

Low-rank Bayesian Tensor Factorization for Hyperspectral Image Denoising

Kaixuan Wei, Ying Fu*

School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

Abstract

In this paper, we present a low-rank Bayesian tensor factorization approach for hyperspectral image (HSI) denoising problem, where zero-mean white and homogeneous Gaussian additive noise is removed from a given HSI. The approach is based on two intrinsic properties underlying a HSI, i.e., the global correlation along spectrum (GCS) and nonlocal self-similarity across space (NSS). We first adaptively construct the patch-based tensor representation for the HSI to extract the NSS knowledge while preserving the property of GCS. Then, we employ the low rank property in this representation to design a hierarchical probabilistic model based on Bayesian tensor factorization to capture the inherent spatial-spectral correlation of HSI, which can be effectively solved under the variational Bayesian framework. Furthermore, through incorporating these two procedures in an iterative manner, we build an effective HSI denoising model to recover HSI from its corruption. This leads to a state-of-the-art denoising performance, consistently surpassing recently published leading HSI denoising methods in terms of both comprehensive quantitative assessments and subjective visual quality.

Keywords: Hyperspectral image denoising, full Bayesian CP factorization, nonlocal self-similarity, global correlation along spectrum, variational Bayesian inference, tensor rank auto determination.

1. Introduction

Hyperspectral image (HSI) is made up of massive contiguous wavebands for each spatial position of real scenes and provides much richer information about scenes than multiple/RGB images. It has been widely used for remote sensing, including mineral identification [1, 2], land cover classification [3], vegetation studies [4], and atmospheric studies [5]. Besides, in the computer vision field, the availability of detailed physical representation of HSI has been substantiated to significantly enhance the performance of numerous computer vision tasks, such as inpainting [6], tracking [7], unmixing [8], super-resolution [9], and face recognition [10].

However, in real cases, a HSI is always corrupted by noise, which severely degrades the quality of the imagery, and negatively impacts all subsequent HSI processing tasks aforementioned. Noise is inevitable during the acquisition, and caused at different stages in both the optics and photodetector [11]. Therefore, HSI denoising plays a vital role in the typical workflow of HSI analysis and processing.

From our observations of several state-of-the-art HSI denoising methods [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22], we find that the essence of successful design of HSI denoising algorithm is to reasonably extract useful prior structures knowledge underlying a HSI. The most commonly employed prior structures for HSI recovery including its global correlation along spectrum (GCS) and nonlocal self-similarity across space

(NSS). More specific, the GCS prior denotes a huge amount of redundancy across the spectral dimension. The high correlation can be observed among images located in adjacent bands of a HSI generally. And the NSS prior indicates the enhancement of sparsity can be achieved by grouping similar images fragments (i.e. blocks), which can further improve the performance of various HSI recovery methods [6, 16, 20].

Since the traditional 2D image denoising is a well studied yet still active topic, the simplest way of denoising a HSI is to apply these off-the-shelf techniques [23, 24, 25, 26, 27] band by band. However, this kind of coarsely extended methods ignores the GCS prior completely, which leads to a relatively low-quality result. To address this issue, carefully designed extension of several high performance 2D image denoising methods was proposed recently [12, 15, 14]. One notable example is a nonlocal transform-domain filter, generally referred to BM4D [12], which is a non-trivial but straightforward extension of well-known image denoising method BM3D [24]. Besides, a sparse representation based reconstruction method [15] which jointly utilize the global and local redundancy and correlation in spatial/spectral domain, is inspired by previous outstanding work 3D cubic K-SVD [23]. Similarly, a spectral/spatial adaptive hyperspectral total variance (SSAHTV) denoising algorithm [14], in which the spectral noise differences and spatial information differences are both considered in the process of noise reduction, is a HSI-oriented variation of the spatially adaptive TV (SATV) model [28].

Instead of constructing method directly from existed one, low-rank matrix recovery (LRMR) techniques are employed for HSI denoising, including convex relaxation based approaches

*Corresponding author

Email address: kaixuan_wei@outlook.com, fuying@bit.edu.cn
(Kaixuan Wei, Ying Fu)

[29, 30, 31, 32] and Bayesian inference approaches [33, 34, 35, 36, 37]. For example, Zhang *et al.* [19] directly adopt the Go Decomposition (GoDec) algorithm [32], which is a optimization algorithm aimed to solve LRMR model, to estimate the low-rank HSI patch. Chen *et al.* [38] employed a modified version of robust principal component analysis (RPCA) algorithm which models noise with mixture of Gaussian in the context of LRMR model, originally proposed in [37] to deal with HSI with non-i.i.d. noise structure. Recently, He *et al.* [21] proposed a local matrix recovery method while considering the global spatial-spectral smoothness through TV regularization, which achieves great performance especially in complex noise removal of hyperspectral remote sensing images.

Albeit these LRMR-based approaches are effective to certain HSI denoising cases, but they only consider the GCS prior knowledge. Since the HSI can be naturally represented as a 3D tensor instead of a 2D matrix, one obvious extension of these LRMR-based approaches is to facilitate the power of tensor decomposition [39, 40] which attracts growing attention recent years. The two popular tensor factorization framework, named as Tucker and CANDECOMP/PARAFAC (CP), are used to denoise HSI and permit to appreciate the denoising efficiency respectively[41, 17]. Nevertheless, the input tensor of their methods is just the original form of the HSI, including two spatial dimensions and one spectral dimension. This definitely degrades their denoising performance, since the two separated spatial dimensions of the HSI are generally not low-rank representable. This intuition can be easily comprehended by Figure 1, which exhibits typical output bands by directly applying LRTF method and our integrated model. It can be seen in the LRTF recovery (top right corner) that some vertical or horizontal artifact textures are introduced due to the unsuitable assumption that the two spatial dimensions of a HSI are low-rank representable.

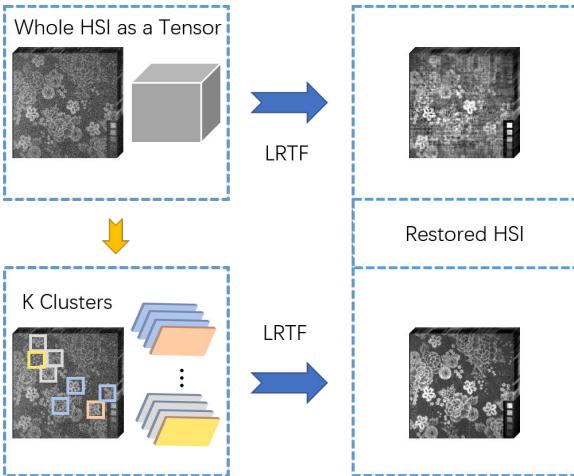


Figure 1: Typical output bands by the LRTF based methods, which show it is unreasonable to directly apply LRTF based methods on HSI denoising.

This property of HSI limits the way to straightly apply the low-rank tensor factorization (LRTF) based methods. To alleviate this problem, one question has been raised naturally: Can we transfer the initial form of a HSI into a more low-rank rep-

resentable one without destroying the spatial and spectral structure informations? Fortunately, this issue can be efficiently addressed by a remarkable technique called block matching, originally introduced by the benchmark BM3D [24] for image denoising. There are two approaches to extend the block matching strategy into volumetric data in general. One approach is to impose the NSS prior on the spectral dimension which further stacks the cubes of similar voxels into 4D "groups" respectively [12]. Another approach, which considers the GCS prior, is to simply extend a normal 2-D patch to a full band patch (FBP), then follows the same block matching procedure as the BM3D [16, 13]. Given the spectral property of a HSI, the FBP based approach is more efficient than the voxel based approach while preserving effectiveness.

Once the FBP clusters, which can be viewed as a set of 3D tensors, have been constructed using the extended block matching strategy, the family of tensor decomposition methods can be employed to maximize the use of underlying knowledge including these aforementioned priors. Xie *et al.* built a new HSI denoising model ITS-Reg[13] by applying LRTF based method on tensors formed by nonlocal similar patches within the HSI. Zhuang *et al.* [22] combined the power of LRMR and LRTF based techniques to construct a global local factorization model called GLF. However, as the noise level varies drastically, these convex relaxation based low rank approximation approaches are prone to overfitting due to the incorrect specified regularization parameters, resulting in severe deterioration of recovery performance. It is also worth noting that the rank minimization based on convex optimization of nuclear norm is affected by the tuning parameters, which may tend to over/underestimate the HSI.

In this paper, we present a hierarchical probabilistic model for HSI denoising based on full Bayesian CP tensor factorization (LBTF)[42], which can not only fit the underlying noise adaptively without knowing the specific noise intensity, but also determine the tensor rank automatically to address the overfitting issues. We first adaptively transform the original HSI into patch-based tensor representations (clusters) to extract the NSS knowledge while preserving the property of GCS in these new representations. Then we regard each cluster as a low-rank noisy observation in our hierarchical probability model in order to obtain the inherent spatial-spectral correlation of HSI. This model is effectively solved by an elegant deterministic algorithm based on variational inference. The empirical study demonstrates the superiority of our method, which consistently outperforms other state-of-the-art HSI denoising methods both quantitatively and visually.

The rest of this paper is organized as follows. Section 2 presents preliminary multilinear operators and notations. Section 3 introduces our Bayesian tensor factorization approach for HSI denoising. The extensive experimental results on both synthetic and real data are presented in Section 4, followed by conclusions in Section 5.

2. Notions and Preliminaries

The order of a tensor is the number of dimensions (a.k.a ways or modes). For clarification, scalars (zero order tensors) are denoted by lowercase letters, e.g., a . Vectors (first order tensors) are denoted by boldface lowercase letters, e.g., \mathbf{a} . Matrices (second order tensors) are denoted by boldface capital letters, e.g., \mathbf{A} . And the k -th row or column vector are denoted by \mathbf{A}_k , $\mathbf{A}_{\cdot k}$ respectively. General tensors (without order constraint) are denoted by boldface calligraphic letters, e.g., \mathcal{A} . Given a N -th order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, its (i_1, \dots, i_N) -th entry is denoted by $\mathcal{A}_{i_1 \dots i_N}$ without boldface.

