	3
$\sigma = 0.1$	0.005	0.04	4

TABLE 2. PQIs of different denoising methods on the Pavia city center dataset under different noise levels.

Noise Level	PQIs	LRRSDS	BM4D	SRLR	LRTV	LLRSR
$\sigma = 0.025$	MPSNR	41.12	40.74	41.29	41.03	**42.40**
	MSSIM	0.9918	0.9913	0.9917	0.9908	**0.9940**
	MSAD	1.782	1.822	1.693	1.790	**1.488**
$\sigma = 0.05$	MPSNR	36.67	36.22	36.47	36.70	**37.60**
	MSSIM	0.9766	0.9766	0.9755	0.9725	**0.9824**
	MSAD	2.812	3.010	2.881	3.002	**2.514**
$\sigma = 0.075$	MPSNR	33.29	33.66	33.59	34.13	**34.58**
	MSSIM	0.9416	0.9587	0.9543	0.9485	**0.9657**
	MSAD	3.584	4.009	3.959	4.124	**3.532**
$\sigma = 0.1$	MPSNR	31.76	31.88	31.53	32.32	**32.41**
	MSSIM	0.9420	0.9381	0.9290	0.9236	**0.9442**
	MSAD	4.580	4.917	4.981	5.128	**4.553**

A. SIMULATED EXPERIMENTS ON DIFFERENT NOISE LEVELS

1) DATA DISCRIPTION

This subsection compares LLRSR with other four HSI denoising methods under different noise levels. Three commonly used datasets are tested. They are the Washington DC Mall dataset[1], the San Diego dataset [18], and the Pavia City Center dataset[1]. The size of the Washington DC Mall subdataset which is selected is $256 \times 256 \times 68$. We choose a subdataset of the San Diego dataset whose size is $200 \times 200 \times 68$. In the experiment of the Pavia City Center dataset, a subdataset whose size is $300 \times 300 \times 40$ is selected. Spectral bands which are heavily contaminated by noise are removed in these subdatasets. Four different levels of Gaussian noise are added to the datasets, i.e., $\sigma = 0.025, 0.05, 0.075,$ and 0.1. For each dataset, the gray values of each HSI band are normalized between [0, 1].

2) PARAMETERS SETTING

All the experiments were implemented in MATLAB 2016a with the platform of Intel(R) Core(TM) i5-3570 CPU @ 3.40GHz and 8 GB memory.

Notice that there are four parameters in (14). Under the same noise level, we use the same parameters for all three datasets. Table 1 lists the parameters $\gamma, \mu,$ and ρ that we use

---

[1]Available from http://lesun.weebly.com/hyperspectral-data-set.html

TABLE 3. PQIs of different denoising methods on the san diego dataset under different noise levels.

Noise Level	PQIs	LRRSDS	BM4D	SRLR	LRTV	LLRSR
	MPSNR	41.27	41.35	_41.97_	41.71	**42.37**
$\sigma = 0.025$	MSSIM	0.9876	0.9889	_0.9900_	0.9894	**0.9909**
	MSAD	1.380	1.350	_1.254_	1.318	**1.225**
	MPSNR	37.18	36.86	37.58	_37.67_	**37.88**
$\sigma = 0.05$	MSSIM	0.9662	0.9713	_0.9737_	0.9705	**0.9757**
	MSAD	2.215	2.257	_2.076_	2.161	**2.024**
	MPSNR	34.62	34.27	34.87	**35.16**	_35.01_
$\sigma = 0.075$	MSSIM	0.9408	0.9509	_0.9543_	0.9481	**0.9562**
	MSAD	2.988	3.038	_2.825_	3.001	**2.785**
	MPSNR	32.64	32.49	32.75	**33.54**	_32.89_
$\sigma = 0.1$	MSSIM	0.9126	0.9297	_0.9311_	0.9306	**0.9331**
	MSAD	3.763	3.726	_3.582_	3.620	**3.532**

TABLE 4. PQIs of different denoising methods on the Washington DC mall dataset under different noise levels.

Noise Level	PQIs	LRRSDS	BM4D	SRLR	LRTV	LLRSR
	MPSNR	39.87	39.74	_40.53_	40.34	**41.06**
$\sigma = 0.025$	MSSIM	0.9909	0.9899	0.9916	_0.9918_	**0.9923**
	MSAD	1.850	1.902	_1.710_	1.785	**1.619**
	MPSNR	35.90	35.33	36.00	_36.52_	**36.62**
$\sigma = 0.05$	MSSIM	0.9759	0.9735	_0.9774_	0.9766	**0.9802**
	MSAD	2.875	3.113	2.845	_2.821_	**2.629**
	MPSNR	33.27	32.83	33.34	**34.17**	_33.75_
$\sigma = 0.075$	MSSIM	0.9562	0.9545	0.9601	_0.9607_	**0.9635**
	MSAD	3.913	4.123	3.827	_3.669_	**3.618**
	MPSNR	31.29	31.11	31.47	**32.52**	_31.60_
$\sigma = 0.1$	MSSIM	0.9331	0.9341	0.9410	**0.9435**	_0.9422_
	MSAD	4.956	5.006	4.716	**4.448**	_4.635_

for LLRSR under different noise levels for all three datasets. It is worth noting that the second parameter η in (14) is set implicitly.

3) COMPARISON WITH OTHER METHODS

Tables 2, 3, and 4 present the PQIs of different methods on the the Pavia City Center dataset, the San Diego dataset, and the Washington DC Mall dataset, respectively. The best and the second best results of each PQI are highlighted as bold and underlined, respectively.

From Table 2, we see that the PQIs of LLRSR are higher than those of LRRSDS, LRTV, BM4D, and SRLR. In Tables 3 and 4, most of the LLRSR's PQIs are the best,

although LRTV has better PQIs in a few cases. In Table 4, the PQIs of LRTV are better than those of LLRSR when $\sigma = 0.1$. It can be expected since the TV term imposes spatial consistency and has a good performance when the HSIs admit a piecewise smooth character, especially under heavy noise [17]. To sum up, LLRSR obtains comparable or better results compared with other four methods. This indicates the advantage of the proposed method for HSI denoising.

