Adaptive Spatial-Spectral Dictionary Learning for Hyperspectral Image Denoising

Ying Fu[1], Antony Lam[2], Imari Sato[3], Yoichi Sato[1]

[1]The University of Tokyo   [2]Saitama University   [3]National Institute of Informatics

Abstract

Hyperspectral imaging is beneficial in a diverse range of applications from diagnostic medicine, to agriculture, to surveillance to name a few. However, hyperspectral images often times suffer from degradation due to the limited light, which introduces noise into the imaging process. In this paper, we propose an effective model for hyperspectral image (HSI) denoising that considers underlying characteristics of HSIs: sparsity across the spatial-spectral domain, high correlation across spectra, and non-local self-similarity over space. We first exploit high correlation across spectra and non-local self-similarity over space in the noisy HSI to learn an adaptive spatial-spectral dictionary. Then, we employ the local and non-local sparsity of the HSI under the learned spatial-spectral dictionary to design an HSI denoising model, which can be effectively solved by an iterative numerical algorithm with parameters that are adaptively adjusted for different clusters and different noise levels. Experimental results on HSI denoising show that the proposed method can provide substantial improvements over the current state-of-the-art HSI denoising methods in terms of both objective metric and subjective visual quality.

1. Introduction

Hyperspectral imaging is the process of capturing images of a scene over multiple bands of the electromagnetic spectrum. When captured, a hyperspectral image (HSI) can be thought of as a set of 2D spatially organized "pixels" where each pixel contains an entire spectral distribution over wavelengths. This allows us to see the spectral distribution of any given surface point in a scene and has led to numerous applications in remote sensing [1], classification and detection [2, 3], diagnostic medicine [4], surveillance [5], biometrics [6], and more. However, conventional hyperspectral imaging suffers from issues such as limited light in individual bands, which introduces noise into the imaging process. The degradation of HSIs not only influences the visual appearance of these images but also limits the performance of classification and detection methods.

Therefore, HSI denoising is a critical step in HSI based system pipelines.

Recently, sparse representation based models have shown their popularity and effectiveness in image restoration. The success of sparse representation models owes to the development of l_1-norm optimization techniques [7, 8, 9, 10] and the finding that natural images are intrinsically sparse in many domains [11]. These domains can be represented by a dictionary, which can be constructed from basis functions (e.g. DCT, wavelet) and adaptive learning (e.g. K-SVD [12], stochastic approximations [13]). In addition, due to natural images containing self-repeating patterns, the concept of non-local self-similarity in images is introduced in many papers on topics such as denoising [14, 15], demosaicing [16], super-resolution[17, 18], and frame rate upconversion [19].

Dong *et al.* [20, 21] was able to combine both sparse representation and non-local self-similarity to restore grayscale natural images to great effect. However, their grayscale image approach is not ideally suited to HSIs since the relationship between bands is not considered. On the other hand, Lam *et al.* [22] showed that spectral domain statistics can help HSI denoising. This suggests there is rich information in the spectral domain that can be utilized. Therefore, the potential use of sparse representation, non-local self-similarity, and spectral domain statistics jointly should be studied.

In this paper, we propose a novel HSI denoising method based on adaptive spatial-spectral dictionary learning. We first adaptively learn a spatial-spectral dictionary for each overlapping cubic patch of the noisy HSI, which can sparsely represent its own cubic patches across both spatial and spectral domains as well as consider the non-local self-similarity in of the HSI. In the dictionary learning process, the cubic patches are clustered into groups in terms of the their similarity across the full HSI. The dictionary is then learned for each group. In addition, we exploit the the local and non-local sparsity of the HSI to design a denoising model for it. We show that this model can be effectively solved by a numerical algorithm with its regularization parameters, adaptively adjusted for different clusters and dif-

ferent noise levels.

In summary, our main contributions are that we

1. Exploit the high correlation across spectra and non-local self-similarity over space to learn an adaptive spatial-spectral dictionary for efficient HSI denoising,

2. Take into account the local and non-local sparsity of the HSI under the learned dictionary to design the HSI denoising model,

3. Develop a numerical algorithm to effectively solve the proposed model and adaptively adjust the regularization parameters for different clusters and different noise levels.

We show our method is accurate and demonstrate its effectiveness through experimental results on HSI denoising, which outperforms many state-of-the-art methods under several comprehensive quantitative assessments.

2. Related Work

A number of algorithms such as K-SVD [12], Non-local means [14] and BM3D [15] have been developed for 2D image denoising. These methods can be directly applied to HSI denoising by denoising each band image independently. However, considering the bands to be independent limits performance. Qian *et al.* [23] and Maggioni *et al.* [24] denoise HSIs by considering 3D cubes of the HSI instead of the 2D patches of traditional image restoration but these methods ignore the high correlation across spectra and limits their performance.

There are several methods based on Wavelet and/or PCA for HSI denoising. Atkinson *et al.* [25] and Othman and Qian [26] both proposed wavelet-based HSI denoising algorithms. Chen *et al.* [27, 28] proposed performing dimensionality reduction and HSI denoising using wavelet shrinkage and principal component analysis (PCA). Lam *et al.* [22] used PCA for dimensionality reduction in the spectral domain and then performed denoising in local spatial neighborhoods to further improve denoising results.

Several methods used tensor decompositions. Letexier and Bourennane [29] proposed an HSI denoising method based on a generalized multidimensional Wiener filter by using a third-order tensor. Renard *et al.* [30] used a low rank tensor approximation to jointly achieve denoising and spectral dimensionality reduction. Karami *et al.* [31] reduced the noise of HSIs by using their Genetic Kernel Tucker Decomposition. Guo *et al.* [32] denoised the HSIs based on a high-order rank-1 tensor decomposition.

In addition, Wang *et al.* [1, 33] utilized an alternative hyperspectral anisotropic diffusion scheme to denoise HSIs. Yuan *et al.* [34] employed a spatial-spectral adaptive total variation model for their denoising scheme. Zhong and Wang [35] simultaneously model and utilize spatial and

spectral dependence in a unified probabilistic framework by multiple-spectral-band conditional random fields. Also, Zhang *et al.* [36] explored the low-rank property of HSIs for image restoration. Many of these methods exploited both spatial and spectral information. However, none have utilized the non-local self-similarity present in HSIs of natural scenes.

In [37], Qian and Ye introduced the non-local self-similarity and spatial-spectral structure of HSIs into a sparse representation framework, but their method used a global 3D dictionary built from 3D DCT and 3D OWT, which cannot be adapted to the specific characteristics of a given scene. Peng *et al.* [38] proposed an HSI denoising approach based on decomposable non-local tensor dictionary learning, but their model did not explicitly take the neighboring non-local self-similarity into account.