The denotations of tensor operations follow [40]. The Kronecker product of matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{M \times N}$ is a matrix of size $IM \times JN$, denoted by $\mathbf{A} \otimes \mathbf{B}$. The Khatri-Rao product of matrices $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$ is a matrix of size $IJ \times K$ defined by a columnwise Kronecker product, and denoted by $\mathbf{A} \odot \mathbf{B}$. For convenience, the Khatri-Rao product of a set of matrices $\{\mathbf{A}^{(n)} | n = 1, \dots, N\}$ in a reverse order is defined by

$$\bigodot_n \mathbf{A}^{(n)} = \mathbf{A}^{(N)} \odot \mathbf{A}^{(N-1)} \odot \dots \odot \mathbf{A}^{(1)} \quad (1)$$

while the Khatri-Rao product of a set of matrices except the k -th matrix, is defined by

$$\bigodot_{n \neq k} \mathbf{A}^{(n)} = \mathbf{A}^{(N)} \odot \dots \odot \mathbf{A}^{(k+1)} \odot \mathbf{A}^{(k-1)} \dots \odot \mathbf{A}^{(1)} \quad (2)$$

The Hadamard product (a.k.a the Schur product or the entrywise product) of two tensors $\{\mathcal{A}, \mathcal{B}\} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ is denoted by $\mathcal{A} \circ \mathcal{B}$. It is a tensor with same dimensions as $\{\mathcal{A}, \mathcal{B}\}$ where each element indexed by $i_1 i_2 \dots i_N$ is the product of elements indexed by $i_1 i_2 \dots i_N$ of the original two tensors.

The inner product of two tensors $\{\mathcal{A}, \mathcal{B}\}$ is defined by $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1, \dots, i_N} \mathcal{A}_{i_1, \dots, i_N} \mathcal{B}_{i_1, \dots, i_N}$. For a more general case, we define

$$\langle \mathcal{A}^{(1)}, \dots, \mathcal{A}^{(N)} \rangle = \sum_{i_1, \dots, i_N} \prod_n \mathcal{A}_{i_1, \dots, i_N}^{(n)} \quad (3)$$

Our framework for LRTF is based on the CP decomposition which can be viewed as a higher-order generalization of the widely used matrix singular value decomposition (SVD). Given a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, it can be exactly factorized by a CP model, giving by

$$\mathcal{X} = \sum_{r=1}^R \mathbf{A}_{\cdot r}^{(1)} \odot \dots \odot \mathbf{A}_{\cdot r}^{(N)} = [\![\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]\!] \quad (4)$$

where \odot denotes the outer product of vectors, $[\![\dots]\!]$ denotes a Kruskal operator of a set of matrices having the same number of columns, $\mathbf{A}^{(n)}$ is a mode- n factor matrix of size $I_n \times R$ and R is assumed to be the upper bound the rank of tensor \mathcal{X} . Then each element of the tensor can be described as

$$\mathcal{X}_{i_1, \dots, i_N} = \sum_{r=1}^R A_{i_1 r}^{(1)} A_{i_2 r}^{(2)} \dots A_{i_N r}^{(N)} \quad (5)$$

3. The Proposed Method

3.1. Formulation

The HSI can be mathematically viewed as a 3D tensor. Given the additive noise, the noisy observation \mathbf{Y} can be described as

$$\mathbf{Y} = \mathcal{X} + \boldsymbol{\epsilon} \quad (6)$$

where \mathcal{X} is the original, unknown clean HSI, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau^{-1})$ is the additive noise. $\mathbf{Y} \in \mathbb{R}^{H \times W \times B}$, where H, W, B are stand for the corresponding spatial height, spatial width and the number of spectral band of this HSI respectively. \mathcal{X} and $\boldsymbol{\epsilon}$ have the same dimension with \mathbf{Y} . In this paper, we mainly consider the independent and identically distributed (i.i.d.) Gaussian noise with unknown noise intensity.

For clarification and simplicity, we denote the proposed low-rank Bayesian tensor factorization technique as LBTF, and the whole HSI denoising algorithm as LBTF-HSI. The objective of the proposed LBTF-HSI is to provide an estimation $\hat{\mathcal{X}}$ of the original \mathcal{X} from the noisy observation \mathbf{Y} , so that $\hat{\mathcal{X}}$ should be similar to \mathcal{X} as much as possible under a commonly adopted error measure (e.g. l_2 norm).

3.2. Iterative Denoising Framework

The LBTF-HSI is implemented in an iterative denoising framework motivated by [24], generally consisted of one to five near duplicated stages. Each stage comprises three steps: grouping, low-rank tensor recovery, and aggregation. The grouping and aggregation steps are required by the block matching technique, and the low-rank tensor recovery is what the step we apply the proposed LBTF algorithm. The flow-diagram of the LBTF-HSI implementation is illustrated in Figure 2.

In the grouping step, we first separate the noisy HSI \mathbf{Y} into a set of FBPs with overlap. Then for each local FBP, we construct a FBP cluster by performing block matching. To be specific, given a reference FBP, the P_n most similar FBP over all FBPs will be matched then form together into a FBP cluster, where P_n denotes the number of patches in a cluster, and the similarity is measured in a l_2 norm form for simplicity. Note this procedure is highly related to the GCS and NSS prior, and after this operation, both GCS and NSS knowledge are well preserved and reflected by such new representation, along its spectral and nonlocal-patch-number modes, respectively.

In the recovery step, we first initialize two zero-entries tensors denoted by \mathbf{C} and \mathbf{W} with the same size of the noisy observation \mathbf{Y} , then straightly apply the LBTF technique on each of FBP clusters to exploit the intrinsic low-rank property underlying this new representation. After we reconstruct all clean FBP clusters, we map every estimation to \mathbf{C} like the original form \mathbf{Y} using a cumulative scheme. It also does deserve to be noticed that \mathbf{W} can be explained as weight with respect to the cumulative \mathbf{C} , which is obtained during the calculation process of \mathbf{C} .

In the aggregation step, like a inverse Hadamard product, we simply divide the cumulative \mathbf{C} by its corresponding weight \mathbf{W} elementwise to generate an estimate $\hat{\mathcal{X}}$ of the original \mathcal{X} . To regard this estimate $\hat{\mathcal{X}}$ as a regularization, we can construct

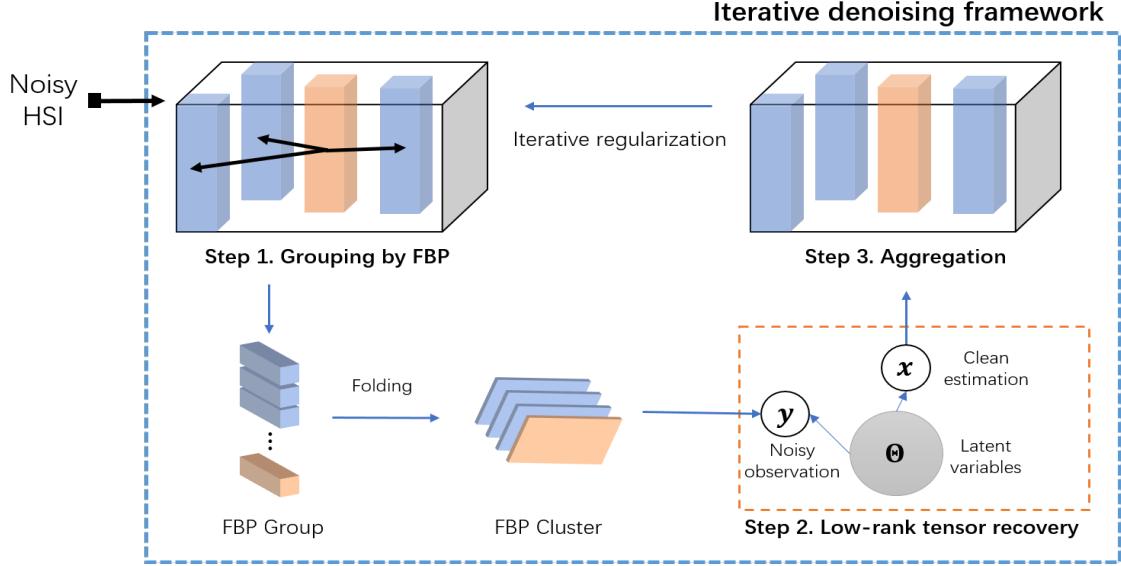


Figure 2: Flowchart of the proposed LBTF-HSI method. An noisy HSI is firstly splitted into a set of overlapping full bands patch, after grouping by each reference FBP (step 1), each FBP cluster is fed into the LBTF model to acquire its clean estimation (step 2), then each clean estimation is mapped onto the corresponding locations as it in the original noisy HSI by an accumulative scheme (step 3). Then either follow an iterative regularization to repeat this process or output the estimated HSI as a final result.

a updated noisy observation \mathbf{Y} by iterative regularization then repeat the same steps described above. This produces an iterative denoising framework. The whole detailed procedures of this framework are summarized in Algorithm 1.

Algorithm 1 Iterative Denoising Framework

Input: Noisy HSI \mathbf{Y} ;

- 1: Initialize $\hat{\mathbf{X}}^{(0)} = \mathbf{Y}$;
- 2: **for** $k = 1 : K$ **do**
- 3: Iterative regularization $\mathbf{Y}^{(k)} = \hat{\mathbf{X}}^{(k-1)} + \delta(\mathbf{Y} - \hat{\mathbf{X}}^{(k-1)})$;
- 4: Construct the entire FBP set Ω_k ;
- 5: Group matching FBP clusters $\{\mathbf{Y}_i\}_{i=1}^L$;
- 6: **for** Each FBP cluster \mathbf{Y}_i **do**
- 7: Recover \mathbf{X}_i from \mathbf{Y}_i by LBTF;
- 8: **end for**
- 9: Aggregate $\{\mathbf{X}_i\}_{i=1}^L$ to form the clean estimate $\hat{\mathbf{X}}^{(k)}$;
- 10: **end for**
- 11: Assign $\hat{\mathbf{X}} = \hat{\mathbf{X}}^{(K)}$;

Output: Estimation $\hat{\mathbf{X}}$ of \mathbf{X} ;

3.3. Hierarchical Probabilistic Model

Now we present the LBTF algorithm used in the recovery step, which is the key component of our method. In this step, each observation is a noisy FBP cluster formed by block matching. Given such noisy observation, we apply the LBTF to infer the underlying clean cluster.