Fig. 2 shows the denoising results of band 38 of the San Diego dataset under $\sigma = 0.025$. We also show the corresponding PSNR values. The magnified red blocks are displayed in Fig. 3. From Fig. 3, we can observe that all five methods remove noise effectively. Clearly, BM4D and
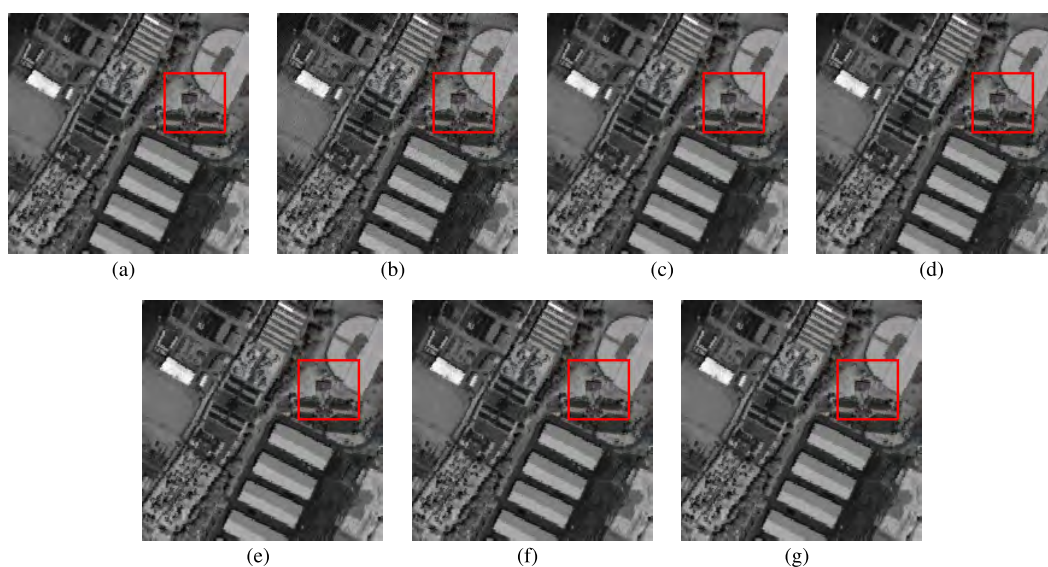
FIGURE 2. Band 38 of the San Diego dataset before and after denoising via five different methods under $\sigma = 0.025$. (a) Original noise-free image. (b) Noise image under $\sigma = 0.025$. (c) Result of LRTV (PSNR = 41.48 dB). (d) Result of LRRSDS (PSNR = 42.25 dB). (e) Result of BM4D (PSNR = 41.77 dB). (f) Result of SRLR (PSNR = 42.53 dB). (g) Result of LLRSR (PSNR = 43.39 dB).
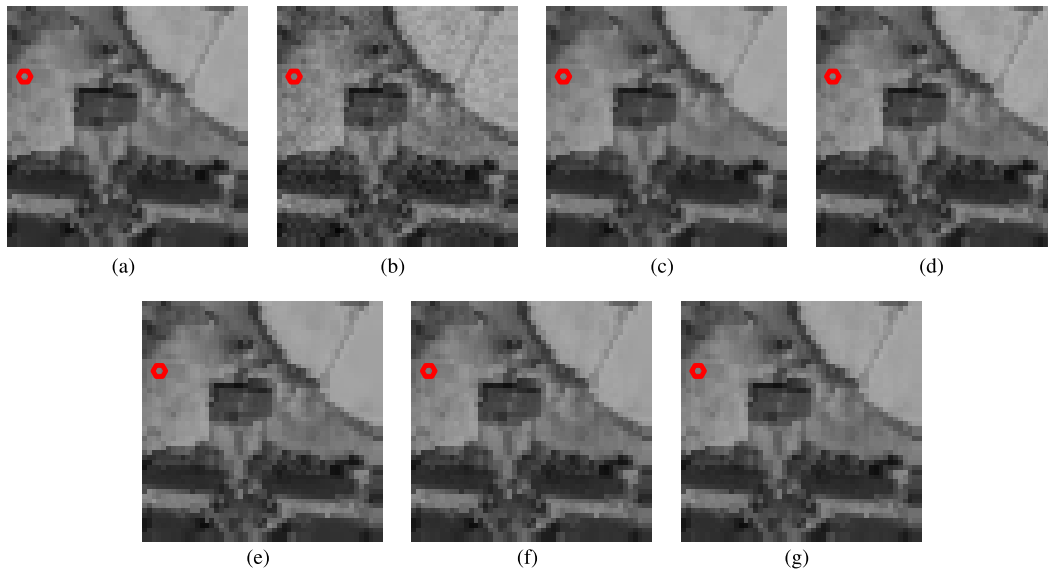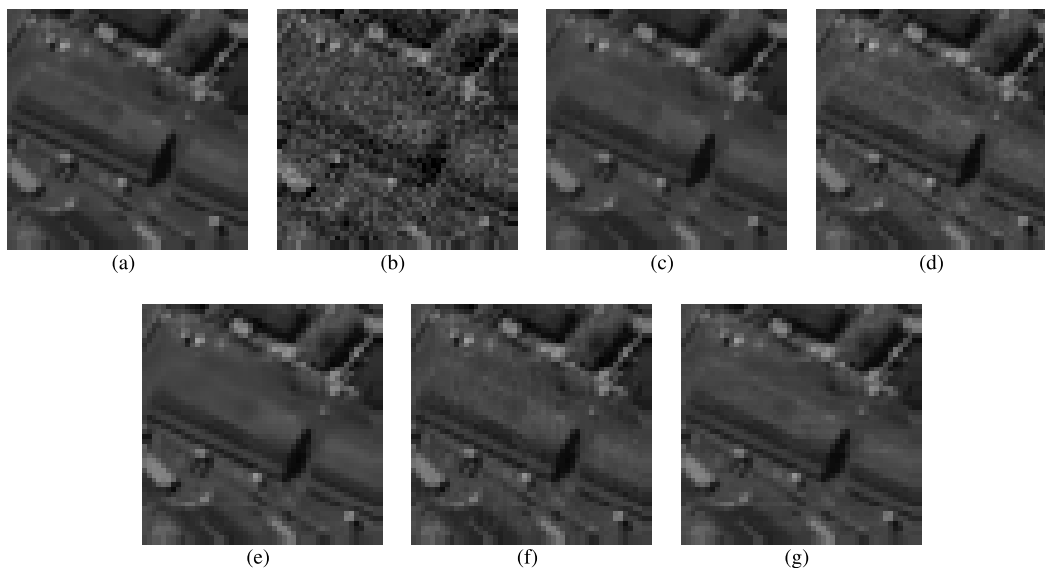
FIGURE 3. Details of the red blocks in Fig. 2. (a) Original noise-free image. (b) Noise image under $\sigma = 0.025$. (c) Result of LRTV. (d) Result of LRRSDS. (e) Result of BM4D. (f) Result of SRLR. (g) Result of LLRSR.
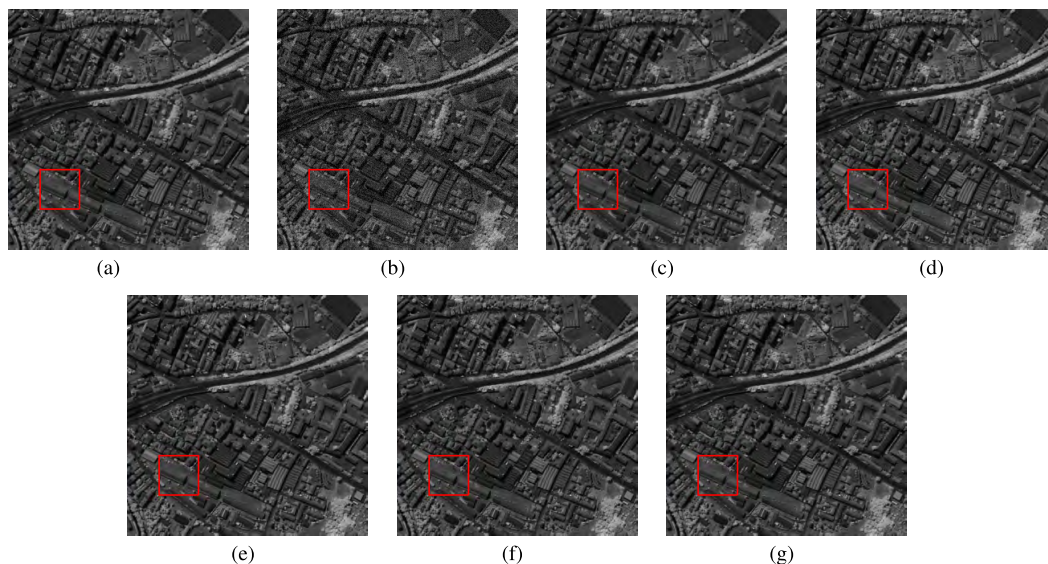
FIGURE 4. Band 22 of the Pavia city center dataset before and after denoising via five different methods under $\sigma = 0.05$. (a) Original noise-free image. (b) Noise image under $\sigma = 0.05$. (c) Result of LRTV (PSNR = 37.52 dB). (d) Result of LRRSDS (PSNR = 34.66 dB). (e) Result of BM4D (PSNR = 35.83 dB). (f) Result of SRLR (PSNR = 37.18 dB). (g) Result of LLRSR (PSNR = 38.15 dB).