3. Hyperspectral Image Denoising Model

An HSI degraded by additive noise can be modeled as
$$Y = X + n, \tag{1}$$
where X is the original clean HSI, with a size of $M \times N \times B$, in which M stands for the rows of the image, N represents the number of columns, and B is the number of bands. Y is the degraded HSI and n is additive noise with the same size as X. The overall framework of the proposed method is shown in Fig. 1.

3.1. Adaptive Spatial-Spectral Dictionary Learning

We first show how we build the adaptive spatial-spectral dictionary, which considers both non-local self-similarities and the high correlation across spectra. The HSI is first divided into overlapping cubic patches of size $P \times P \times B$, where $P < M$ and $P < N$. In our method, we consider the two spatial dimensions as a whole. Each cubic patch centered at the spatial location (i, j) can be described as $Z_{i,j} \in \mathbb{R}^{P^2 \times B}$. Each element in $Z_{i,j}$ can be denoted as $Z_{i,j}[n, b]$, where $n \in 1, 2, \cdots, P^2$ and $b \in 1, 2, \cdots, B$ index the spatial pixel location and spectral band, respectively.

In studies of the human visual system [11, 39], the receptive fields of the eye's cells code natural images using a small number of structural primitives sparsely chosen from an over-complete code set. This suggests that natural images are intrinsically sparse in some domains, which can be described as a dictionary and constructed by basis functions. Therefore, we can seek a representation that allows us to analyze the high-dimensional $Z_{i,j}$ using a smaller number of components. Then, $Z_{i,j}$ can be represented in terms of a linear superposition of basis functions,
$$Z_{i,j}[n, b] = \sum_m \alpha_m \phi_m[n, b], \tag{2}$$
where $\{\phi_m\}$ is a basis set, and α_m are the corresponding scalar coefficients.
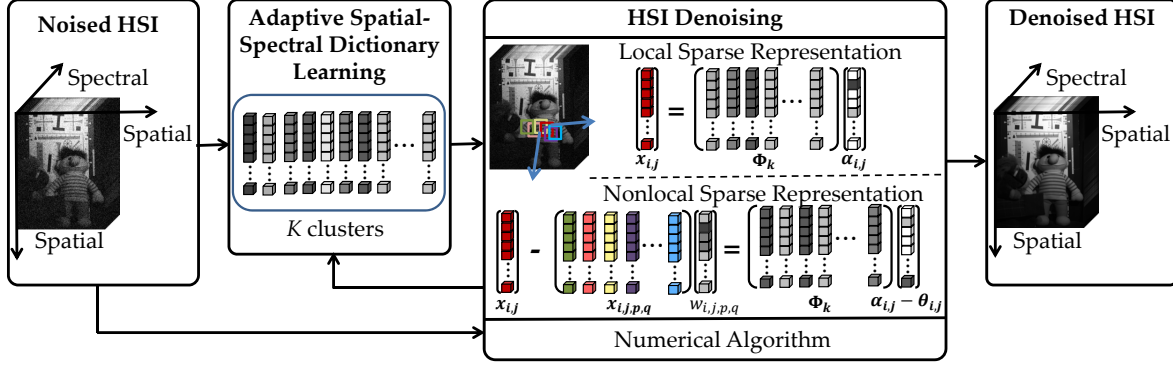
Figure 1. Overview of the proposed method. The noised HSI is first used to learn the adaptive spatial-spectral dictionary. Then, the trained dictionary is employed to sparsely represent the HSI using both local details and non-local self-similarities of the image. These two terms are integrated into a unified variational framework for optimization, which is solved by an iterative numerical algorithm.
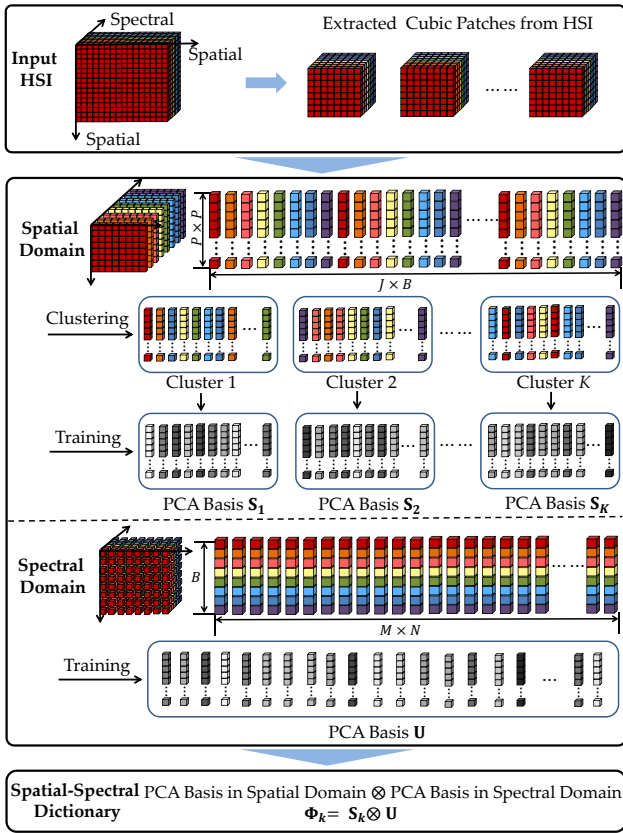

Figure 2. Illustration of adaptive spatial-spectral dictionary learning. The cubic patches are extracted from the HSI in the overlapping pattern first. Then, all the the band images of the cubic patches are clustered in the spatial domain. The spatial PCA bases for each cluster are then learned independently. In the spectral domain, all spectral distributions from the entire HSI are extracted as vectors and used to train the spectral PCA basis. Finally, the adaptive spatial-spectral dictionary is derived from the spatial and spectral PCA bases in which each cluster has its own spatial-spectral dictionary.

According to previous research [40], an HSI's dictionary

can be decomposed into a product of separate spatial and spectral components, that is to say,

$$\phi_m[n, b] = S_d[n]U_r[b], \quad (3)$$

where $\{S_d\}_1^{P^2}$ and $\{U_r\}_1^{B}$ are orthonormal bases spanning the space of monochrome $P \times P$ spatial patches and the space of B-channel spectral distributions, respectively. In practice, $\phi_m \in \mathbb{R}^{P^2 \times B}$ is reshaped to a vector $\Phi_m \in \mathbb{R}^{P^2 B}$. This means that spatial bases $\{S_d\}_1^{P^2}$ and spectral bases $\{U_r\}_1^{B}$ can be separately trained, and then combined to obtain the joint spatial-spectral dictionary.