From a Bayesian perspective, the CP tensor factorization can be formulated by a hierarchical probabilistic model which is actually an instance of probabilistic graphical model (PGM) [43].

The CP generative model in Equation (4) together with the observation model in Equation (6) directly gives rise to the following hierarchical probabilistic model. As we discussed in iterative denoising framework, after we acquire a set of FBP clusters $\{\mathbf{Y}_i\}_{i=1}^L$, for each cluster $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ (for brevity, the subscript i is omitted), its probability density can be derived through being factorized over tensor elements

$$p(\mathbf{Y} | \{A^{(n)}\}_{n=1}^3, \tau) = \prod_{i_1=1}^{I_1} \prod_{i_2=1}^{I_2} \prod_{i_3=1}^{I_3} \mathcal{N}(\mathbf{y}_{i_1 i_2 i_3} | \langle A_{i_1 \cdot}^{(1)}, A_{i_2 \cdot}^{(2)}, A_{i_3 \cdot}^{(3)} \rangle, \tau^{-1}) \quad (7)$$

where $A^{(n)}$ is the latent mode- n factor matrix of size $I_n \times R$, we note $n = 1, 2, 3$ for spatial, spectral, and nonlocal-patch-number modes respectively. $R_t \leq R$ denotes the ground-truth rank of tensor \mathbf{X} . $\mathcal{N}(y|\mu, \tau^{-1})$ denotes a Gaussian distribution of the form

$$\mathcal{N}(y|\mu, \tau^{-1}) = \left(\frac{\tau}{2\pi} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\tau}{2}(y - \mu)^2 \right\} \quad (8)$$

In order to further build our hierarchical probabilistic model, we need to enforce a suitable probabilistic structure on the underlying factor matrices $\{A^{(n)}\}_{n=1}^3$. From the CP model in Equation (4), notice that each outer product contributes at most one to the rank of \mathbf{X} . Since a low-rank estimation of \mathbf{X} is sought, our goal is to achieve column sparsity in $A^{(n)}$, such that most column in $A^{(n)}$ are set equal to zero. To enforce this constraint, we associate the columns of $A^{(n)}$ with Gaussian priors of precisions (inverse variances) λ_r , that is,

$$p(A^{(n)} | \lambda) = \prod_{i_n} \mathcal{N}(A_{i_n \cdot}^{(n)} | \mathbf{0}, \Lambda^{-1}), \quad \forall n \in [1, 3] \quad (9)$$

where $\Lambda = \text{diag}(\lambda)$ denotes an inverse variance matrix and is shared by latent factor matrices in all modes. Thus, the r -th columns of $\{\mathbf{A}^{(n)}\}_{n=1}^3$ have the same sparsity profile enforced by the common precisions λ_r . As shown later, many of the precisions λ_r will assume very large values during inference, which effectively removes the corresponding outer-products from \mathcal{X} , and hence reduces the rank of the estimation. We can further define a hyperprior over hyperparameter λ , which is factorized over latent dimensionality due to the independent assumption

$$p(\lambda) = \prod_{r=1}^R \text{Gam}(\lambda_r | c_0^r, d_0^r) \quad (10)$$

where $\text{Gam}(x|a, b)$ denotes a Gamma distribution of the form

$$\text{Gam}(x|a, b) = \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)} \quad (11)$$

$\Gamma(\cdot)$ is the Gamma function.

Using the similar technique, we also place a hyperprior over the noise precision τ , i.e.

$$p(\tau) = \text{Gam}(\tau | a_0, b_0) \quad (12)$$

Combining Equations (7) and (9) to (12) together, we can complete our hierarchical probabilistic model as a PGM, the whole graph representation is illustrated in Figure 3.

For brevity of notations, we denote all unknowns including both latent variable and hyperparameters by $\Theta = (\{\mathbf{A}^{(n)}\}_{n=1}^3, \lambda, \tau)$. From Figure 3, we can write the joint distribution of observed data and all model parameters as

$$p(\mathbf{Y}, \Theta) = p(\mathbf{Y}_\Omega | \{\mathbf{A}^{(n)}\}_{n=1}^3, \tau) \prod_{n=1}^3 p(\mathbf{A}^{(n)} | \lambda) p(\lambda) p(\tau). \quad (13)$$

The goal turns to infer the posterior of all involved parameters, which can be done by maximizing Equation (13) without loss of generality.

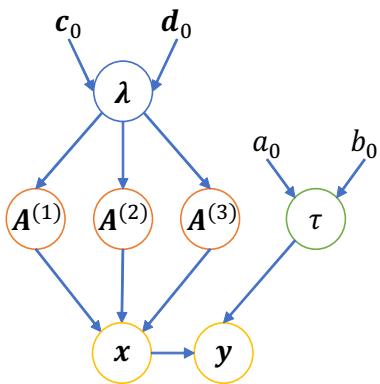


Figure 3: The probabilistic graphical model representation of Bayesian CP tensor factorization.

3.4. Variational Inference

However, in contrast to the point estimation, we aim to compute the full posterior distribution of all parameters in Θ .

Since that, a deterministic approximate inference method under the variational Bayesian (VB) framework [43] is developed to learn the aforementioned hierarchical probabilistic model. To achieve this goal, we therefore seek a distribution $q(\Theta)$ to approximate the true posterior distribution $p(\Theta|\mathbf{Y})$ by solving the following optimization problem

$$\min_q \text{KL}(q(\Theta) \| p(\Theta|\mathbf{Y})) = - \int q(\Theta) \ln \left\{ \frac{p(\Theta|\mathbf{Y})}{q(\Theta)} \right\} d\Theta \quad (14)$$

where $\text{KL}(q||p)$ represents the KL divergence between two distribution q and p . Since the posterior distribution $p(\Theta|\mathbf{Y})$ is computational intractable in our model, it makes our problem that cannot be reduced from the VB framework into the expectation maximization (EM) framework. Thus, some constraints need to be imposed on the variational distribution $q(\Theta)$ to make this optimization feasible. Specifically, it will be assumed that the variational distribution is factorized w.r.t. each parameter Θ_j , so that

$$q(\Theta) = q(\lambda) q(\tau) \prod_{n=1}^3 q(\mathbf{A}^{(n)}). \quad (15)$$

This factorized form of variational inference corresponds to an approximation framework developed in physics called mean field theory [44]. After that, the closed-form optimal solution $q_j^*(\Theta_j)$ can be obtained by

$$\ln q_j^*(\Theta_j) = \langle \ln p(\mathbf{Y}, \Theta) \rangle_{\Theta \setminus \Theta_j} + \text{const} \quad (16)$$

where $\langle \cdot \rangle$ is a unary operator denoting expectation and $\Theta \setminus \Theta_j$ denotes the set of Θ with Θ_j removed. Since the distributions of all variables are drawn from the distributions over their parent variables, we can analytically infer the posterior distributions of model parameters using Equations (13), (15) and (16).

Estimation of mode-n factors $\mathbf{A}^{(n)}$.

$$q^*(\mathbf{A}^{(n)}) = \prod_{i_n=1}^{I_n} \mathcal{N}\left(\mathbf{A}_{i_n}^{(n)} \mid \langle \mathbf{A}_{i_n}^{(n)} \rangle, \Sigma_{i_n}^{(n)}\right) \quad \forall n \in [1, 3] \quad (17)$$

where the posterior parameters can be updated by

$$\langle \mathbf{A}_{i_n}^{(n)} \rangle = \langle \tau \rangle \Sigma_{i_n}^{(n)} \langle \mathbf{B}^{(\setminus n)T} \rangle \text{vec}[\mathbf{Y}_{\cdot i_n \cdot}] \quad (18)$$

$$\Sigma_{i_n}^{(n)} = (\langle \tau \rangle \langle \mathbf{B}^{(\setminus n)T} \mathbf{B}^{(\setminus n)} \rangle + \langle \Lambda \rangle)^{-1} \quad (19)$$

$$\mathbf{B}^{(\setminus n)} = \bigodot_{k \neq n} \mathbf{A}^{(k)} \quad (20)$$

The most complex term is related to $\mathbf{B}^{(\setminus n)}$, which is of size $\prod_{k \neq n} I_k \times R$, and denotes the Khatri-Rao product of latent factors in all modes except nth-mode. $\text{vec}[\mathbf{Y}_{\cdot i_n \cdot}]$ denotes the vectorized FBP cluster of size $\prod_{k \neq n} I_k$, whose mode-n index is i_n .