LRTV over-smooth the image. Though the denoised image of LRRSDS is better than that of BM4D in Fig. 3, it still remains more noise than that of LLRSR. The results of SRLR and LLRSR are close, but SRLR loses more details than LLRSR in the denoised result. In conclusion, LLRSR achieves the best result among the five methods.

Fig. 4 displays the denoising results of band 22 of the Pavia City Center dataset under $\sigma = 0.05$ and the corresponding PSNR values. The magnified red blocks are displayed in Fig. 5. It is shown that the results of BM4D and LRTV are over-smoothed in Fig. 5 and they lose some texture details. And the results of SRLR and LRRSDS are noisy compared to the result of LLRSR.

Fig. 6 presents the denoising results of band 5 of the Washington DC Mall dataset under $\sigma = 0.075$. The magnified red blocks are displayed in Fig. 7. The result of BM4D is blurred. For the original image Fig. 7 (a), LLRSR removes noise and restores details more accurately than LRTV, LRRSDS, and SRLR, as presented in Figs. 7 (g), (c), (d), and (f), respectively.

Fig. 8 shows the denoising results of band 4 of the San Diego dataset under $\sigma = 0.1$. We also list the corresponding PSNR values. The magnified red blocks are displayed in Fig. 9. We can observe from Fig. 9 that LRRSDS and SRLR still remain some noise. Clearly, BM4D over-smooths the image with the loss of some texture details. Compared with

FIGURE 5. Details of the red blocks in Fig. 4. (a) Original noise-free image. (b) Noise image under $\sigma = 0.05$. (c) Result of LRTV. (d) Result of LRRSDS. (e) Result of BM4D. (f) Result of SRLR. (g) Result of LLRSR.
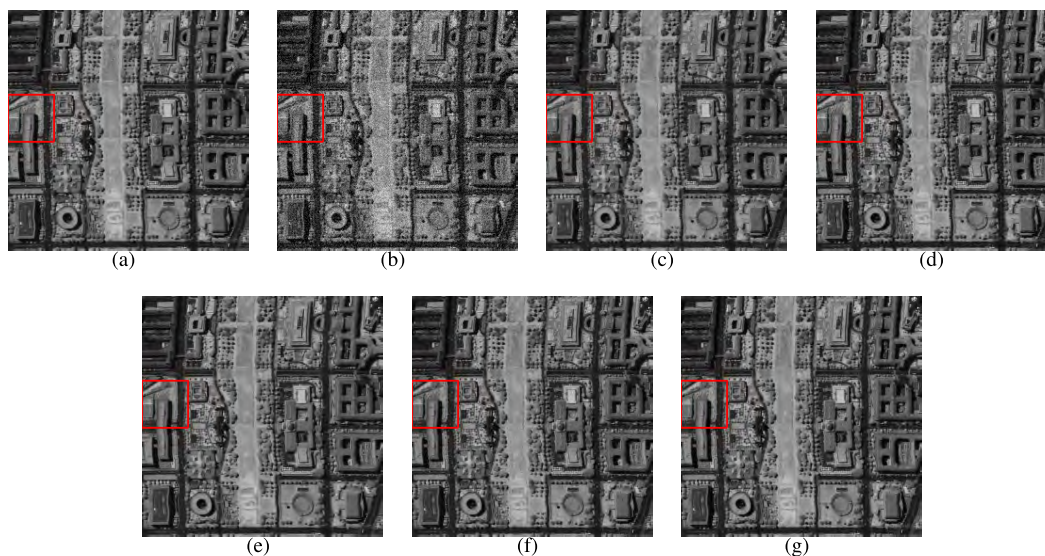
FIGURE 6. Band 5 of the Washington DC Mall dataset before and after denoising via five different methods under $\sigma = 0.075$. (a) Original noise-free image. (b) Noise image under $\sigma = 0.075$. (c) Result of LRTV (PSNR = 32.44 dB). (d) Result of LRRSDS (PSNR = 32.80 dB). (e) Result of BM4D (PSNR = 33.41 dB). (f) Result of SRLR (PSNR = 33.05 dB). (g) Result of LLRSR (PSNR = 34.00 dB).

LRTV, LLRSR better restores the image while suppressing noise.

In order to show the preservation of spectral information in denoising results, we display the spectral curves and the spectral difference curves of original full-band pixels taken from the San Diego Dataset and its estimations by different methods in Fig. 10, similarly as in [52]. We note that the pixels of the band 38 are spatially located at the center of the red circles displayed in Fig. 3. Fig. 10 shows that LLRSR is an effective denoising method preserving the spectral information.

Fig. 11 shows the PSNR, SSIM, and SAD values of each band of the denoising results of the San Diego dataset under