As to how we specifically represent the bases for the spatial and spectral domains, we first provide details for the spectral domain. It is well known that large sets of spectra consisting of natural scenes or extensive color samples can be represented compactly in just $6 \sim 8$ principal components [41]. So in the spectral domain, we reshape the original HSI into $B \times MN$ and learn the spectral bases $\{U_r\}_1^{B}$ by PCA [40, 41]. On the other hand, the spatial patches can also be represented by PCA bases [20, 40]. A simple test on the Columbia Multispectral Image Dataset [42] reveals that 10,000 random 7×7 spatial patches require only 6 principal components to capture 99% of the variance. We also train the spatial bases $\{S_d\}_1^{P^2}$ using PCA from monochrome patches pooled across all bands in the spatial domain. The spatial-spectral dictionary $\{\phi_m\}$ can then be constructed by different combinations of S_p and U_r as shown in Eq. (3).

At this point, we could learn a single universal dictionary for all spatial patches like in the cases of the analytically designed wavelet dictionary [37] and the learned K-SVD [12]. Indeed, one dictionary can be used to represent any image patch. However, it would lack sufficient flexibility to sparsely represent a given local patch. In [21] [38], they collected patches that were similar to each other to learn a local dictionary for each cluster instead of a universal dictionary, thus allowing for more sparse representations. Inspired by their methods, instead of learning a single universal dictio-

nary, we first cluster 2D monochrome patches pooled across all bands in the spatial domain by k-means. This clustering process considers the non-local self-similarity of the patches across the full HSI, and we called it the *global non-local self-similarity*. In practice, we cluster spatial image patches by their high frequency patterns (rich textures and edges) as opposed to clustering directly based on the pixel values in the patches to avoid the influence from overly similar intensities. The image of high frequency spatial patterns can be obtained by

$$\boldsymbol{X}_g = \boldsymbol{X} - \boldsymbol{X} * g, \tag{4}$$

where g is a Gaussian kernel convolved with the original image. Note that the patches from high frequency image $\boldsymbol{X}_g$ are only used to cluster similar patches into clusters. The corresponding original patches $\boldsymbol{x}_{i,j}$ from $\boldsymbol{X}$ are used to learn the PCA basis for each cluster.

Let $\mathbf{S}_k = [S_1^{(k)}, S_2^{(k)}, \cdots, S_{P^2}^{(k)}]$ represent the basis in the spatial domain for cluster k ($k = 1, 2, \cdots, K$) and $\mathbf{U} = [U_1, U_2, \cdots, U_B]$ be the basis in the spectral domain. The combined spatial-spectral basis vectors are then formed by $S_d^{(k)} \otimes U_r$, for all pairs of d and r in cluster k. We obtain

$$\boldsymbol{\Phi}_k = \mathbf{S}_k \otimes \mathbf{U}, \tag{5}$$

where $\otimes$ is the Kronecker product and $\boldsymbol{\Phi}_k = [\Phi_1^{(k)}, \Phi_2^{(k)}, \cdots, \Phi_{P^2B}^{(k)}]$ is the spatial-spectral dictionary, which is adaptively learned for each cluster and considers the spatial and spectral dimensions together in an explicit manner. Fig. 2 illustrates the adaptive spatial-spectral dictionary learning process.

3.2. Local and Non-local Sparse Representation Model

Adopting the notational convention in [12], the HSIs $\boldsymbol{Y}$ and $\boldsymbol{X}$ are reshaped as vectors $\mathbf{Y} \in \mathbb{R}^{MNB}$ and $\mathbf{X} \in \mathbb{R}^{MNB}$ in the following derivations. Let vector $\boldsymbol{x}_{i,j} \in \mathbb{R}^{P^2B}$ denote a cubic patch extracted from HSI $\mathbf{X}$ and centered at the spatial location (i, j),

$$\boldsymbol{x}_{i,j} = \mathbf{R}_{i,j}\mathbf{X}, \tag{6}$$

where $\mathbf{R}_{i,j} \in \mathbb{R}^{P^2B \times MNB}$ is the matrix extracting patch $\boldsymbol{x}_{i,j}$ from $\mathbf{X}$. The cubic patches are extracted by shifting over 2D spatial locations such that we obtain J cubic patches, where $J = (M - P + 1)(N - P + 1)$. Please note that vector $\boldsymbol{x}_{i,j}$ is obtained by reshaping matrix $Z_{i,j}$ as a vector.

As mentioned previously, the learned adaptive spatial-spectral dictionary $\boldsymbol{\Phi}_k \in \mathbb{R}^{BP^2 \times m}$[1] can be used to sparsely code cubic patches $\boldsymbol{x}_{i,j}$ and denoted as $\boldsymbol{\Phi}_k \boldsymbol{\alpha}_{i,j}$, when the cubic patch at location (i, j) belongs to k-th cluster C_k. Given the adaptive spatial-spectral dictionary $\boldsymbol{\Phi}_k$, the

sparse codes $\boldsymbol{\alpha}$ of image $\mathbf{X}$ can be recovered by

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{i,j}\sum_{k=1}^{K}$$
$$\left(\gamma\|\boldsymbol{x}_{i,j} - \boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}\|_2^2 + \lambda\|\boldsymbol{\alpha}_{i,j}\|_1\right), \tag{7}$$

where $\boldsymbol{\alpha}$ denotes the concatenation of all $\boldsymbol{\alpha}_{i,j}$, and $\hat{\boldsymbol{\alpha}}$ is the estimation of $\boldsymbol{\alpha}$.

The cubic patches in the HSI also have rich self-similarity with its neighboring patches [23], which implies the cubic patch can be effectively described by its neighboring patches. We call this the *neighboring non-local self-similarity* of the HSI to distinguish it with the *global non-local self-similarity in Section 3.1*. In practice, we employ non-local means to describe the *neighboring non-local self-similarity* of the HSI. The non-local means $\hat{\boldsymbol{x}}_{i,j}$ can be computed by

$$\hat{\boldsymbol{x}}_{i,j} = \sum_{(p,q)\in\boldsymbol{\Omega}_{i,j}} w_{i,j,p,q}\boldsymbol{x}_{i,j,p,q}, \tag{8}$$

where $\boldsymbol{x}_{i,j,p,q}$ denotes a found similar cubic patch $\boldsymbol{x}_{p,q}$ within set $\boldsymbol{\Omega}_{i,j}$, which represents a large neighborhood around cubic patch $\boldsymbol{x}_{i,j}$, and $w_{i,j,p,q}$ is the corresponding weight. Like in [14], $w_{i,j,p,q}$ can be set as