Estimation of hyperparameters λ .

$$q^*(\lambda) = \prod_{r=1}^R \text{Gam}(\lambda_r | c^r, d^r) \quad (21)$$

where

$$c^r = c_0^r + \frac{1}{2} \sum_{n=1}^3 I_n \quad (22)$$

$$d^r = d_0^r + \frac{1}{2} \sum_{n=1}^3 \langle \mathbf{A}_{\cdot r}^{(n)T} \mathbf{A}_{\cdot r}^{(n)} \rangle \quad (23)$$

Estimation of hyperparameter τ .

$$q^*(\tau) = \text{Gam}(\lambda_r | a, b) \quad (24)$$

where

$$a = a_0 + \frac{1}{2} \prod_{n=1}^3 I_n \quad (25)$$

$$b = b_0 + \frac{1}{2} \left\langle \|\mathbf{Y} - [\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]\|_F^2 \right\rangle \quad (26)$$

Algorithm 2 Low-rank Bayesian Tensor Factorization

Input: A FBP cluster \mathbf{Y}_i ;

- 1: Initialize factor matrices and their covariance $\mathbf{A}_{i_n}^{(n)}, \Sigma_{i_n}^{(n)}$, hyperpriors $a_0, b_0, \mathbf{c}_0, \mathbf{d}_0$ and hyperparameters $\tau = \frac{a_0}{b_0}, \lambda_r = \frac{c_0^r}{d_0^r}$;
- 2: **while** not converge **do**
- 3: **for** $n = 1$ to 3 **do**
- 4: Update the posterior $q(\mathbf{A}^{(n)})$ using Equations (18) to (20);
- 5: **end for**
- 6: Update the posterior $q(\lambda)$ using Equations (22) and (23);
- 7: Update the posterior $q(\tau)$ using Equations (25) and (26);
- 8: Update the estimated Rank R by $\max_n \text{Rank}(\mathbf{A}^{(n)})$
- 9: **end while**

Output: Estimate FBP cluster $\hat{\mathbf{X}}_i$ and Rank R;

The whole procedure of model inference is summarized in Algorithm 2, It's worth noting that tensor rank is determined automatically and implicitly. To be specific, during inference, most of the hyperparameters λ_i are driven to very large values, which will force the posterior means of the columns to go to zero, effectively removing them from the model and reducing the rank. For implementation of the algorithm, we keep the size of $\{\mathbf{A}^{(n)}\}$ unchanged during iterations, while an alternative way is to remove the zero components of $\{\mathbf{A}^{(n)}\}$ after each iteration.

4. Experiment and Analysis

In this section, extensive simulated and real data experiments are conducted to validate the denoising capabilities of the proposed LBTF-HSI algorithm, and qualitative and visual results are illustrated. The detailed analysis about our method is presented in final.



Figure 4: Simulated pseudo color images from Columbia Dataset

4.1. Simulated HSI Denoising

Columbia Dataset. The Columbia HSI Dataset [46]¹ is employed in our simulated experiment, which is commonly used in other algorithms verification [13, 16]. This dataset consists of 32 real-world scenes of a wide variety of real-world materials and objects, with spatial resolution 512×512 and spectral bands 31. Each HSI includes full spectral resolution reflectance data collected from 400 nm to 700 nm with 10 nm interval. The simulated pseudo color images from this dataset are shown in Figure 4. In our experiments, the intensity of these HSIs is scaled into $[0, 1]$.

Implementation Details. Additive white Gaussian noise (AWGN), which comes from many natural sources, is added into these testing HSIs to generate \mathbf{Y} corresponding to our observation model with noise intensity ranging from 15 to 100 (It's need to be clarified we denote the noise intensity with a base 255, i.e. 15 means the standard deviation of Gaussian noise is $\frac{15}{255}$, similarly hereinafter). Unlike other methods, which require specific noise intensity as a input parameter, we do not feed this information into our method since the internal noise intensity can be automatically learned during its denoising process. Consequently, except particularly mentioned, we provide the real noise intensity to comparison methods while our method learns the noise model automatically.

For parameters setting, we need to care about the initialization strategy in LBTF (Algorithm 2). There are two parameters which are closely relevant to initialization. One is a binary parameter which can choose the low-rank components initialization scheme between SVD and random generation (follow a standard normal distribution). Though the theory of VB framework [43] can guarantee every initialized point converges to a local minimum, we find using random generation rather than SVD will achieve better performance in the context of HSI denoising. This phenomenon can be interpreted by grouping and aggregation operations involved in our method, which appreciate miscellaneous initialized points rather relatively stable ones. Another parameter which dominantly affects the denoising capability of LBTF is the upper bound of rank R of the low-rank components. It's worth noting that we only need to provide a roughly estimation of the upper bound of objective rank rather

¹<http://www1.cs.columbia.edu/CAVE/databases/multispectral>

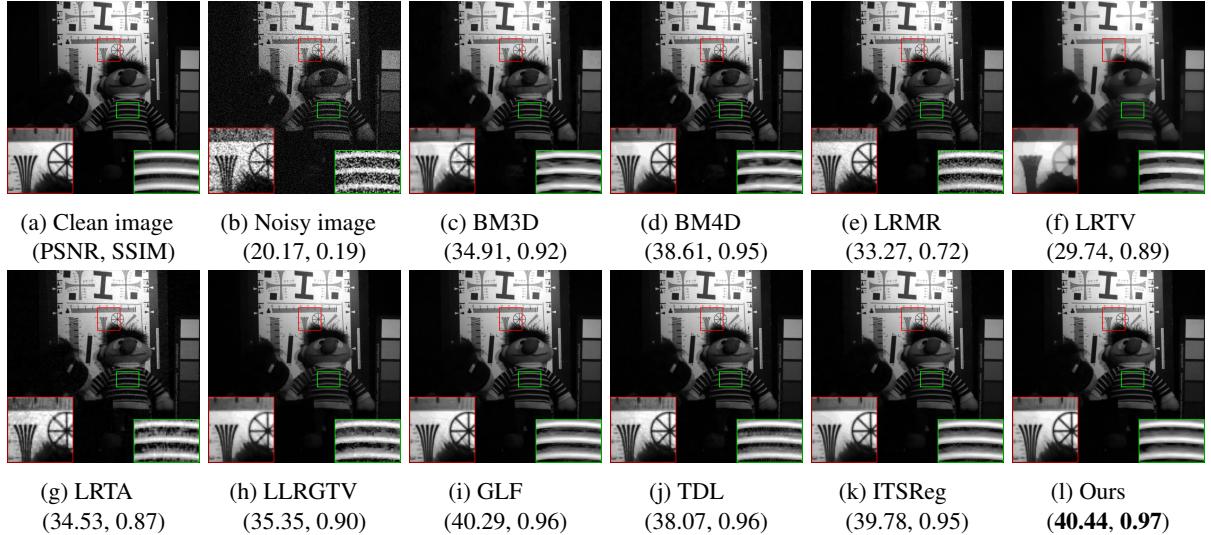


Figure 5: The images at band 590 nm of *chart and stuffed toy* under noise level $\sigma = 25$ on CAVE dataset. Two demarcated areas in each image are amplified at a 3 times larger scale for easy observation of details.

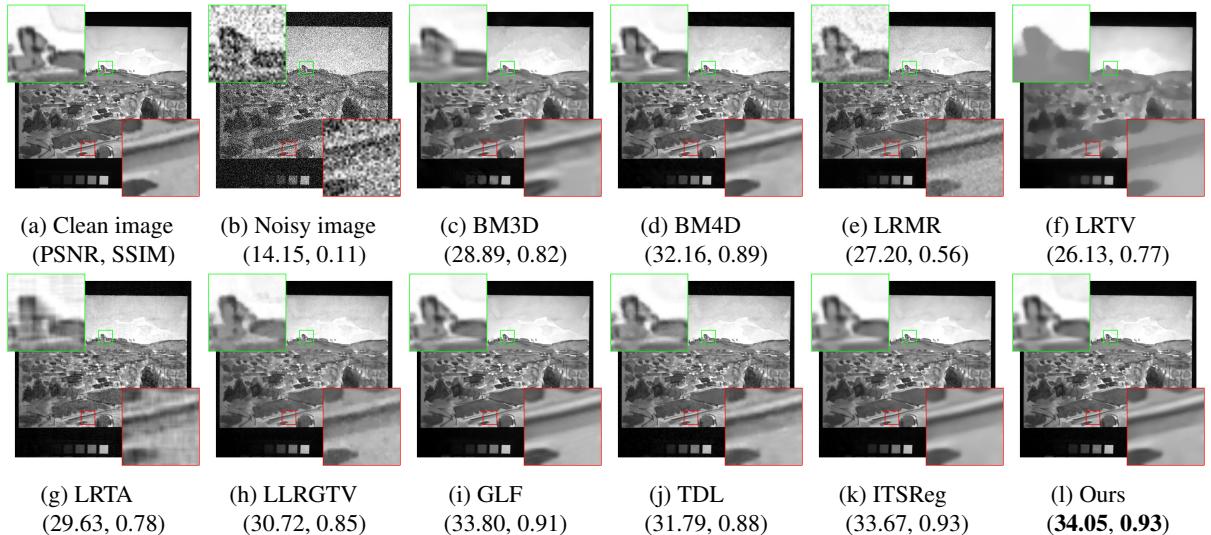


Figure 6: The images at band 490 nm of *watercolors* under noise level $\sigma = 50$ on CAVE dataset. Two demarcated areas in each image are amplified at a 6 times larger scale for easy observation of details.

than the indeed objective rank required by other low-rank based methods [19, 45]. After one iteration of our algorithm, the truth rank can be automatically estimated. We simply set R in the first iteration to 15, and keep track of mean of the truth rank of all clusters as R of the next iteration in all of our experiments.

Comparison Methods. The comparison methods include: band-wise BM3D [24]², which represents the state-of-the-art for the 2D extended band-wise approach; BM4D [12]², which represents state-of-the-arts for the 2D extended 3D-cube-based approach; LRMR [19], LRTV [45] and LLRGTV [21] which represent state-of-the-arts for the low-rank matrix-based approach; LRTA [41], GLF [22], TDL [16]³ and ITS-Reg [13]³, which

represent state-of-the-arts for the tensor-based approach. All parameters involved in the competing algorithms were manually tuned optimally or automatically chosen as described in the reference papers.