noise level $\sigma = 0.025$, the Pavia City Center dataset under noise level $\sigma = 0.05$, and the Washington DC Mall dataset under noise levels $\sigma = 0.075$ and 0.1, respectively. Clearly, LLRSR achieves the best PSNR, SSIM, and SAD values in most bands in Figs. 11 (a)-(f). In Figs. 11 (g)-(i), LLRSR has better PSNR, SSIM, and SAD values than BM4D, SRLR, and LRRSDS in most bands and has comparable PSNR, SSIM, and SAD values to LRTV. As for Figs. 11 (j)-(l), PSNR, SSIM, and SAD values of LLRSR are better than those of BM4D in most bands and comparable to those of SRLR and LRRSDS. LRTV has better PSNR, SSIM, and SAD values than LLRSR in most bands because of the piecewise smooth
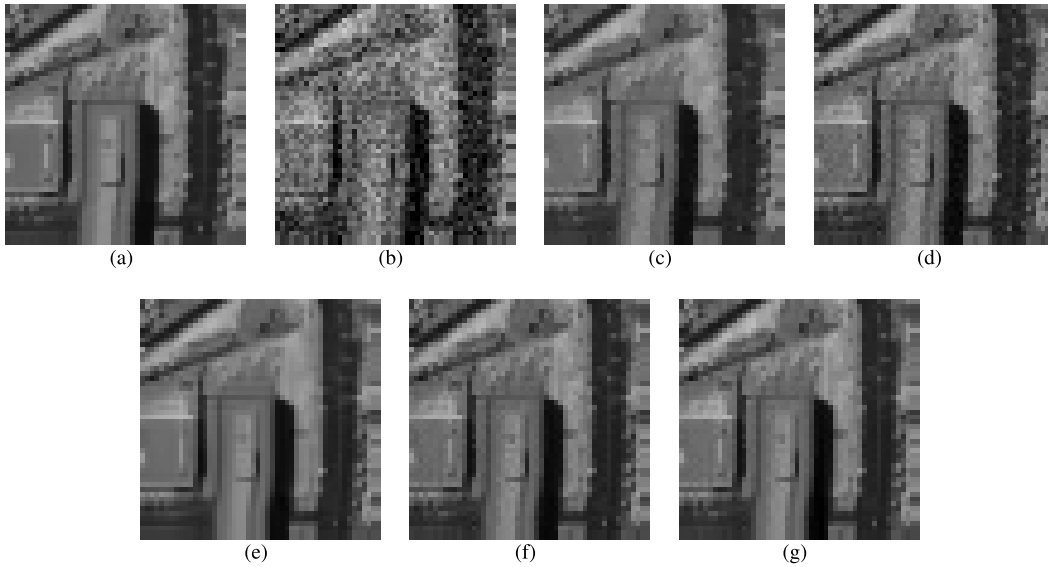
FIGURE 7. Details of the red blocks in Fig. 6. (a) Original noise-free image. (b) Noise image under $\sigma = 0.075$. (c) Result of LRTV. (d) Result of LRRSDS. (e) Result of BM4D. (f) Result of SRLR. (g) Result of LLRSR.
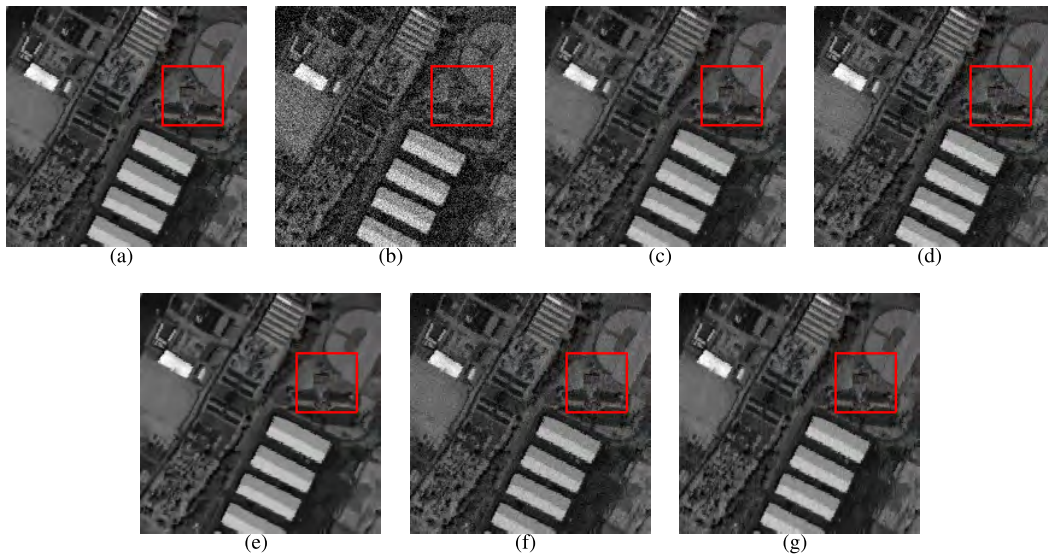
FIGURE 8. Band 4 of the San Diego dataset before and after denoising via five different methods under $\sigma = 0.1$. (a) Original noise-free image. (b) Noise image under $\sigma = 0.1$. (c) Result of LRTV (PSNR = 31.00 dB). (d) Result of LRRSDS (PSNR = 30.12 dB). (e) Result of BM4D (PSNR = 32.00 dB). (f) Result of SRLR (PSNR = 29.98 dB). (g) Result of LLRSR (PSNR = 32.66 dB).

character of the HSI. In addition, the PQIs of LLRSR are more stable than those of SRLR, LRRSDS, and LRTV.

Fig. 12 shows the relative change (RelCha) [53] in each iteration loop of the Washington DC Mall dataset under $\sigma = 0.025$ and the San Diego dataset under $\sigma = 0.075$. Here, the RelCha is defined as

$$\text{RelCha} = \frac{\|X^{k+1} - X^k\|_F}{\|X^k\|_F}. \tag{22}$$

It is shown in Fig. 12 that the convergence history of LLRSR in the experiments is stable.

B. EXPERIMENT WITH DIFFERENT NOISE LEVELS IN EACH BAND

In this section. we do an experiment on the Washington DC Mall dataset to show the effectiveness of LLRSR when the noise level varies in different bands. Fig. 13 shows the noise level of the noise we have added on each band of the Washington DC Mall dataset.

Recall that LLRSR assumes that each band has the same noise level. Therefore, in order to apply LLRSR to the above dataset, we first do a preprocessing step such that the noise levels of each band are equal. The preprocessing step
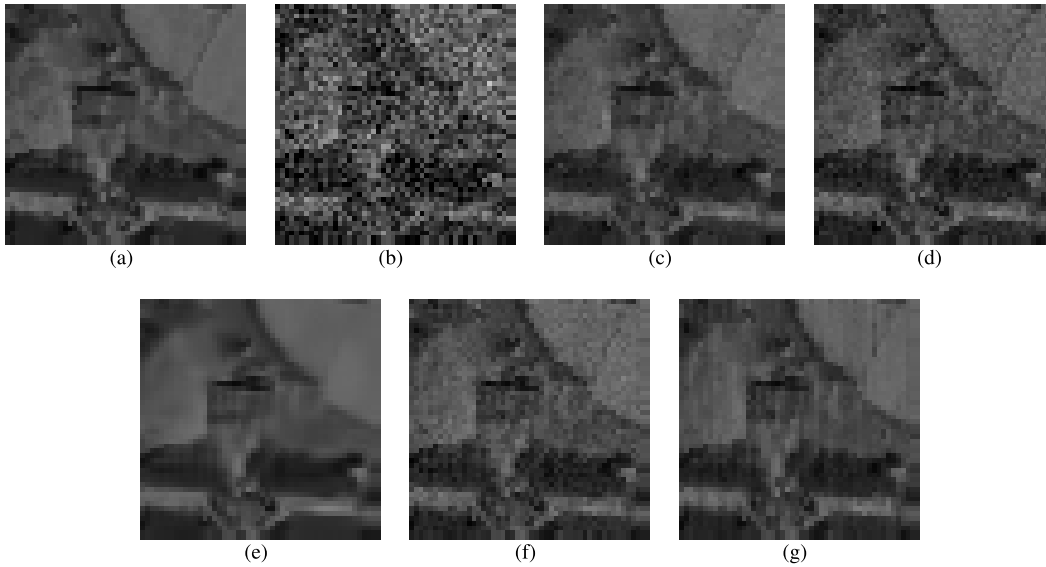
FIGURE 9. Details of the red blocks in Fig. 8. (a) Original noise-free image. (b) Noise image under $\sigma = 0.1$. (c) Result of LRTV. (d) Result of LRRSDS. (e) Result of BM4D. (f) Result of SRLR. (g) Result of LLRSR.
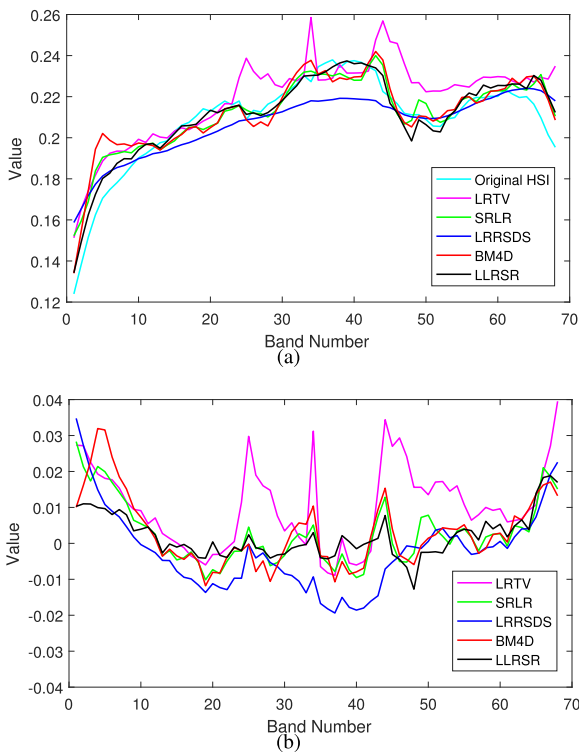
FIGURE 10. (a) The spectral curves and (b) the spectral difference curves of the choosen pixels in San Diego dataset via five different methods under $\sigma = 0.025$.

proposed in [18] is:

$$\tilde{Y}_{:,:,l} = \frac{Y_{:,:,l}}{\sigma_l} \times \sigma, \tag{23}$$

where $Y_{:,:,l}$ is the intensity of the lth band, σ_l is the noise level of the lth band, and σ is a given constant that the noise levels in each band change to after the preprocessing step. We set

TABLE 5. PQIs of different denoising methods on the Washington DC mall dataset under the noise which varies in different bands.