$$w_{i,j,p,q} = \frac{1}{W}\exp\left(-\|\boldsymbol{x}_{i,j} - \boldsymbol{x}_{i,j,p,q}\|_2^2/h\right) \tag{9}$$

where W is a normalization factor and h is a pre-determined scalar. Then, we can obtain

$$\boldsymbol{x}_{i,j} - \hat{\boldsymbol{x}}_{i,j} = \boldsymbol{x}_{i,j} - \sum_{(p,q)\in\boldsymbol{\Omega}_{i,j}} w_{i,j,p,q}\boldsymbol{x}_{i,j,p,q}$$
$$= \boldsymbol{\Phi}_k\left(\boldsymbol{\alpha}_{i,j} - \sum_{(p,q)\in\boldsymbol{\Omega}_{i,j}} w_{i,j,p,q}\boldsymbol{\alpha}_{i,j,p,q}\right)$$
$$= \boldsymbol{\Phi}_k\left(\boldsymbol{\alpha}_{i,j} - \boldsymbol{\theta}_{i,j}\right), \tag{10}$$

where $\boldsymbol{\theta}_{i,j} = \sum_{(p,q)\in\boldsymbol{\Omega}_{i,j}} w_{i,j,p,q}\boldsymbol{\alpha}_{i,j,p,q}$ is the sparse representation of the non-local means $\hat{\boldsymbol{x}}_{i,j}$. Since $\hat{\boldsymbol{x}}_{i,j}$ approximates $\boldsymbol{x}_{i,j}$, the sparse codes $\boldsymbol{\theta}_{i,j}$ of $\hat{\boldsymbol{x}}_{i,j}$ should be similar to the sparse codes $\boldsymbol{\alpha}_{i,j}$ of $\boldsymbol{x}_{i,j}$. According to [37, 43], the locations of the non-zero elements in the sparse codes are similar for the similar patches. Therefore, it is reasonable to assume that sparse code $\boldsymbol{\alpha}_{i,j}$ minus sparse code $\boldsymbol{\theta}_{i,j}$ should be sparse. Then we employ $\|\boldsymbol{\alpha}_{i,j} - \boldsymbol{\theta}_{i,j}\|_1$ as the non-local sparse constraint like [44]. In our method, we add this non-locally sparse constraint into Eq (7) as a regularization,

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{i,j}\sum_{k=1}^{K}$$
$$\left(\gamma\|\boldsymbol{x}_{i,j} - \boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}\|_2^2 + \lambda\|\boldsymbol{\alpha}_{i,j}\|_1 + \eta\|\boldsymbol{\alpha}_{i,j} - \boldsymbol{\theta}_{i,j}\|_1\right). \tag{11}$$

We can see that our model (Eq. (11)) unifies the local sparsity (i.e. $\|\boldsymbol{\alpha}_{i,j}\|_1$) and neighboring self-similarity induced non-local sparsity (i.e. $\|\boldsymbol{\alpha}_{i,j} - \boldsymbol{\theta}_{i,j}\|_1$) into a variational

---

[1]When the dictionary $\boldsymbol{\Phi}_k$ is obtained from PCA, $m = BP^2$.

formulation. The model also clusters the HSI cubic patches across the spatial domain to learn the adaptive spatial-spectral dictionary by considering the global non-local self-similarity over the entire HSI and the high correlation across spectra (i.e. we perform a summation $\sum_{k=1}^{K}$ and $\boldsymbol{\Phi}_k$ over the K clusters for the cost function). Therefore, our model exploits both local and non-local redundancies across the spatial-spectral domain to restore the HSI.

3.3. Numerical Algorithm

Here, we mainly show how to solve our model (Eq. (11)) in practice. In Eq. (11), we can see that there are 3 variables, i.e. $\mathbf{X}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\theta}$. $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are the concatenation of all $\boldsymbol{\alpha}_{i,j}$ and $\boldsymbol{\theta}_{i,j}$, respectively. To solve Eq. (11), we derive update rules for $\mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\theta}$ and adopt an alternating minimization scheme to reduce the original problem into simpler sub-problems.

Update $\boldsymbol{\theta}$. We initialize $\boldsymbol{\theta}_{i,j}$ to 0, i.e. $\boldsymbol{\theta}_{i,j}^{(0)} = 0$, where $a^{(t)}$ represents the t-th iteration of variable a. Then, setting $\mathbf{X}^{(0)} = \mathbf{Y}$ and $\eta = 0$, the sparse coding $\boldsymbol{\alpha}_{i,j}^{(0)}$ can be obained by $\boldsymbol{\alpha}_{i,j}^{(0)} = \boldsymbol{\Phi}_k^T \boldsymbol{x}_{i,j}^{(0)}$. The similar patches for each local sparse representation $\boldsymbol{\alpha}_{i,j}^{(0)}$ are selected from its neighborhood in terms of the squared difference between them, i.e. $\|\boldsymbol{\alpha}_{i,j} - \boldsymbol{\alpha}_{i,j,p,q}\|^2$. Hence the non-local mean $\boldsymbol{\theta}_{i,j}^{(t)}$ can be updated by

$$\boldsymbol{\theta}_{i,j}^{(t)} = \sum_{(p,q)\in\boldsymbol{\Omega}_{i,j}} w_{i,j,p,q}\boldsymbol{\alpha}_{i,j,p,q}^{(t-1)}. \tag{12}$$

In the t-th iteration, the model can be described as

$$\{\hat{\boldsymbol{\alpha}}^{(t)}, \hat{\mathbf{X}}^{(t)}\} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{i,j}\sum_{k=1}^{K}$$

$$\left(\gamma\|\boldsymbol{x}_{i,j} - \boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}\|_2^2 + \lambda\|\boldsymbol{\alpha}_{i,j}\|_1 + \eta\|\boldsymbol{\alpha}_{i,j} - \boldsymbol{\theta}_{i,j}^{(t)}\|_1\right). \tag{13}$$

Update $\boldsymbol{\alpha}$. In this sub-problem, we fix all other variables and optimize sparse code $\boldsymbol{\alpha}$ by

$$\hat{\boldsymbol{\alpha}}^{(t)} = \arg\min_{\boldsymbol{\alpha}} \sum_{i,j}\sum_{k=1}^{K}\left(\gamma\|\boldsymbol{x}_{i,j}^{(t-1)} - \boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}\|_2^2\right.$$

$$\left. + \lambda\|\boldsymbol{\alpha}_{i,j}\|_1 + \eta\|\boldsymbol{\alpha}_{i,j} - \boldsymbol{\theta}_{i,j}^{(t)}\|_1\right). \tag{14}$$

We employ a bi-variate shrinkage algorithm [20] to solve the function in Eq. (14).