Performance Metrics. To comprehensively assess the performance of all competing methods, we employ five quantitative picture quality indices (PQI) for performance evaluation, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM [47]), feature similarity (FSIM [48]), erreur relative globale adimensionnelle de synthèse (ERGAS [49]) and spectral angle map (SAM [50]). PSNR and SSIM are two conventional PQIs in image processing and computer vision. They evaluate the similarity between the target image and reference image based on MSE and structural consistency, respectively. FSIM emphasizes the perceptual consistency with the reference

²<http://www.cs.tut.fi/foi/GCF-BM3D/>

³<http://gr.xjtu.edu.cn/web/dymeng/>

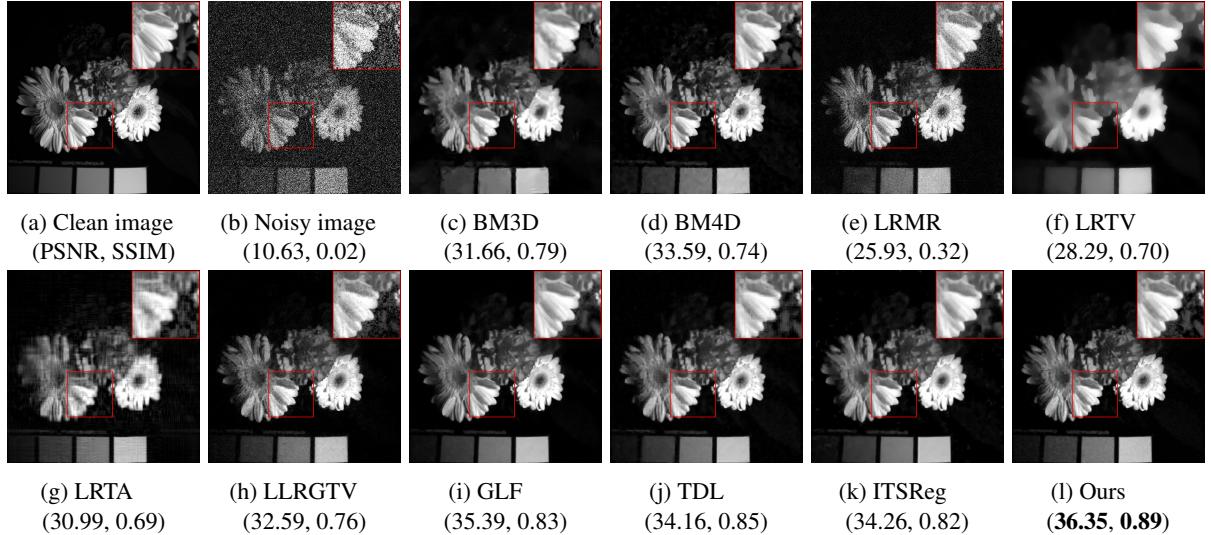


Figure 7: The images at band 640 nm of *flowers* under noise level $\sigma = 75$ on CAVE dataset. One demarcated areas in each image is amplified at a 1.5 times larger scale for easy observation of details.

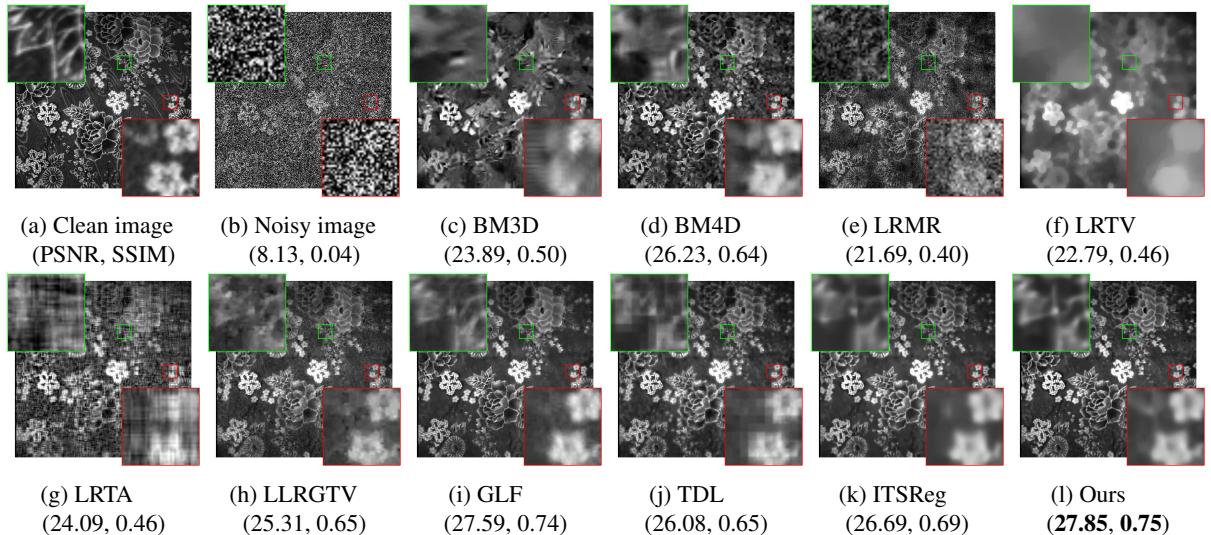


Figure 8: The images at band 590 nm of *cloth* under noise level $\sigma = 100$ on CAVE dataset. Two demarcated areas in each image are amplified at a 6 times larger scale for easy observation of details.

image. The larger these three measures are, the closer the target HSI is to the reference one. ERGAS measures fidelity of the restored image based on the weighted sum of MSE in each band. SAM measures the spectral fidelity between the restored image and the reference image across all spatial positions. Different from the former three measures, the smaller these two measures are, the better does the target HSI estimate the reference one.

Performance Evaluation. For each noise setting, all of the five PQI values for each competing HSI denoising methods on all 32 scenes have been calculated and recorded. Table 1 lists the average performance over different scenes under noise settings of all methods. From these quantitative comparison, the advantage of the proposed method can be evidently observed. Particularly, with the increase of noise intensity, our method surpasses the second best ITS-Reg under the measure of PSNR

by a large margin (e.g. 0.96 dB under $\sigma = 75$, 2.5dB under $\sigma = 75$). This is due to the overfitting issue commonly existed in state-of-the-art methods. Our method successfully address this issue by automatically determining the tensor rank, consequently achieving great performance especially in severe pollution case. Figures 5 to 8 illustrate the visual results of different methods under different noise levels. It can be seen that our method consistently outperform other methods as we measured in Table 1. Specifically, in Figure 6, we can see except GLF and our method, none of the competing methods can successfully recover the exact edge shape of cloud exhibited in the green box. In Figure 8, only GLF, ITS-Reg and our method produce smooth and noise-free results, while the fine-grained details of ours are much clearer and sharper than ITS-Reg's. We also compute the PSNR value of each bands in these four HSIs

Table 1: Average performance of 10 competing methods w.r.t. 5 PQIs. For each specific noise intensity setting, the results are obtained by averaging through the 32 scenes. The best results of each case among these methods are denoted by boldface.

Sigma	Index	Methods										
		Noisy	BM3D [24]	BM4D [12]	LRMR [19]	LRTV [45]	LRTA [41]	LLRGT [21]	GLF [22]	TDL [16]	ITSReg [13]	Ours
15	PSNR	24.61	39.81	42.38	37.21	33.54	39.21	38.46	43.41	42.30	43.43	43.46
	SSIM	0.291	0.951	0.968	0.869	0.912	0.930	0.948	0.977	0.972	0.972	0.976
	FSIM	0.794	0.973	0.981	0.974	0.938	0.971	0.978	0.989	0.987	0.989	0.988
	ERGAS	325.24	56.41	41.35	76.49	124.88	60.89	71.05	38.49	41.98	37.26	36.72
	SAM	0.785	0.157	0.151	0.391	0.204	0.183	0.175	0.128	0.101	0.138	0.103
25	PSNR	20.17	37.03	39.59	33.49	32.42	36.67	36.63	40.96	39.72	40.57	41.21
	SSIM	0.148	0.919	0.943	0.736	0.895	0.893	0.913	0.957	0.957	0.945	0.964
	FSIM	0.661	0.955	0.968	0.952	0.922	0.953	0.969	0.984	0.979	0.980	0.982
	ERGAS	542.05	77.50	57.16	115.39	136.84	81.21	84.89	50.50	56.39	51.45	48.26
	SAM	0.933	0.208	0.215	0.569	0.234	0.218	0.254	0.167	0.123	0.242	0.118
50	PSNR	14.15	33.49	35.65	28.35	29.82	33.16	33.45	37.15	36.16	37.55	37.83
	SSIM	0.052	0.862	0.870	0.470	0.846	0.819	0.812	0.890	0.918	0.919	0.927
	FSIM	0.465	0.922	0.938	0.890	0.891	0.919	0.944	0.970	0.956	0.963	0.966
	ERGAS	1084.15	116.60	90.13	204.78	183.40	120.99	118.41	77.24	84.58	72.85	71.07
	SAM	1.124	0.277	0.340	0.797	0.350	0.278	0.433	0.263	0.186	0.243	0.173
75	PSNR	10.63	31.36	33.28	25.27	27.98	31.17	31.28	34.75	34.08	34.78	35.74
	SSIM	0.026	0.810	0.794	0.310	0.787	0.762	0.716	0.812	0.875	0.881	0.889
	FSIM	0.362	0.894	0.908	0.826	0.870	0.892	0.921	0.957	0.934	0.945	0.951
	ERGAS	1626.14	147.89	118.14	290.62	224.31	152.38	149.57	101.80	107.73	100.36	90.47
	SAM	1.225	0.338	0.429	0.913	0.477	0.318	0.585	0.357	0.243	0.297	0.224
100	PSNR	8.13	29.83	31.56	23.03	26.50	29.69	29.64	33.03	32.56	31.77	34.26
	SSIM	0.015	0.767	0.723	0.214	0.751	0.712	0.635	0.747	0.826	0.835	0.855
	FSIM	0.299	0.871	0.879	0.766	0.853	0.869	0.899	0.944	0.911	0.914	0.938
	ERGAS	2168.26	175.21	143.73	375.29	267.94	180.21	178.72	123.92	128.06	143.74	107.08
	SAM	1.290	0.383	0.496	0.995	0.540	0.350	0.695	0.432	0.299	0.306	0.263

Table 2: Average performance of 10 competing methods w.r.t. 5 PQIs under unknown Gaussian noise level. The results are obtained by averaging through the 32 scenes. The best results of each case among these methods are denoted by boldface.