PQIs	LRRSDS	BM4D	SRLR	LRTV	LLRSR	LLRSR$_\text{p}$
MPSNR	39.86	38.52	39.55	40.38	40.04	**40.89**
MSSIM	0.9908	0.9874	0.9899	0.9917	0.9906	**0.9919**
MSAD	1.855	2.201	1.914	1.782	1.788	**1.657**

$\sigma = 0.025$ in this experiment. After the denoising process, an inverse step is used to get the final recovered result:

$$X_{:,:,l} = \frac{\tilde{X}_{:,:,l}}{\sigma} \times \sigma_l, \tag{24}$$

where $\tilde{X}_{:,:,l}$ is the intensity of the lth band after denoising and $X_{:,:,l}$ is the final recovered result of the lth band. Here, we use LLRSR$_\text{p}$ to represent LLRSR with the preprocessing step.

Table 5 shows the PQIs of different denoising methods on this dataset. We can see from Table 5 that LLRSR$_\text{p}$ gets the best PQIs. It shows that LLRSR$_\text{p}$ achieves an effective denoising result and the preprocessing step works for LLRSR to get a better result.

C. EFFECTIVENESS OF THE SPARSE REPRESENTATION PRIOR AND THE LOCAL LOW-RANK PRIOR

In this experiment, we show the effectiveness of simultaneously utilizing the sparse representation prior and the local low-rank prior in LLRSR on three datasets. For all three datasets, on the one hand, we denoise HSIs only by sparse representation and ignore the local low-rank term. That is to set $\mu = 0$ in (12). And we use SR to represent the resulting method. On the other hand, we denoise HSIs only using the local low-rank constraint and ignore the sparse representation prior. The resulting method is denoted as LLR for short.
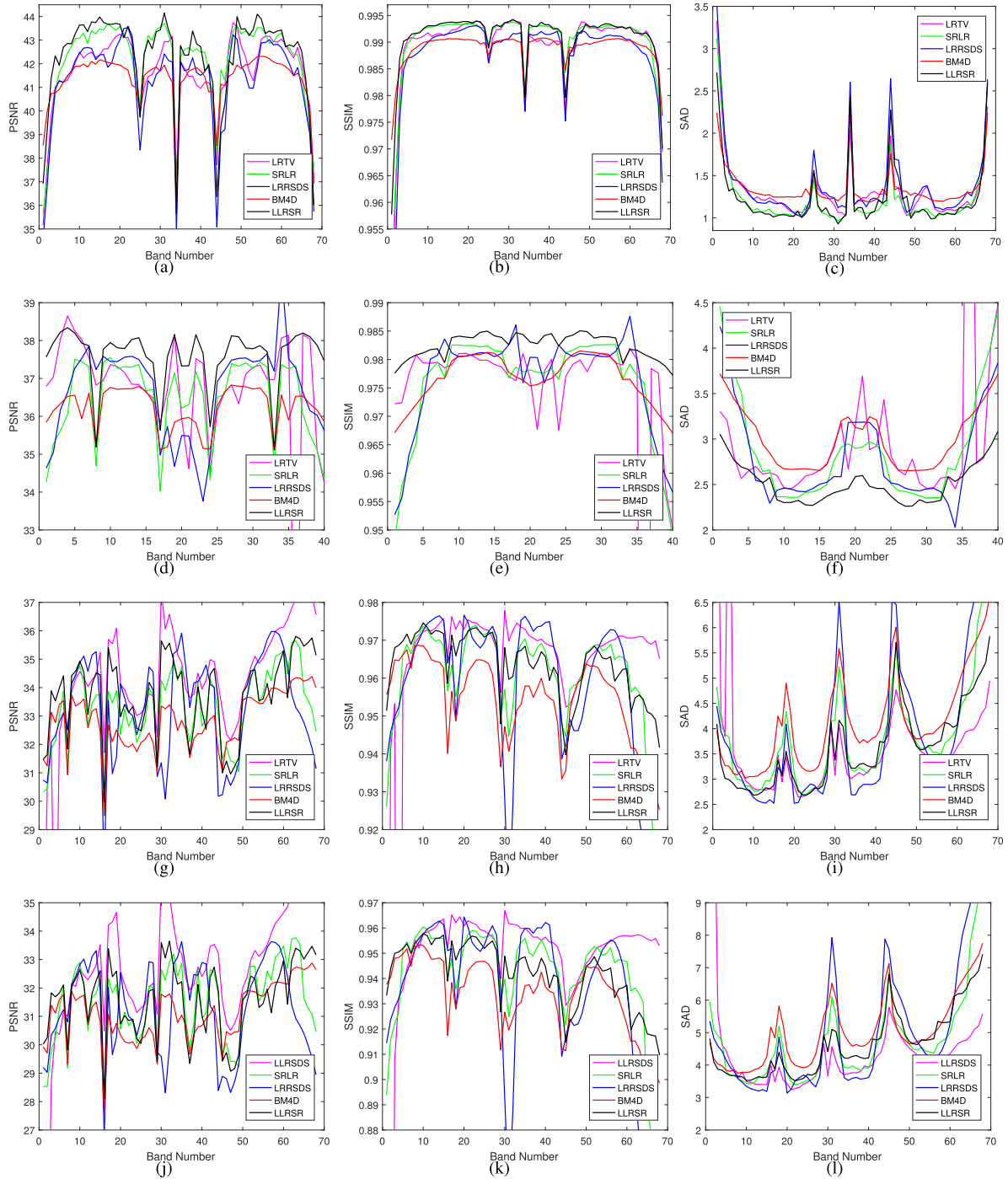
FIGURE 11. PQIs of the denoising results for three different datasets by five different methods. (a)-(c) the San Diego dataset under $\sigma = 0.025$. (d)-(f) the Pavia City Center dataset under $\sigma = 0.05$. (g)-(i) the Washington DC Mall dataset under $\sigma = 0.075$. (j)-(l) the Washington DC Mall dataset under $\sigma = 0.1$.

In this case, the minimization problem (12) becomes

$$\min_{X} \gamma \|X - Y\|_F^2 + \sum_{i=1}^{J} \mu \|R_i X\|_{w_i, *}. \qquad (25)$$

Table 6 shows that the PQIs of LLRSR are better than those of SR and LLR. It means that LLRSR works better than SR and LLR in the experiment. Therefore, it demonstrates the

rationality to simultaneously utilize the sparse representation prior and the local low-rank prior rather than to utilize only one of them.