Update $\mathbf{X}$. For the fixed sparse code $\boldsymbol{\alpha}$, the optimization of the $\mathbf{X}$ can be described as

$$\hat{\mathbf{X}}^{(t)} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_2^2$$

$$+ \gamma\sum_{i,j}\sum_{k=1}^{K}\|\mathbf{R}_{i,j}\mathbf{X} - \boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}^{(t)}\|_2^2. \tag{15}$$

Eq. (15) is a quadratic minimization problem and thus can

be solved in closed-form, i.e.,

$$\hat{\mathbf{X}}^{(t)} = (\mathbf{I} + \gamma\sum_{i,j}\sum_{k=1}^{K}\mathbf{R}_{i,j}^T\mathbf{R}_{i,j})^{-1}$$

$$(\mathbf{Y} + \gamma\sum_{i,j}\sum_{k=1}^{K}\sum_{(i,j)\in C_k}\mathbf{R}_{i,j}^T\boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}^{(t)}), \tag{16}$$

where $\mathbf{I}$ is an identity matrix. In our implementation, we use the conjugate gradients algorithm to compute Eq. (15).

From the above discussion, we can see that neighboring non-local self-similarity $\boldsymbol{\theta}$, the sparse coding $\boldsymbol{\alpha}$, and latent HSI $\mathbf{X}$ are alternatively updated in Eq. (11). During the iterations, the accuracy of sparse code $\boldsymbol{\alpha}^{(t)}$ is improved, which in turn, improves the accuracy of neighboring non-local self-similarity $\boldsymbol{\theta}^{(t)}$. The clustered similarity patches in the HSI across the spatial domain and the dictionary $\boldsymbol{\Phi}_k$ are also updated in terms of the updated $\boldsymbol{\alpha}^{(t)}$. After several iterations, the algorithm converges and the desired sparse code $\boldsymbol{\alpha}$ and latent HSI $\mathbf{X}$ can then be obtained.

3.4. Determination of Parameters λ and η

Previous research [44, 47], presented a well-developed sparse representation in the context of a Bayesian framework. Here, we borrow from their formulation and apply it to provide an explicit way to adaptively set the parameters λ and η for different clusters and noise levels. We consider the case where the observation $\mathbf{Y}$ is contaminated with additive Gaussian noise with standard deviation σ_n. Under the Bayesian framework, the estimation of sparsity vector $\boldsymbol{\alpha}$ can be cast as a maximum a posteriori problem,

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha}|\mathbf{Y}) = \arg\max_{\boldsymbol{\alpha}} P(\mathbf{Y}|\boldsymbol{\alpha})P(\boldsymbol{\alpha})$$

$$= \arg\min_{\boldsymbol{\alpha}}\{-\log P(\mathbf{Y}|\boldsymbol{\alpha}) - \log P(\boldsymbol{\alpha})\}, \tag{17}$$

where the two terms correspond to the likelihood and prior terms, respectively. The likelihood term can be characterized as

$$P(\mathbf{Y}|\boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(-\frac{1}{2\sigma_n^2}\|\mathbf{Y} - \mathbf{X}\|_2^2\right) \times$$

$$\prod_{i,j}\prod_{k=1}^{K}\frac{1}{\sqrt{2\pi/\gamma}\sigma_n}\exp\left(-\frac{\gamma}{2\sigma_n^2}\|\boldsymbol{x}_{i,j} - \boldsymbol{\Phi}_k\boldsymbol{\alpha}_{i,j}\|_2^2\right). \tag{18}$$

The prior term in Eq. (17) can be expressed as

$$P(\boldsymbol{\alpha}) = \prod_{i,j,l}\prod_{k=1}^{K}\frac{c_1}{\sqrt{2}\delta_{kl}}\exp\left(-\frac{\sqrt{2}c_1|\alpha_{ijl}|}{\delta_{kl}}\right)$$

$$\frac{c_2}{\sqrt{2}\varrho_{kl}}\exp\left(-\frac{\sqrt{2}c_2|\alpha_{ijl} - \theta_{ijl}|}{\varrho_{kl}}\right), \tag{19}$$

where α_{ijl} and θ_{ijl} are the l-th elements of $\boldsymbol{\alpha}_{i,j}$ and $\boldsymbol{\theta}_{i,j}$, and δ_{kl} and ϱ_{kl} are the standard deviations of α_{ijl} and $(\alpha_{ijl} - \theta_{ijl})$ when $(i,j) \in C_k$, $l = 1, 2, \cdots, P^2B$. c_1 and c_2 are predefined constants. Substituting Eqs. (18) and

(19) into Eq. (17), we obtain

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{i,j} \sum_{k=1}^{K}$$

$$\left(\gamma \|\boldsymbol{x}_{i,j} - \boldsymbol{\Phi}_k \boldsymbol{\alpha}_{i,j}\|_2^2 + c_1 \sum_l \frac{2\sqrt{2}\sigma_n^2}{\delta_{kl}} \|\alpha_{ijl}\|_1 + \right.$$

$$\left. c_2 \sum_l \frac{2\sqrt{2}\sigma_n^2}{\varrho_{kl}} \|\alpha_{ijl} - \theta_{ijl}\|_1 \right).$$

(20)

Comparing Eq. (20) and Eq. (11), we have

$$\lambda_{kl} = c_1 \frac{2\sqrt{2}\sigma_n^2}{\delta_{kl}}, \quad \eta_{kl} = c_2 \frac{2\sqrt{2}\sigma_n^2}{\varrho_{kl}}. \quad (21)$$

In practice, standard deviations δ_{kl} and ϱ_{kl} for each cluster can be estimated from the sets of α_{ijl} and $(\alpha_{ijl} - \theta_{ijl})$ belonging to cluster C_k, respectively. We set $\gamma = 0.3$, $c_1 = 0.3$, and $c_2 = 0.7$ for the HSI denoising.