	PSNR	SSIM	FSIM	ERGAS	SAM
None	14.03 ± 4.62	0.079 ± 0.108	0.462 ± 0.197	1235.75 ± 613.62	1.124 ± 0.276
BM3D	33.36 ± 3.31	0.857 ± 0.052	0.919 ± 0.030	119.26 ± 35.93	0.292 ± 0.112
BM4D	35.73 ± 3.02	0.877 ± 0.055	0.934 ± 0.033	92.11 ± 27.73	0.320 ± 0.145
LRMR	29.35 ± 3.77	0.603 ± 0.170	0.893 ± 0.061	194.22 ± 71.11	0.610 ± 0.228
LRTV	29.38 ± 3.03	0.841 ± 0.072	0.894 ± 0.043	194.92 ± 63.90	0.361 ± 0.157
LRTA	33.34 ± 3.21	0.844 ± 0.065	0.924 ± 0.029	119.97 ± 37.56	0.236 ± 0.080
LLRGT	32.38 ± 3.14	0.783 ± 0.092	0.931 ± 0.031	135.66 ± 42.20	0.410 ± 0.200
GLF	36.88 ± 3.06	0.859 ± 0.086	0.967 ± 0.015	82.62 ± 27.99	0.287 ± 0.168
TDL	36.20 ± 3.09	0.915 ± 0.035	0.952 ± 0.022	85.57 ± 24.05	0.183 ± 0.085
ITSReg	37.17 ± 3.17	0.916 ± 0.042	0.959 ± 0.023	78.21 ± 25.31	0.218 ± 0.150
Ours	37.70 ± 2.98	0.924 ± 0.034	0.965 ± 0.017	73.20 ± 21.36	0.174 ± 0.084

(i.e. *watercolors*, *cloth*, etc.). It can be seen in Figure 9, the PSNR values of all bands obtained by LBTF-HSI are significantly higher than those of the other methods.

Denoising under Unknown Noise Level. Motivated by appealing noise intensity self-adaptive property aforementioned of our method, we conduct experiments under unknown Gaussian noise level for further demonstrating the advantages of the proposed method. Here, we still adopt 32 real-world scenes HSIs from

the Columbia Dataset described above. Unlike former experiment, which recurrently adds Gaussian noise with intensity from 15 to 100 into 32 clean HSIs to generate 160 corrupted HSIs, we only generate 32 corrupted HSIs with noise intensities randomly sampled from a uniform distribution of range [15, 100] in this experiment. Notice the true noise intensities are not provided, we use an off-the-shelf noise estimation method [51] to estimate it, which is set as the input parameter for all compared methods except ours. Table 2 summarizes the qualitative

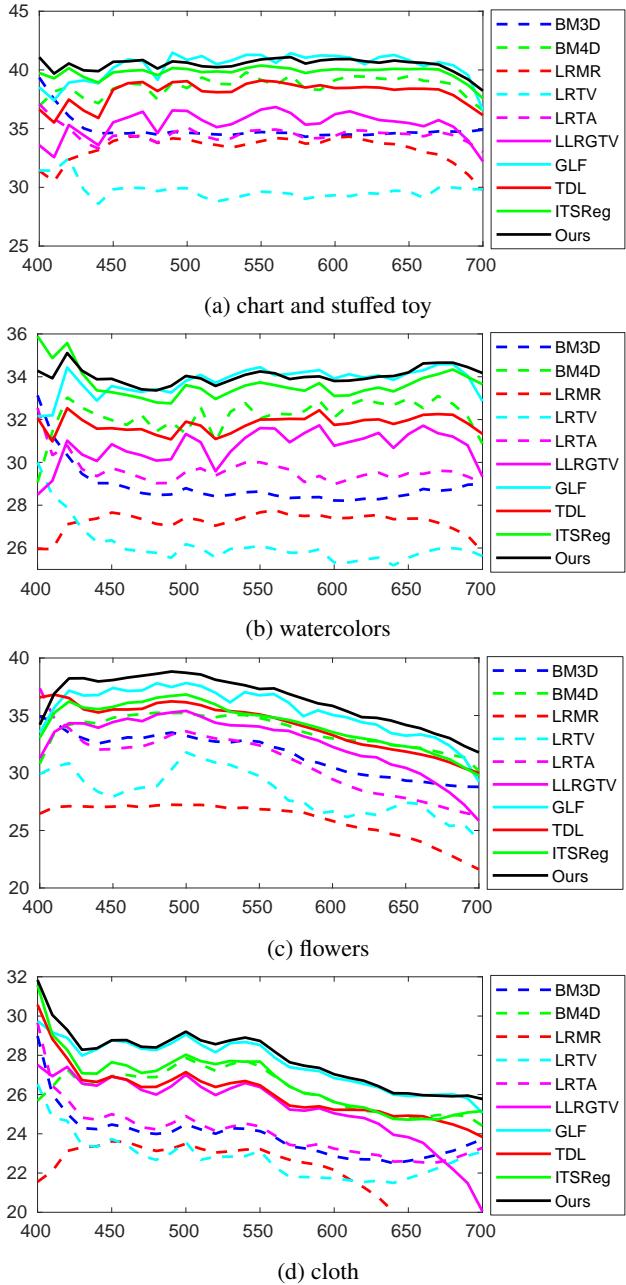


Figure 9: PSNR values across the spectrum corresponding to *chart and stuffed toy* (Fig. 5), *watercolors* (Fig. 6), *flowers* (Fig. 7) and *cloth* (Fig. 8) respectively.

results of this experiment, which shows our method surpasses about 0.48 dB than previous best-performance method ITS-Reg under the measure of PSNR while with the best stability (less variance) among all the competing methods.

Run Time. In addition to visual quality, another important aspect for an HSI denoising method is the run time. We then compare the speed of all competing methods. All experiments are run under the Matlab2016a environment on a machine with Intel(R) Core(TM) i7-7700K CPU of 4.2GHz and 16 GB RAM. Figure 10 shows the *Time v.s. PSNR* of different methods for denoising HSIs of size $512 \times 512 \times 31$. The results are obtained by

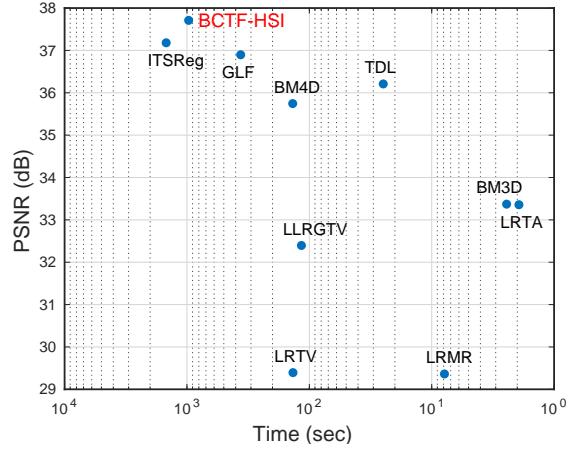


Figure 10: Time (second) v.s. PSNR (dB) of all competing method for HSI denoising.

averaging all 32 scenes with variety of noise intensity. We can see that effectiveness potentially often sacrifices efficiency. Our method is relatively slower than TDL, BM4D and GLF. However, taking the great enhancement in denoising effectiveness into account, our method is still highly completable with these two state-of-the-art methods. On the other hand, our method typically achieves 2 times speed even with better denoising capability compared with ITS-Reg.

4.2. Real HSI Denoising

Here, the Hyperspectral Digital Imagery Collection Experiment (HYDICE) urban dataset⁴ and the Harvard real-world hyperspectral datasets (HHD)[52] are utilized to evaluate our method in real-world noise context. The original HSI in HYDICE is of size $304 \times 304 \times 210$. As the bands 139-155, 201-210 are seriously polluted by the atmosphere and water absorption, and can provide little useful information, we manually remove them and leave the remaining test data with a size of $304 \times 304 \times 183$ like [13]. The whole HHD dataset consisting of 50 noisy hyperspectral images of size $1040 \times 1392 \times 31$ are captured with the wave-lengths in the range of 420-720 nm at an interval of 10. We scale these HSIs into the interval $[0, 1]$, and employ the similar implementation strategies and parameter settings for all competing methods as previous simulated experiments. Noise estimation method [51] used before is also applied in this setting. We illustrate the experimental results in Figure 11 and Figure 12 respectively.

Figure 11 includes the restorations of bands 1, 109 of the urban HSI. We finely choose two demarcated area with specific semantics to conveniently compare the denoising capability of different methods. Specifically, The red box area of band 1 represents the housing estate in urban area. It can be obviously observed that most of competing methods (e.g. BM3D, BM4D, LRTA, TDL, ITS-Reg) cannot remove the inappropriate stripes existed in this area, while some methods (i.e. LRM, LRTV) produce oversmooth results, in some degree destroying