D. A REAL EXPERIMENT
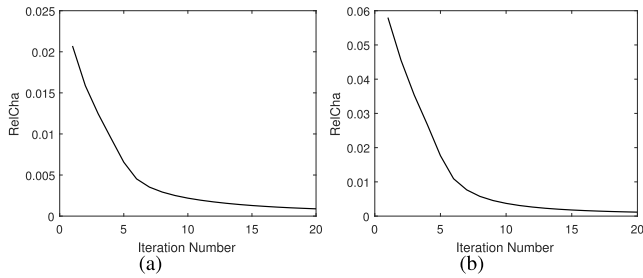In this section, we present the effectiveness of LLRSR on a real dataset. In the experiment, we use the Indian Pines

FIGURE 12. RelCha in each iteration loop of (a) the Washington DC mall dataset under $\sigma = 0.025$ and (b) the San Diego dataset under $\sigma = 0.075$.
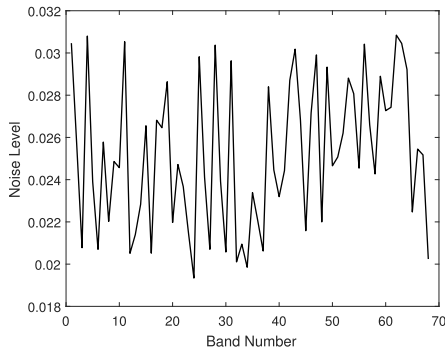

FIGURE 13. Noise level of each band in the Washington, DC, mall dataset.

FIGURE 14. Noise level of each band in the Indian pines dataset before and after the preprocessing step.

dataset[2] of size $145 \times 145 \times 220$. Images in spectral bands 1, 104-109, 150-163, and 219-220 are abandoned because the noise levels in these bands are high and there is nearly no useful information in these bands.

In the real experiment, the noise level of each band of the HSI is different. So we first estimate the noise level of each band by a fast and accurate local noise level estimation method proposed in [54]. Similarly as in Section IV-B,

---
[2]Available from http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

in order to apply LLRSR to the real noisy data, we apply the preprocessing step in (23) to make the noise level of each band be equal. Fig. 14 shows the estimated noise level of each band in the Indian Pines dataset before and after the preprocessing step.

Similarly as in Section IV-B, we denote LLRSR with the preprocessing step as $LLRSR_p$. Figs. 15 and 16 show the images of bands 103 and 142 before and after denoising via the five different methods. It can be observed that BM4D is hard to restore the details. Also, SRLR remains some noise. Clearly, LRRSDS, LRTV, LLRSR, and $LLRSR_p$ achieve better denoising results. They remove most of the noise while preserving many details. In particular, $LLRSR_p$ gets better denoising results comparing to LLRSR. In Table 7, we utilize a blind HSI quality assessment introduced in [55] to evaluate the HSIs before and after denoising. A better denoising result is indicated by a lower blind HSI quality assessment score [55]. As shown in Table 7, $LLRSR_p$ achieves the lowest score, which means that the $LLRSR_p$ gets a better result than other four compared methods. In addition, Table 7 also shows the effectiveness of the preprocessing step in $LLRSR_p$.

TABLE 6. PQIs of SR, LLR, and LLRSR on different datasets under different noise levels.

Noise Level	PQIs	Pavia City Center			San Diego			Washington DC Mall		
		SR	LLR	LLRSR	SR	LLR	LLRSR	SR	LLR	LLRSR
$\sigma = 0.025$	MPSNR	38.89	41.64	**42.40**	39.87	41.15	**42.37**	37.67	40.80	**41.06**
	MSSIM	0.9864	0.9922	**0.9940**	0.9835	0.9876	**0.9909**	0.9841	0.9913	**0.9923**
	MSAD	2.202	1.672	**1.488**	1.598	1.399	**1.225**	2.349	1.715	**1.619**
$\sigma = 0.05$	MPSNR	33.82	36.28	**37.60**	33.02	36.07	**37.88**	33.02	35.83	**36.62**
	MSSIM	0.9595	0.9738	**0.9824**	0.9578	0.9616	**0.9757**	0.9578	0.9743	**0.9802**
	MSAD	3.812	3.084	**2.514**	3.890	2.489	**2.024**	3.890	3.013	**2.629**
$\sigma = 0.075$	MPSNR	30.95	33.17	**34.58**	30.34	33.26	**35.01**	30.34	32.81	**33.75**
	MSSIM	0.9242	0.9522	**0.9657**	0.9257	0.9348	**0.9562**	0.9257	0.9535	**0.9635**
	MSAD	5.284	4.347	**3.532**	5.252	3.405	**2.785**	5.252	4.216	**3.618**
$\sigma = 0.1$	MPSNR	28.91	30.46	**32.41**	29.57	30.69	**32.89**	28.26	30.23	**31.60**
	MSSIM	0.8807	0.9187	**0.9442**	0.8704	0.8958	**0.9331**	0.8850	0.9226	**0.9422**
	MSAD	6.703	5.923	**4.553**	5.218	4.551	**3.532**	6.649	5.673	**4.635**

FIGURE 15. Band 103 of the Indian pines dataset before and after denoising via five different methods. (a) Original image. (b) Result of LRTV. (c) Result of LRRSDS. (d) Result of BM4D. (e) Result of SRLR. (f) Result of LLRSR. (g) Result of LLRSR$_p$.
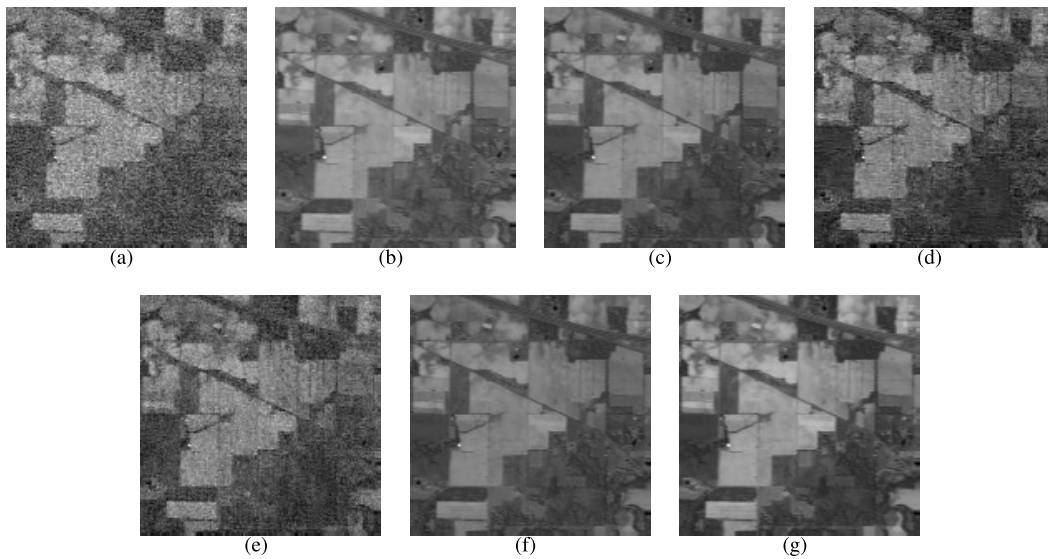

FIGURE 16. Band 142 of the Indian pines dataset before and after denoising via five different methods. (a) Original image. (b) Result of LRTV. (c) Result of LRRSDS. (d) Result of BM4D. (e) Result of SRLR. (f) Result of LLRSR. (g) Result of LLRSR$_p$.

TABLE 7. Blind HSI quality assessment by different denoising methods for the Indian pines dataset.