4. Experimental Results

We selected 6 HSIs[2] from the Columbia Multispectral Image Database [42] to test our method. To demonstrate the effectiveness of the proposed algorithm, we compare our denoising results with 8 recently developed state-of-art denoising methods, including band-wise BM3D [15] , band-wise centralized sparse representation (CSR) [20], adaptive non-local means denoising for volumetric data (ANLM) [45], non-local transform domain filter for volumetric data (BM4D) [24], low rank tensor approximation (LRTA) [30], spectral domain statistics based method (SDS) [22], 3D-cube K-SVD [46] and decomposable non-local tensor dictionary learning (DNTDL) [38]. All parameters involved in the competing algorithms were optimally assigned or automatically chosen as described in the reference papers. The denoising metrics (PSNR, SSIM, or FSIM) for each HSI are calculated as the average of the band-wise results. To show the influence of the number of the clusters on the denoising performance, we also implement our method with a single universal dictionary by setting the number of clusters to one, denoted as Ours-BL. In Tables 1 and 2, we provide the PSNR, SSIM, and FSIM results obtained from various methods for each HSI under different noise levels. We can see that the best result for each set is highlighted in bold. The proposed method outperforms other competitive approaches in most cases and achieves state-of-the-art denoising performance in terms of PSNR, SSIM, and FSIM. Comparing our method with Ours-BL, we see that our method can produce better results, which demonstrate that the proposed adaptive dictionary learning in Section 3.1, which captures the characteristics of the scene, can efficiently improve the denoising results.

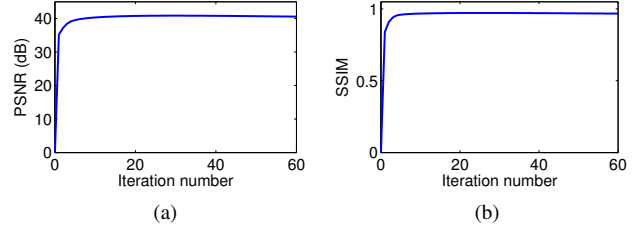[2]Due to space limitations, we choose 6 typical and complex scenes to test our method.

Figure 5. Verification of the convergence of the proposed method. Progression of the PSNR and SSIM for the *chart and stuffed toy* scene under $\sigma_n = 20$ levels of noise.

To further demonstrate the denoising performance of our method, we show two denoised hyperspectral bands from different methods under different levels of noise in Figs. 3 and 4. When the noise level is not very high (σ_n=20), as shown in Fig. 3, all the 8 competing methods achieve good denoising outputs, but when we enlarge the images and compare the details demarcated in Fig. 3 (a). We can see that our method recovers nice texture/edge features with rich details. When the noise level is high ($\sigma_n = 40$), as shown in Fig. 4, LRTA [30], SDS [22] and 3D-cube K-SVD [46] tend to generate many visual artifacts. BM3D [15], CSR [20], ANLM [45] and BM4D [24] are overly smooth and show many block artifacts. DNTDL [38] and our method work better in this case. In particular, the denoised image (Fig. 4(l)) by the proposed method has much less artifacts than other methods, and is visually more appealing, especially for the demarcated parts of the scene.

In addition, we also take the case of HSI denoising when $\sigma_n = 20$ for the *chart and stuffed toy* scene as an example to show the convergence of our method. Fig. 5 plots the evaluations of PSNR and SSIM versus iteration numbers for the tested HSI. We can see that both PSNR and SSIM curves increase monotonically and demonstrate the convergence of our method .

5. Conclusion

In this paper, we presented an effective HSI restoration method based on adaptive spatial-spectral dictionary learning, which considered the underlying characteristics of HSIs: the sparsity across the spatial-spectral domain, the high correlation across spectra, and non-local self-similarity over space. We exploited the high correlation across spectra and non-local self-similarity over space to learn the adaptive spatial-spectral dictionary for each overlapping cubic patch of the HSI. By design, each adaptive spatial-spectral dictionary could sparsely represent its own cubic patch across spatial and spectral domains while sharing features with other similar non-local patches across full HSI. Then, the local and non-local sparsity of the HSI was used to design a denoising model under the trained dictionary. This model was effectively solved by our numerical algorithm with its

Table 1. Denoising results (PSNR(dB)/SSIM/FSIM) of the 6 hyperspectral images from different methods, $\sigma_n = 20$.

Images	Metrics	BM3D [15]	CSR [20]	ANLM [45]	BM4D [24]	LRTA [30]	SDS [22]	K-SVD [46]	DNTDL [38]	Ours -BL	Ours
	PSNR	36.15	36.12	35.57	39.93	35.73	36.21	32.51	39.38	39.77	**40.79**
	SSIM	0.9397	0.9437	0.9497	0.9644	0.8932	0.8711	0.8002	0.9666	0.9625	**0.9718**
	FSIM	0.9723	0.9722	0.9747	0.9861	0.9599	0.9597	0.9603	0.9883	0.9875	**0.9905**
	PSNR	30.48	30.35	29.54	34.74	31.87	34.01	34.70	30.32	35.39	**35.91**
	SSIM	0.8404	0.8359	0.8033	0.9282	0.8376	0.9046	0.9323	0.8067	0.9370	**0.9407**
	FSIM	0.9518	0.9476	0.9342	0.9830	0.9639	0.9801	0.9844	0.9634	0.9836	**0.9870**
	PSNR	36.83	36.82	35.88	39.71	35.57	34.68	32.58	39.59	39.46	**40.23**
	SSIM	0.9384	0.9407	0.9435	0.9606	0.8759	0.8516	0.7947	**0.9646**	0.9520	0.9602
	FSIM	0.9685	0.9682	0.9720	0.9811	0.9521	0.9555	0.9589	0.9844	0.9839	**0.9868**
	PSNR	33.86	33.82	32.00	36.46	34.27	34.79	31.77	36.59	37.53	**37.93**
	SSIM	0.8679	0.8665	0.8031	0.9205	0.8774	0.8626	0.7652	0.9319	0.9373	**0.9421**
	FSIM	0.9554	0.9540	0.9389	0.9782	0.9593	0.9676	0.9522	0.9810	0.9822	**0.9852**
	PSNR	36.39	36.30	35.68	38.32	36.08	33.82	32.66	37.73	38.11	**39.07**
	SSIM	0.9106	0.9090	0.9195	0.9367	0.8898	0.8749	0.7949	0.9323	0.9317	**0.9334**
	FSIM	0.9643	0.9641	0.9712	0.9789	0.9643	0.9734	0.9616	0.9819	0.9792	**0.9848**
	PSNR	33.52	33.41	32.59	36.85	34.34	34.92	31.47	36.58	37.23	**38.07**
	SSIM	0.9118	0.9100	0.9127	0.9496	0.8940	0.9021	0.8164	0.9488	0.9530	**0.9605**
	FSIM	0.9651	0.9643	0.9648	0.9846	0.9730	0.9780	0.9675	0.9863	0.9852	**0.9890**

Table 2. Denoising results (PSNR(dB)/SSIM/FSIM) of the 6 hyperspectral images from different methods, $\sigma_n = 40$.