⁴<http://www.tec.army.mil/hypercube>

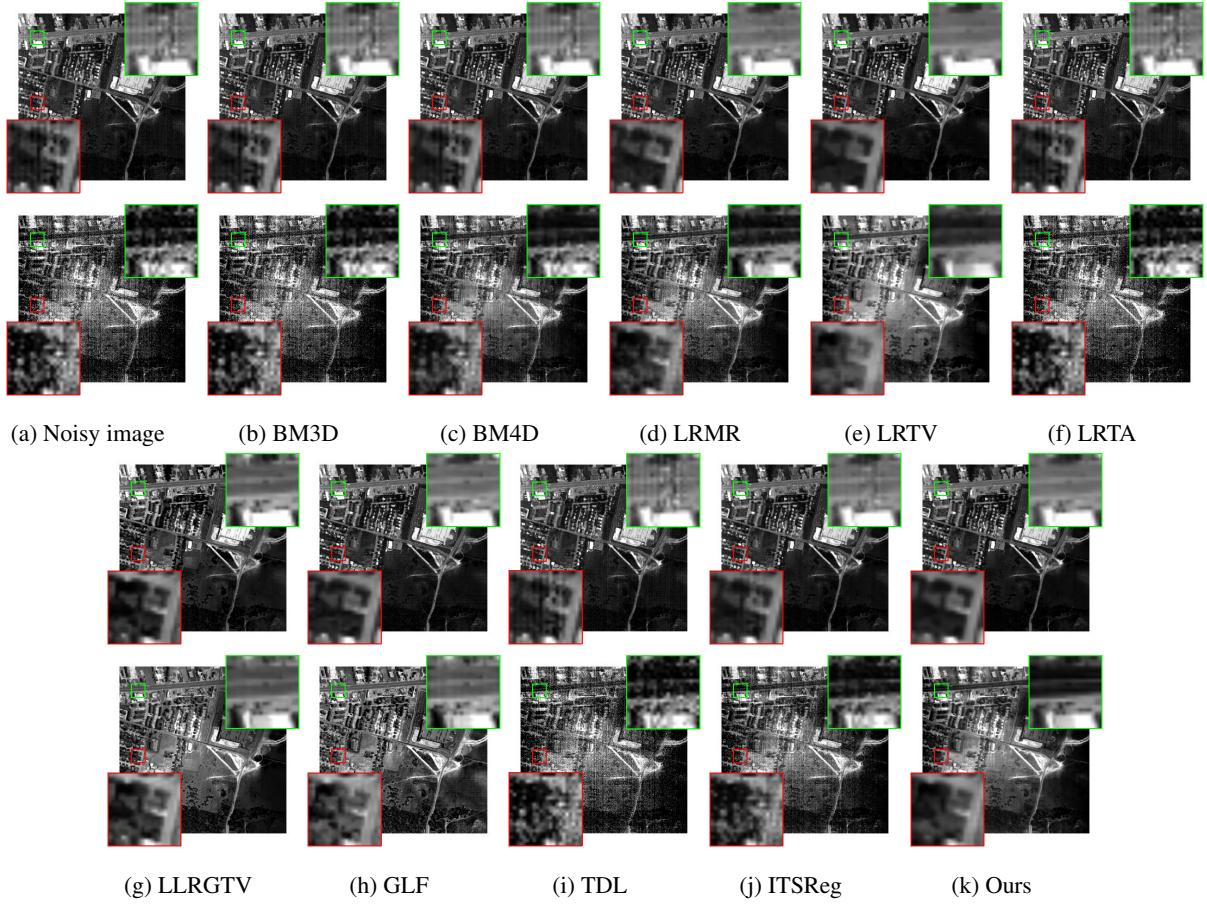


Figure 11: Real complex noise removal results at two bands (indexed by 1, 109 respectively) of HYDICE urban HSI. Two demarcated areas in each image are amplified at a 6 times larger scale for easy observation of details.

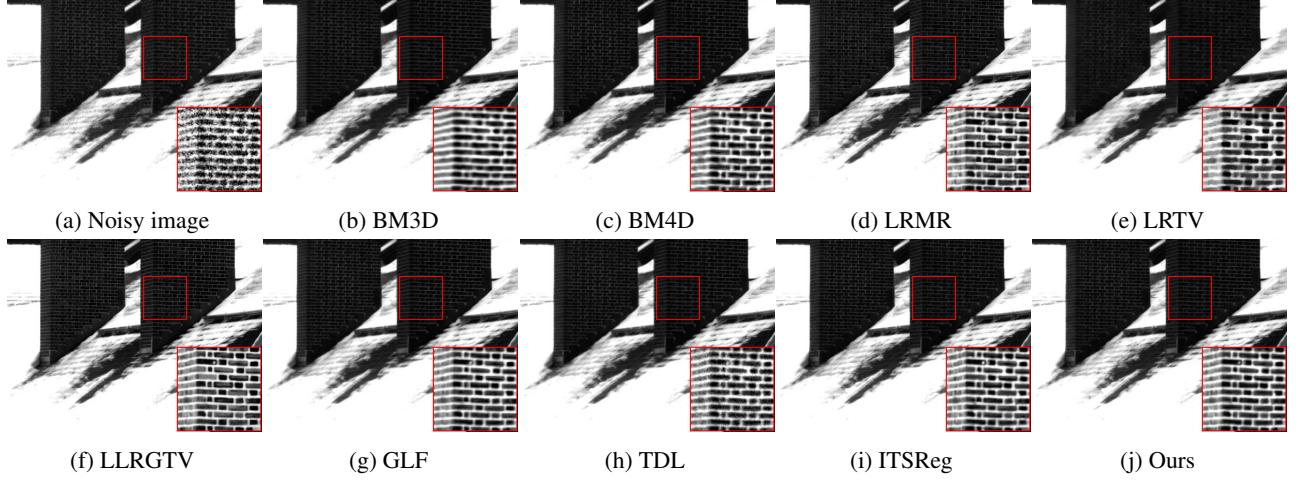


Figure 12: Real random noise removal results on HHD dataset. One demarcated area in each image is amplified at a 2 times larger scale for easy observation of details.

the original structure of objects of this housing estate. LLRGTV, GLF and our LBTF-HSI successfully gets rid of the stripe noise while preserving the topology structure of this housing estate. At band 109, the image is highly corrupted by miscellaneous complex noise. Obvious artefacts are still remained in the results of many competing methods (i.e. BM3D, BM4D,

LRTA, TDL, ITS-Reg). While LLRGTV and GLF do produce appealing results with good perceptual quality, these results apparently deviate from the underlying ground truth (see green box region at band 109). This phenomena may be caused by the incorrect specified subspace dimension (i.e. objective rank required by their low rank approximation techniques). As a com-

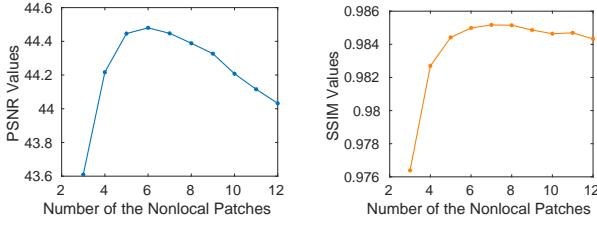


Figure 13: Effects of patch sizes on denoising performance.

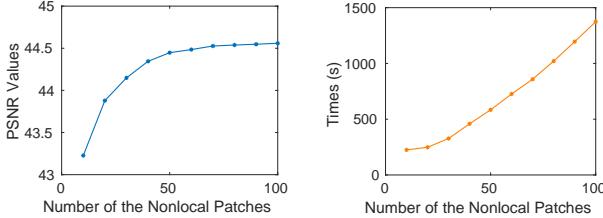


Figure 14: Effects of the number of nonlocal patches on denoising performance.

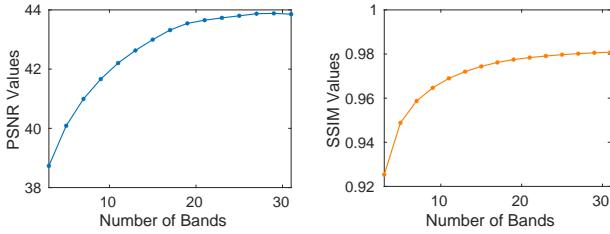


Figure 15: Effects of the number of bands on denoising performance.

parison, Our method does not suffer from the rank determination issue, thus it not only recovers the de facto semantics of the demarcated area (i.e. the scene of neighbourhood of highway), but also produces results with high fidelity.

Figure 12 displays the real random noise removal results On HHD dataset. From the demarcated window, we can observe that our LBTF-HSI method obtains artifact-free image with clearer texture and line pattern. In summary, LBTF-HSI has obtained better performance in terms of noise suppression, detail preserving, visual pleasure and PSNR value under different noise level, even in the real-world unknown noise context.

4.3. Discussion

Besides the initialization strategy aforementioned, there are other parameters introduced by different stages of our model, i.e. patch size, numbers of nonlocal patches (for grouping) and numbers of iterations (for iterative framework). Figure 13 shows the PSNR/SSIM values with respect to different patch size. Patch size 6 (6x6) and 7 achieve best PSNR and SSIM values respectively, among all candidates. Figure 14 illustrates how PSNR/Times value varies with respect to the number of nonlocal patches. We can see the denoising results become gradually better with large number of nonlocal patches, inferring the nonlocal self-similarity could be sufficiently utilized by our model, even in a relaxed condition. Nevertheless, given the computational cost and marginal enhancement through increas-

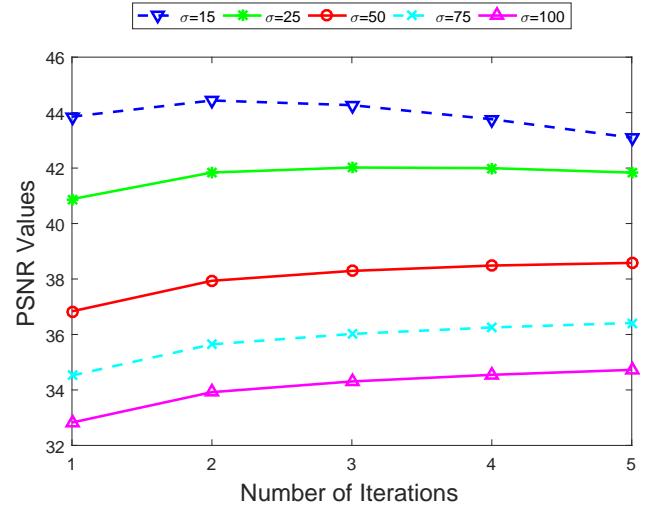


Figure 16: Effects of the number of iterations on denoising performance with respect to different noise levels.

ing the number of nonlocal patches, we set it to 50 in all of our experiment.

We also show how the number of bands of HSI influences the denoising capacity of our model. From Figure 15, we can observe that the denoising results become gradually better with larger number of bands. This suggests the information contained in one band could be utilized to recover other bands, such that the global correlation along the spectrum can be effectively exploited by our model.