Method	LRRSDS	BM4D	SRLR	LRTV	LLRSR	LLRSR$_p$
Score	16.02	15.94	15.85	15.62	15.67	**15.59**

E. DISCUSSION

1) PARAMETERS DISCUSSION

Recall that the first parameter γ in (14) is the weight of the fidelity term. Parameter γ determines the influence degree of noisy HSI Y on the process of denoising. Large γ guarantees that the recovered HSI X is close to the noisy HSI Y. In consequence, when the noise level is lower, γ is larger, and vice versa, as Table 1 lists. In Fig. 17, γ changes from
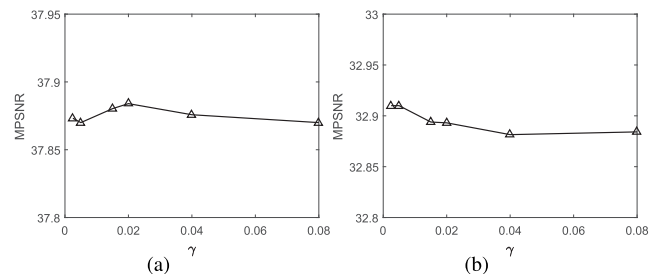
FIGURE 17. MPSNR of the san diego dataset with different γ under two noise levels (a) $\sigma = 0.05$ and (b) $\sigma = 0.1$.

the set of {0.0025, 0.005, 0.015, 0.02, 0.04, 0.08}. From the figure, results are relatively stable, and get the best with the chosen value in Table 1. As we can see, the second
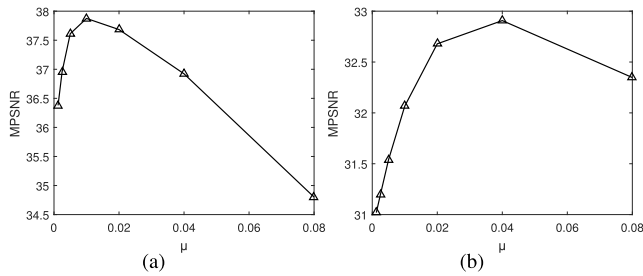
IEEE Access

FIGURE 18. MPSNR of the san diego dataset with different μ under two noise levels (a) $\sigma = 0.05$ and (b) $\sigma = 0.1$.
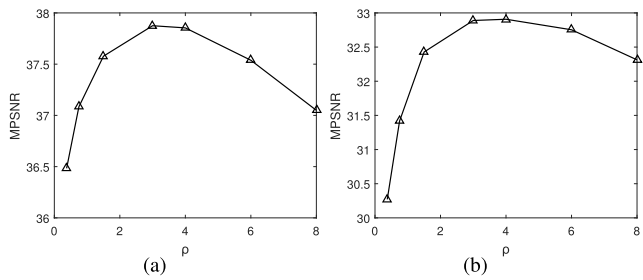

FIGURE 19. MPSNR of the san diego dataset with different ρ under two noise levels (a) $\sigma = 0.05$ and (b) $\sigma = 0.1$.
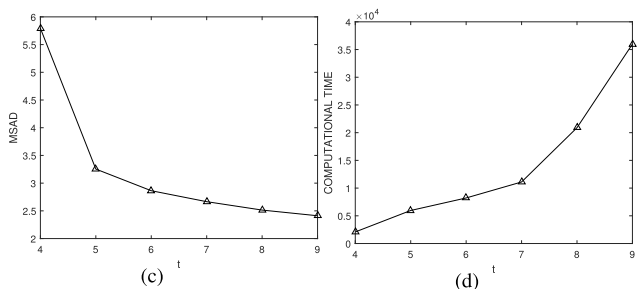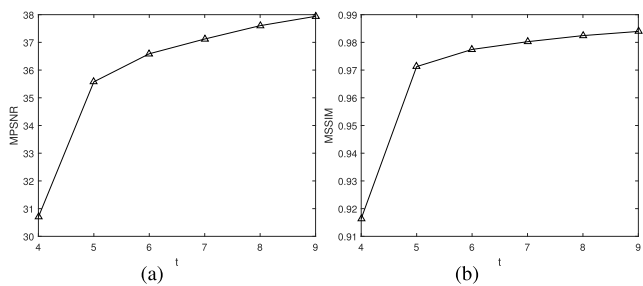

FIGURE 20. PQIs of denoising the Washington DC mall dataset with different patch sizes $t \times t$ under $\sigma = 0.05$. (a) MPSNR. (b) MSSIM. (c) MSAD. (d) Computational Time.

FIGURE 21. PQIs in each iteration loop of denoising the Washington, DC, mall dataset under different noise levels. (a) MPSNR and (b) MSSIM.

FIGURE 22. PQIs in each iteration loop of denoising the san diego dataset under different noise levels. (a) MPSNR and (b) MSSIM.

FIGURE 23. PQIs in each iteration loop of denoising the pavia city center dataset under different noise levels. (a) MPSNR and (b) MSSIM.

fourth parameter ρ constrains auxiliary variable P_i to be close to the local patch $R_i X$. In Fig. 19, ρ changes from the set of {0.375, 0.75, 1.5, 3, 4, 6, 8}. We choose ρ from the set for the best MPSNR value.

As shown in Fig. 20, we see that there is a tradeoff between PQIs and the computational time for the same patch size. In all of our experiments, we set the patch size to be 8×8 to obtain comparable results.

2) MAXIMUM ITERATION NUMBER SETTING
We note that MPSNR and MSSIM values in each iteration loop of the Washington DC Mall dataset, the San Diego dataset, and the Pavia city center dataset under four different noise levels are recorded in Figs. 21, 22, and 23, respectively. These figures present the relation between the iteration number and the denoising performance. As the optimization procedure only needs a few iterations to get a valid result, it is sufficient to set the iteration number as 20, similarly as in [17]–[19], [28]. It is worth noting that a larger iteration number may cause over-fitting and therefore a poor denoised performance [18].

parameter is η, which is set implicitly in the process of sparse coding in (15) [28]. It keeps the balance between the representation error and sparsity. Then the third parameter is μ, which weights the local low-rank constraint. As Table 1 indicates, parameter μ is set according to the noise level. During the optimization process, the noise level decreases iteratively. So μ is divided by 10 in each iteration loop of our optimization process. In Fig. 18, μ changes from the set of {0.00125, 0.0025, 0.05, 0.01, 0.02, 0.04, 0.08}. From the figure, the result changes with different μ. Notice that the
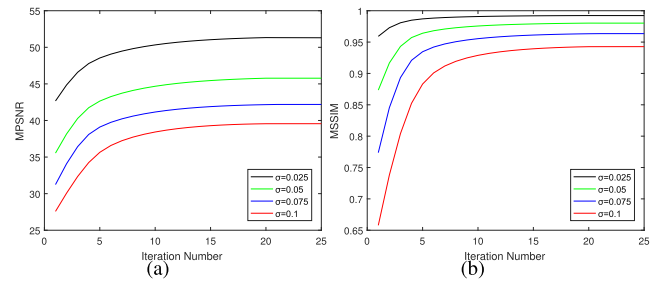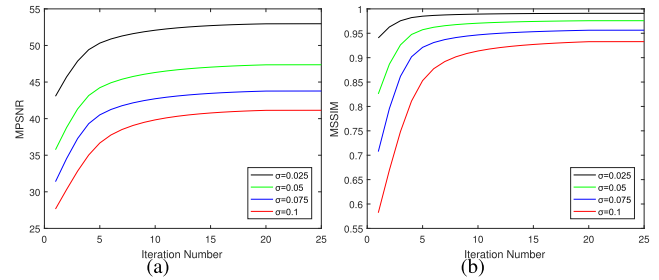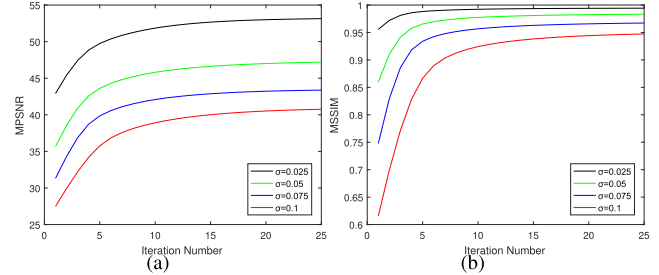
V. CONCLUSION

In this paper, we have proposed a local low-rank and sparse representation method for HSI denoising problem. We utilize the local low-rank property in our model to preserve local spectral-spatial information of HSIs. We solve the proposed model under the alternative minimization framework and obtain an algorithm called LLRSR. The weighted nuclear norm is adopted to enhance the local low-rank property. We also show the convergence of the proposed algorithm numerically. The experimental results indicate that our method removes HSI noise effectively, comparing with other four state-of-the-art methods.