Images	Metrics	BM3D [15]	CSR [20]	ANLM [45]	BM4D [24]	LRTA [30]	SDS [22]	K-SVD [46]	DNTDL [38]	Ours -BL	Ours
	PSNR	32.21	32.14	31.51	35.82	32.28	29.57	29.02	35.30	35.52	**36.57**
	SSIM	0.8778	0.8900	0.8983	0.9105	0.8112	0.5617	0.7157	0.9334	0.9172	**0.9372**
	FSIM	0.9382	0.9404	0.9513	0.9670	0.9223	0.8737	0.9200	0.9717	0.9695	**0.9760**
	PSNR	27.06	26.97	26.99	30.94	28.32	27.37	25.90	31.05	31.22	**32.03**
	SSIM	0.6968	0.6926	0.6990	0.8483	0.7077	0.6649	0.6041	0.8593	0.8638	**0.8839**
	FSIM	0.8833	0.8718	0.8775	0.9553	0.9146	0.9406	0.8564	0.9634	0.9537	**0.9635**
	PSNR	33.03	33.17	32.62	35.82	32.16	29.28	29.25	35.96	35.55	**36.38**
	SSIM	0.8801	0.8951	0.9032	0.9106	0.7883	0.5616	0.7185	**0.9370**	0.9032	0.9233
	FSIM	0.9358	0.9384	0.9509	0.9580	0.9128	0.8718	0.9203	0.9674	0.9643	**0.9696**
	PSNR	30.93	31.11	30.48	33.07	31.40	28.59	30.49	33.30	34.23	**34.81**
	SSIM	0.7760	0.7881	0.7501	0.8417	0.8081	0.5780	0.7501	0.8761	0.8837	**0.8965**
	FSIM	0.9158	0.9134	0.9207	0.9527	0.9223	0.9023	0.9207	0.9563	0.9600	**0.9664**
	PSNR	33.11	33.43	33.59	35.22	32.62	28.08	29.63	34.97	34.58	**35.66**
	SSIM	0.8607	0.8745	0.8936	0.8961	0.8098	0.5547	0.7248	0.8984	0.8760	**0.9012**
	FSIM	0.9351	0.9392	0.9583	0.9569	0.9322	0.9152	0.9367	0.9650	0.9553	**0.9652**
	PSNR	29.93	29.83	29.57	32.32	30.87	28.48	28.06	32.89	33.30	**34.18**
	SSIM	0.8416	0.8473	0.8605	0.9087	0.8177	0.6230	0.7249	0.9069	0.9040	**0.9279**
	FSIM	0.9265	0.9232	0.9363	0.9658	0.9418	0.9253	0.9790	0.9663	0.9618	**0.9706**

regularization parameters adaptively adjusted for different clusters and different noise levels. The experimental results on HSI denoising showed that the proposed method could effectively denoise HSIs and provides significant improvement over many state-of-the-art methods.

References

[1] Y. Wang and R. Niu, "Hyperspectral urban remote sensing image smoothing and enhancement using forward-and-backward diffusion," in *Joint Urban Remote Sensing Event*, May 2009, pp. 1–5. 1, 2

[2] N. Renard and S. Bourennane, "Improvement of target detection methods by multiway filtering," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2407–2417, Aug. 2008. 1

[3] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of classification for hyperspectral images based on tensor modeling," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 801–805, Oct. 2010. 1

[4] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, Jan. 2014. 1
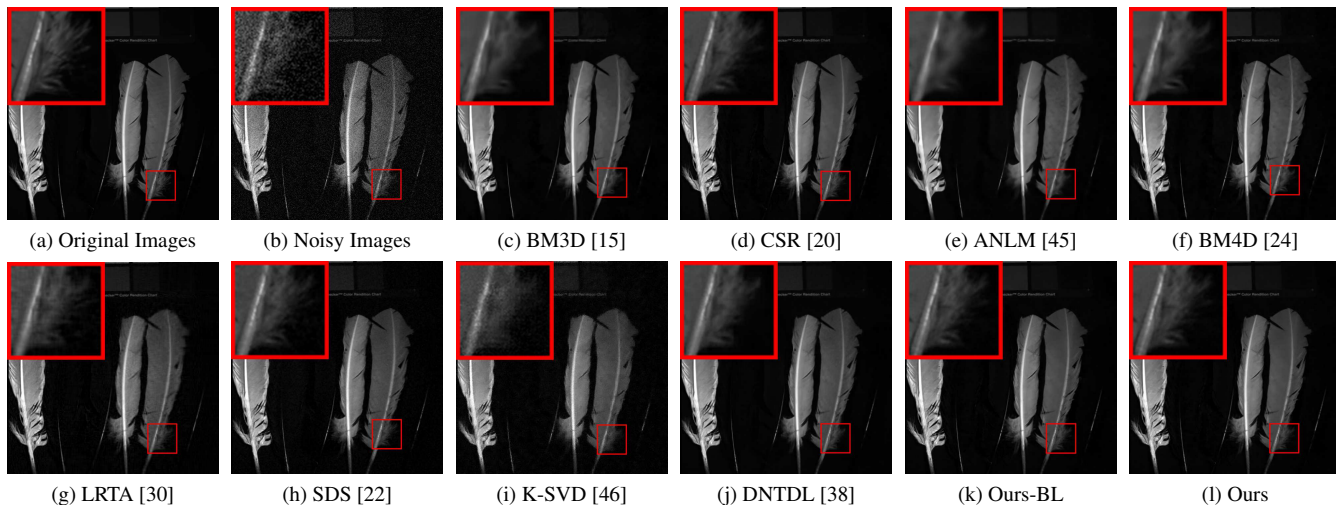
(a) Original Images   (b) Noisy Images   (c) BM3D [15]   (d) CSR [20]   (e) ANLM [45]   (f) BM4D [24]

(g) LRTA [30]   (h) SDS [22]   (i) K-SVD [46]   (j) DNTDL [38]   (k) Ours-BL   (l) Ours

Figure 3. Visual quality comparison for the HSI *feathers* under noise level $\sigma_n = 20$. The $620\ nm$ band is shown.