Figure 16 displays the effects of numbers of iteration on denoising performance with respect to different noise levels. Generally, more stronger noise intensity will require more iteration times to achieve better performance, while at a expense of computing efficiency. we can see when noise intensity is relatively small (e.g. $\sigma = 15$), running algorithm in more than 2 iterations would successively degenerate the performance. Though the degradation issue is not observed during 5 iterations in strong corruption cases (e.g. $\sigma = 50, 75, 100$), the performance increment through iterations becomes limited while significantly increasing the computational cost. Therefore, we suggest the use of {1, 2, 3, 4, 5} for $\sigma = \{15, 25, 50, 75, 100\}$ in the simulated data experiments respectively.

5. Conclusion

In this paper, we presented an effective Low-rank Bayesian Tensor Factorization based HSI denoising method, which considered two intrinsic characteristics of HSIs: the nonlocal self-similarity across space and the global correlation across spectrum. In order to sufficiently embed these useful priors into our model, the LBTF is utilized to describe the spatial-spectral correlation of each FBP formed by block matching. This model was effectively solved by our deterministic algorithm derived under the variational Bayesian framework. Besides, an iterative denoising framework was introduced for the purpose of further enhancing the denoising capability of our method. The experimental results on simulated and real HSI denoising showed that

the proposed method outperformed many state-of-the-art methods and demonstrated the effectiveness of the proposed method.

We encode the noise structure as Gaussian distribution in our hierarchical probabilistic model. Since in real case, the statistical distribution of noise structure may be hard to be determined, it is worth investigating more effective noise model to model the noise from the real world in future.

6. Acknowledgements

We thank the anonymous reviewers for their helpful comments and suggestions to improve this paper. This work was supported by the National Science Foundation of China under Grants no. 61672096.

References

- [1] J. F. Mustard, C. M. Pieters, Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra, *Journal of Geophysical Research: Solid Earth* 94 (B10) (1989) 13619–13634.
- [2] R. Neville, Automatic endmember extraction from hyperspectral data for mineral exploration, in: International Airborne Remote Sensing Conference and Exhibition, 4 th/21 st Canadian Symposium on Remote Sensing, Ottawa, Canada, 1999.
- [3] M. Gianiello, G. Lechi, The development of superspectral approaches for the improvement of land cover classification, *IEEE Transactions on Geoscience and Remote Sensing* 42 (11) (2004) 2670–2679.
- [4] M. Lewis, V. Jooste, A. A. de Gasparis, Discrimination of arid vegetation with airborne multispectral scanner hyperspectral imagery, *IEEE Transactions on Geoscience and Remote Sensing* 39 (7) (2001) 1471–1479.
- [5] R. Marion, R. Michel, C. Faye, Measuring trace gases in plumes from hyperspectral remotely sensed data, *IEEE Transactions on Geoscience and Remote Sensing* 42 (4) (2004) 854–864.
- [6] A. Chen, The inpainting of hyperspectral images: A survey and adaptation to hyperspectral data, *SPIE Remote Sensing. International Society for Optics and Photonics* (2012) 85371–85371.
- [7] H. Van Nguyen, A. Banerjee, R. Chellappa, Tracking via object reflectance using a hyperspectral video camera, in: The IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 44–51.
- [8] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (2) (2012) 354–379.
- [9] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, K. Ikeuchi, High-resolution hyperspectral imaging via matrix factorization, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 2329–2336.
- [10] M. Uzair, A. Mahmood, A. Mian, Hyperspectral face recognition with spatirospectral information fusion and pls regression, *IEEE Transactions on Image Processing* 24 (3) (2015) 1127–1137.
- [11] F. Deger, A. Mansouri, M. Pedersen, J. Y. Hardeberg, Y. Voisin, A sensor-data-based denoising framework for hyperspectral images, *Optics express* 23 (3) (2015) 1938–1950.
- [12] M. Maggioni, V. Katkovnik, K. Egiazarian, A. Foi, Nonlocal transform-domain filter for volumetric data denoising and reconstruction, *IEEE Transactions on Image Processing* 22 (1) (2013) 119–133.
- [13] Q. Xie, Q. Zhao, D. Meng, Z. Xu, S. Gu, W. Zuo, L. Zhang, Multispectral images denoising by intrinsic tensor sparsity regularization, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1692–1700.
- [14] Q. Yuan, L. Zhang, H. Shen, Hyperspectral image denoising employing a spectral–spatial adaptive total variation model, *IEEE Transactions on Geoscience and Remote Sensing* 50 (10) (2012) 3660–3677.
- [15] Y.-Q. Zhao, J. Yang, Hyperspectral image denoising via sparse representation and low-rank constraint, *IEEE Transactions on Geoscience and Remote Sensing* 53 (1) (2015) 296–308.
- [16] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, B. Zhang, Decomposable nonlocal tensor dictionary learning for multispectral image denoising, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2949–2956.
- [17] X. Liu, S. Bourennane, C. Fossati, Denoising of hyperspectral images using the parafac model and statistical performance analysis, *IEEE Transactions on Geoscience and Remote Sensing* 50 (10) (2012) 3717–3724.
- [18] G. Chen, S.-E. Qian, Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage, *IEEE Transactions on Geoscience and Remote Sensing* 49 (3) (2011) 973–980.
- [19] H. Zhang, W. He, L. Zhang, H. Shen, Q. Yuan, Hyperspectral image restoration using low-rank matrix recovery, *IEEE Transactions on Geoscience and Remote Sensing* 52 (8) (2014) 4729–4743.
- [20] Y. Fu, A. Lam, I. Sato, Y. Sato, Adaptive spatial-spectral dictionary learning for hyperspectral image restoration, *International Journal of Computer Vision* 122 (2) (2017) 228–245.
- [21] W. He, H. Zhang, H. Shen, L. Zhang, Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 (3) (2018) 713–729.
- [22] L. Zhuang, J. M. Bioucas-Dias, Hyperspectral image denoising based on global and non-local low-rank factorizations, in: Image Processing (ICIP), 2017 IEEE International Conference on, IEEE, 2017, pp. 1900–1904.
- [23] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Transactions on Image Processing* 15 (12) (2006) 3736–3745.
- [24] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, *IEEE Transactions on Image Processing* 16 (8) (2007) 2080–2095.
- [25] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, *IEEE Transactions on Image Processing*.
- [26] S. Gu, L. Zhang, W. Zuo, X. Feng, Weighted nuclear norm minimization with application to image denoising, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2862–2869.
- [27] J. Xu, L. Zhang, W. Zuo, D. Zhang, X. Feng, Patch group based nonlocal self-similarity prior learning for image denoising, in: The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 244–252.
- [28] A. Chopra, H. Lian, Total variation, adaptive total variation and nonconvex smoothly clipped absolute deviation penalty for denoising blocky images, *Pattern Recognition* 43 (8) (2010) 2609–2619.
- [29] E. J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, *Journal of the ACM* 58 (3) (2011) 11.
- [30] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, Y. Ma, Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix, *Computational Advances in Multi-Sensor Adaptive Processing* 61 (6).
- [31] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, *arXiv preprint arXiv:1009.5055*.
- [32] T. Zhou, D. Tao, Godoc: Randomized low-rank & sparse matrix decomposition in noisy case, in: International Conference on Machine Learning (ICML), Omnipress, 2011.
- [33] X. Ding, L. He, L. Carin, Bayesian robust principal component analysis, *IEEE Transactions on Image Processing* 20 (12) (2011) 3419–3430.
- [34] Y. J. Lim, Y. W. Teh, Variational bayesian approach to movie rating prediction, in: Proceedings of KDD cup and workshop, Vol. 7, 2007, pp. 15–21.
- [35] V. Y. Tan, C. Févotte, Automatic relevance determination in nonnegative matrix factorization, in: SPARS'09-Signal Processing with Adaptive Sparse Structured Representations, 2009.
- [36] S. D. Babacan, M. Luessi, R. Molina, A. K. Katsaggelos, Sparse bayesian methods for low-rank matrix estimation, *IEEE Transactions on Signal Processing* 60 (8) (2012) 3964–3977.
- [37] Q. Zhao, D. Meng, Z. Xu, W. Zuo, L. Zhang, Robust principal component analysis with complex noise, in: International Conference on Machine Learning (ICML), 2014, pp. 55–63.
- [38] Y. Chen, X. Cao, Q. Zhao, D. Meng, Z. Xu, Denoising hyperspectral

- image with non-iid noise structure, arXiv preprint arXiv:1702.00098.
- [39] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM review 51 (3) (2009) 455–500.
- [40] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, C. Faloutsos, Tensor decomposition for signal processing and machine learning, IEEE Transactions on Signal Processing 65 (13) (2017) 3551–3582.
- [41] N. Renard, S. Bourennane, J. Blanc-Talon, Denoising and dimensionality reduction using multilinear tools for hyperspectral images, IEEE Geoscience and Remote Sensing Letters 5 (2) (2008) 138–142.
- [42] Q. Zhao, L. Zhang, A. Cichocki, Bayesian cp factorization of incomplete tensors with automatic rank determination, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (9) (2015) 1751–1763.
- [43] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.
- [44] A. Georges, G. Kotliar, W. Krauth, M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Reviews of Modern Physics 68 (1) (1996) 13.
- [45] W. He, H. Zhang, L. Zhang, H. Shen, Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration, IEEE Transactions on Geoscience and Remote Sensing 54 (1) (2016) 178–188.
- [46] F. Yasuma, T. Mitsunaga, D. Iso, S. K. Nayar, Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum, IEEE Transactions on Image Processing 19 (9) (2010) 2241–2253.
- [47] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.
- [48] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: A feature similarity index for image quality assessment, IEEE Transactions on Image Processing 20 (8) (2011) 2378–2386.
- [49] L. Wald, Data fusion: definitions and architectures: fusion of images of different spatial resolutions, Presses des MINES, 2002.
- [50] R. H. Yuhas, J. W. Boardman, A. F. Goetz, Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques, in: Summaries of the 4th Annual JPL Airborne Geoscience Workshop, 1993.
- [51] X. Liu, M. Tanaka, M. Okutomi, Single-image noise level estimation for blind denoising, IEEE Transactions on Image Processing 22 (12) (2013) 5226–5237.
- [52] A. Chakrabarti, T. Zickler, Statistics of real-world hyperspectral images, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 193–200.