We notice from the numerical experiments that dictionary learning has high computational complexity. In our future work, we will consider accelerating LLRSR by reducing the computational complexity of dictionary learning.

ACKNOWLEDGEMENT

REFERENCES

[1] K. C. Tiwari, M. K. Arora, and D. Singh, "An assessment of independent component analysis for detection of military targets from hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 5, pp. 730–740, Oct. 2011.

[2] M. T. Eismann, *Hyperspectral Remote Sensing*. Bellingham, WA, USA: SPIE, 2012.

[3] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017.

[4] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jul. 2014.

[5] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013.

[6] C. Li, X. Chen, and Y. Jiang, "On diverse noises in hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5388–5402, Oct. 2015.

[7] L. Tong, J. Zhou, Y. Qian, X. Bai, and Y. Gao, "Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6531–6544, Nov. 2016.

[8] P. Du, L. Gan, J. Xia, and D. Wang, "Multikernel adaptive collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4664–4677, Aug. 2018.

[9] F. Lv, M. Han, and T. Qiu, "Remote sensing image classification based on ensemble extreme learning machine with stacked autoencoder," *IEEE Access*, vol. 5, pp. 9021–9031, 2017.

[10] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[11] T. Skauli, "Sensor noise informed representation of hyperspectral data, with benefits for image storage and processing," *Opt. Express*, vol. 19, no. 14, pp. 13031–13046, 2011.

[12] 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.

[13] Y. Chen, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Hyperspectral image restoration using framelet-regularized low-rank nonnegative matrix factorization," *Appl. Math. Model.*, vol. 63, pp. 128–147, Nov. 2018.

[14] 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.

[15] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising with a spatial–spectral view fusion strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2314–2325, May 2014.

[16] G. Chen and S.-E. 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.

[17] 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.

[18] 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.

[19] L. Sun, B. Jeon, Y. Zheng, and Z. Wu, "Hyperspectral image restoration using low-rank representation on spectral difference image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1151–1155, Jul. 2017.

[20] 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.

[21] L. Sun, T. Zhan, Z. Wu, L. Xiao, and B. Jeon, "Hyperspectral mixed denoising via spectral difference-induced total variation and low-rank approximation," *Remote Sens.*, vol. 10, no. 12, p. 1956, 2018.

[22] L. Jia, S. Song, L. Yao, H. Li, Q. Zhang, Y. Bai, and Z. Gui, "Image denoising via sparse representation over grouped dictionaries with adaptive atom size," *IEEE Access*, vol. 5, pp. 22514–22529, 2017.

[23] Y. Chen, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, and J. Huang, "Stripe noise removal of remote sensing images by total variation regularization and group sparsity constraint," *Remote Sens.*, vol. 9, no. 6, p. 559, Jun. 2017.

[24] L. Ma, J. Yu, and T. Zeng, "Sparse representation prior and total variation–based image deblurring under impulse noise," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 2258–2284, 2013.

[25] L. Sun, B. Jeon, B. N. Soomro, Y. Zheng, Z. Wu, and L. Xiao, "Fast superpixel based subspace low rank learning method for hyperspectral denoising," *IEEE Access*, vol. 6, pp. 12031–12043, 2018.

[26] 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.

[27] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019.

[28] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[29] 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.

[30] R. Abergel, C. Louchet, L. Moisan, and T. Zeng, "Total variation restoration of images corrupted by Poisson noise with iterated conditional expectations," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.* Cham, Switzerland: Springer, 2015, pp. 178–190.

[31] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[32] X. Liu, X. Lu, H. Shen, Q. Yuan, Y. Jiao, 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.

[33] S. Jia, X. Zhang, and Q. Li, "Spectral–spatial hyperspectral image classification using $\ell_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2473–2484, Jun. 2015.

[34] J. Huang, T.-Z. Huang, L.-J. Deng, and X.-L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2419–2438, Apr. 2019.

[35] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4775–4789, Aug. 2016.

[36] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

[37] 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.

[38] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "Learning stable multilevel dictionaries for sparse representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1913–1926, Sep. 2015.

[39] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[40] J. Huang and T.-Z. Huang, "A nonstationary accelerating alternating direction method for frame-based Poissonian image deblurring," *J. Comput. Appl. Math.*, vol. 352, pp. 181–193, May 2019.

[41] J. Huang, M. Donatelli, and R. H. Chan, "Nonstationary iterated thresholding algorithms for image deblurring," *Inverse Problems Imag.*, vol. 7, no. 3, pp. 717–736, 2013.

[42] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[43] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3618–3632, Aug. 2014.

[44] J. Yang, Y.-Q. Zhao, J. C.-W. Chan, and S. G. Kong, "Coupled sparse denoising and unmixing with low-rank constraint for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1818–1833, Mar. 2016.

[45] Q. Wang, Z. Wu, J. Jin, T. Wang, and Y. Shen, "Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising," *Signal Process.*, vol. 142, pp. 11–26, Jan. 2018.

[46] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vols. 397–398, pp. 48–68, Aug. 2017.

[47] J. Xue, Y. Zhao, W. Liao, and S. G. Kong, "Joint spatial and spectral low-rank regularization for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1940–1958, Apr. 2018.

[48] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, and Z. Xu, "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4075–4090, Sep. 2016.

[49] Z. Wu, Q. Wang, J. Jin, and Y. Shen, "Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *Signal Process.*, vol. 131, pp. 202–219, Feb. 2017.

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[51] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 38–42, Jan. 2015.

[52] J. Hu, Y. Li, and W. Xie, "Hyperspectral image super-resolution by spectral difference learning and spatial error correction," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1825–1829, Oct. 2017.

[53] Y.-B. Zheng, T.-Z. Huang, T.-Y. Ji, X.-L. Zhao, T.-X. Jiang, and T.-H. Ma, "Low-rank tensor completion via smooth matrix factorization," *Appl. Math. Model.*, vol. 70, pp. 677–695, Jun. 2019.

[54] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[55] J. Yang, Y. Zhao, C. Yi, and J. C.-W. Chan, "No-reference hyperspectral image quality assessment via quality-sensitive features learning," *Remote Sens.*, vol. 9, no. 4, p. 305, Mar. 2017.
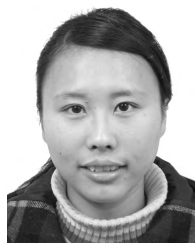
GUANQUN MA received the B.S. degree in mathematics and applied mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China, in 2016, where he is currently pursuing the M.S. degree in computational mathematics. His research interests include hyperspectral image processing and image quality improvement.

TING-ZHU HUANG received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xian Jiaotong University, Xi'an, China.

He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Dr. Huang is an Editor of *The Scientific World Journal*, *Advances in Numerical Analysis*, the *Journal of Applied Mathematics*, the *Journal of Pure and Applied Mathematics: Advances in Applied Mathematics*, and the *Journal of Electronic Science and Technology*, China.

JIE HUANG received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013, where she is currently an Associate Professor with the School of Mathematical Sciences. Her research interests include remote sensing and image processing.

CHAO-CHAO ZHENG received the B.S. degree in mathematics and applied mathematics and the M.S. degree in computational mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Department of Applied Mathematics, Hong Kong Baptist University. His research interests include computer vision and image process.

. . .