(a) Original Images   (b) Noisy Images   (c) BM3D [15]   (d) CSR [20]   (e) ANLM [45]   (f) BM4D [24]

(g) LRTA [30]   (h) SDS [22]   (i) K-SVD [46]   (j) DNTDL [38]   (k) Ours-BL   (l) Ours

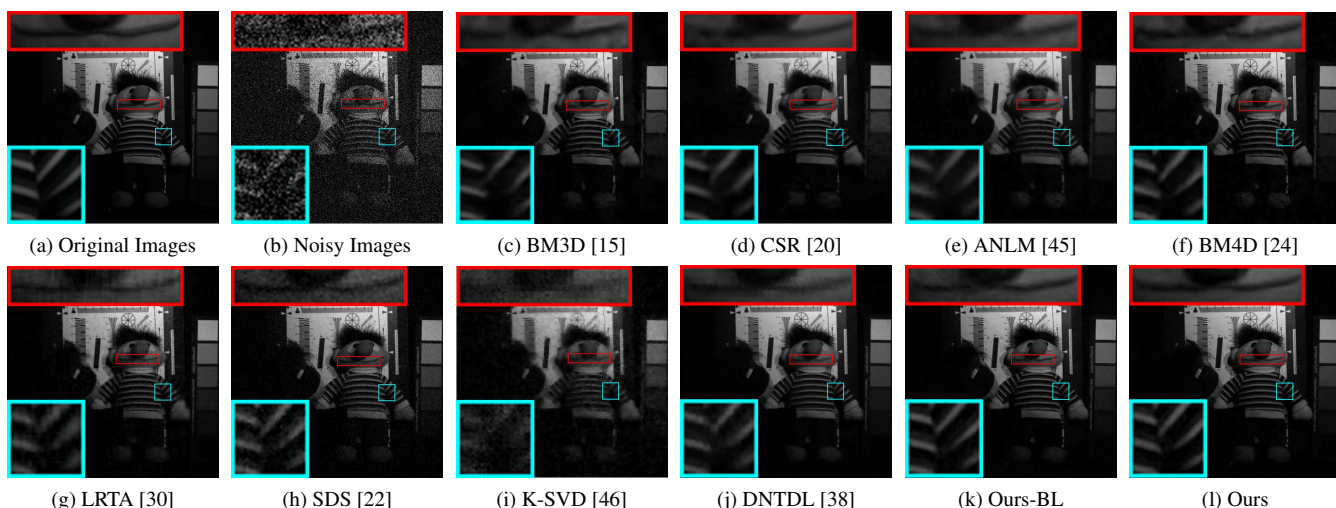Figure 4. Visual quality comparison for the HSI *chart and stuffed toy* under noise level $\sigma_n = 40$. The $600\ nm$ band is shown.

[5] A. Banerjee, P. Burlina, and J. Broadwater, "Hyperspectral video for illumination-invariant tracking," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHIS-PERS)*, Aug. 2009, pp. 1–4. 1

[6] R. Rowe, K. Nixon, and S. Corcoran, "Multispectral fingerprint biometrics," in *Proc. of IEEE Information Assurance Workshop (IAW)*, Jun. 2005, pp. 14–20. 1

[7] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006. 1

[8] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jan. 1998. 1

[9] M. D. Ingrid Daubechies, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math*, vol. 57, no. 11, pp. 1413–1457, 2004. 1

[10] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math*, vol. 59, pp. 797–829, 2004. 1

[11] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996. 1, 2

[12] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006. 1, 2, 3, 4

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. of International Conference on Machine Learning (ICML)*, 2009, pp. 689–696. 1

[14] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, Jun. 2005, pp. 60–65. 1, 2, 4

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007. 1, 2, 6, 7, 8

[16] A. Buades, B. Coll, J. M. Morel, and C. Sbert, "Self-similarity driven color demosaicking," *IEEE Trans. Image Processing*, vol. 18, no. 6, pp. 1192–1202, Jun. 2009. 1

[17] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. of International Conference on Computer Vision (ICCV)*, Sep. 2009, pp. 349–356. 1

[18] C.-Y. Yang, J.-B. Huang, and M.-H. Yang, "Exploiting self-similarities for single frame super-resolution," in *Proc. of Asian Conference on Computer Vision (ACCV)*, 2011, pp. 497–510. 1

[19] M. Shimano, T. Okabe, I. Sato, and Y. Sato, "Video temporal super-resolution based on self-similarity," in *Proc. of Asian Conference on Computer Vision (ACCV)*, 2011, pp. 93–106. 1

[20] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 457–464. 1, 3, 5, 6, 7, 8

[21] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. of International Conference on Computer Vision (ICCV)*, Nov. 2011, pp. 1259–1266. 1, 3

[22] A. Lam, I. Sato, and Y. Sato, "Denoising hyperspectral images using spectral domain statistics," in *Proc. of International Conference on Pattern Recognition (ICPR)*, 2012, pp. 477–480. 1, 2, 6, 7, 8

[23] Y. Qian, Y. Shen, M. Ye, and Q. Wang, "3-d nonlocal means filter with noise estimation for hyperspectral imagery denoising," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2012, pp. 1345–1348. 2, 4

[24] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Processing*, vol. 22, no. 1, pp. 119–133, Jan. 2013. 2, 6, 7, 8

[25] I. Atkinson, F. Kamalabadi, and D. Jones, "Wavelet-based hyperspectral image estimation," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2003, pp. 743–745 vol.2. 2

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

[27] S.-E. Q. Guangyi Chen, "Denoising and dimensionality reduction of hyperspectral imagery using wavelet packets, neighbour shrinking and principal component analysis," *International Journal of Remote Sensing*, vol. 30, no. 18, pp. 4889–4895, 2009. 2

[28] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 973–980, 2011. 2

[29] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008. 2

[30] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 138–142, Apr. 2008. 2, 6, 7, 8

[31] A. Karami, M. Yazdi, and A. Zolghadre Asli, "Noise reduction of hyperspectral images using kernel non-negative tucker decomposition," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 487–493, Jun. 2011. 2

[32] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 83, pp. 50–63, Sep. 2013. 2

[33] Y. Wang, R. Niu, and X. Yu, "Anisotropic diffusion for hyperspectral imagery enhancement," *IEEE Sensors Journal*, vol. 10, no. 3, pp. 469–477, Mar. 2010. 2

[34] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, 2012. 2

[35] P. Zhong and R. Wang, "Multiple-spectral-band CRFs for denoising junk bands of hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2260–2275, Apr. 2013. 2

[36] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014. 2

[37] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 499–515, 2013. 2, 3, 4

[38] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 2, 3, 6, 7, 8

[39] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997. 2

[40] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 193–200. 3

[41] M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Research & Application*, vol. 19, no. 1, pp. 4–9, Feb. 1994. 3

[42] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum," Tech. Rep., Nov 2008. 3, 6

[43] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Nonlocal sparse models for image restoration," in *Proc. of International Conference on Computer Vision (ICCV)*, Sep. 2009, pp. 2272–2279. 4

[44] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Processing*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013. 4, 5

[45] J. V. Manjn, P. Coup, L. Mart-Bonmat, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *Journal of magnetic resonance imaging JMRI*, vol. 31, no. 1, pp. 192–203, Jan. 2010. 6, 7, 8

[46] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Image Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006. 6, 7, 8

[47] L. Sendur and I. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Image Processing*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002. 5