Nonlocal Sparse Tensor Factorization for Semiblind Hyperspectral and Multispectral Image Fusion

Renwei Dian, *Student Member, IEEE*, Shutao Li, *Fellow, IEEE*, Leyuan Fang, *Senior Member, IEEE*, Ting Lu, *Member, IEEE*, and José M. Bioucas-Dias, *Fellow, IEEE*

***Abstract*—Combining a high-spatial-resolution multispectral image (HR-MSI) with a low-spatial-resolution hyperspectral image (LR-HSI) has become a common way to enhance the spatial resolution of the HSI. The existing state-of-the-art LR-HSI and HR-MSI fusion methods are mostly based on the matrix factorization, where the matrix data representation may be hard to fully make use of the inherent structures of 3-D HSI. We propose a nonlocal sparse tensor factorization approach, called the NLSTF_SMBF, for the semiblind fusion of HSI and MSI. The proposed method decomposes the HSI into smaller full-band patches (FBPs), which, in turn, are factored as dictionaries of the three HSI modes and a sparse core tensor. This decomposition allows to solve the fusion problem as estimating a sparse core tensor and three dictionaries for each FBP. Similar FBPs are clustered together, and they are assumed to share the same dictionaries to make use of the nonlocal self-similarities of the HSI. For each group, we learn the dictionaries from the observed HR-MSI and LR-HSI. The corresponding sparse core tensor of each FBP is computed via tensor sparse coding. Two distinctive features of NLSTF_SMBF are that: 1) it is blind with respect to the point spread function (PSF) of the hyperspectral sensor and 2) it copes with spatially variant PSFs. The experimental results provide the evidence of the advantages of the NLSTF_SMBF method over the existing state-of-the-art methods, namely, in semiblind scenarios.**

***Index Terms*—Hyperspectral imaging, hyperspectral super-resolution, nonlocal spatial self-similarity, semiblind fusion, sparse tensor factorization.**

I. Introduction

HYPERSPECTRAL imaging has recently shown promising performance in various computer vision tasks [1], [2] and remote sensing [3]–[5] tasks. The advantages of the HSI come from its high spectral resolution, which capacitates an accurate recognition of the materials. Due to the limited sun irradiance, there is a tradeoff between spectral resolution and spatial resolution for the imaging sensors. HSIs with high spectral resolution suffer from low-spatial resolution. On the contrary, MSIs with much lower spectral resolution can be acquired with higher spatial resolution. In this way, combining a high-spatial-resolution MSI (HR-MSI) and a low-spatial-resolution HSI (LR-HSI) is an economical way to obtain a high-spatial-resolution HSI (HR-HSI), which has the same spectral resolution and spatial resolution as the LR-HSI and HR-MSI [6], [7], respectively. The fusion procedure breaks the limitations of the imaging sensors and has demonstrated to be very effective in practice. The fusion procedure is different from the single image super-resolution [8], which aims at recovering an HR image only from an LR image.

Recently, HSI and MSI fusion, which fuses an LR-HSI with an HR-MSI, has received increasing attention. The recent HSI and MSI fusion methods can be categorized as nonblind and semiblind. The nonblind fusion methods assume that the point spread function (PSF) of a hyperspectral imaging sensor is known, which allows to model the LR-HSI as a linear operation applied to the HR-HSI. Yokoya *et al.* [9] used the coupled non-negative matrix factorization scheme to solve the nonblind fusion problem, which alternatively estimates the two factors of the HR-HSI. Lanaras *et al.* [10] proposed the coupled spectral unmixing (CSU) model for the fusion problem, where the LR-HSI and HR-MSI are alternatively unmixed to estimate the spectral basis and abundances. To further make use of the correlations among spectral bands, Simões *et al.* [11] and Wei *et al.* [12] used a low-dimensional subspace representation model to approximate the target HR-HSI, and then use the total variation and learned sparse prior to regularize the problem, respectively. Dong *et al.* [13] proposed a non-negative structured sparse representation model to solve the fusion problem, which exploits the nonlocal HSI spatial

Manuscript received April 10, 2019; revised July 25, 2019 and September 30, 2019; accepted October 28, 2019. This work was supported in part by the National Natural Science Fund of China under Grant 61890962, Grant 61520106001, and Grant 61801179, in part by the Science and Technology Plan Project Fund of Hunan Province under Grant CX2018B171, Grant 2017RS3024, and Grant 2018TP1013, in part by the Science and Technology Talents Program of Hunan Association for Science and Technology under Grant 2017TJ-Q09, in part by the Hunan Provincial Innovation Foundation for Postgraduate, in part by the Portuguese Science and Technology Foundation under Projects under Grant UID/EEA/50008/2019, in part by the Hunan Provincial Innovation Foundation for Postgraduate, and in part by the China Scholarship Council. This article was recommended by Associate Editor S. Cruces. *(Corresponding author: Shutao Li.)*

R. Dian, S. Li, L. Fang, and T. Lu are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China (e-mail: drw@hnu.edu.cn; shutao_li@hnu.edu.cn; fangleyuan@gmail.com; tingluhnu@gmail.com).

J. M. Bioucas-Dias is with the Instituto de Telecomunicacões, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2019.2951572

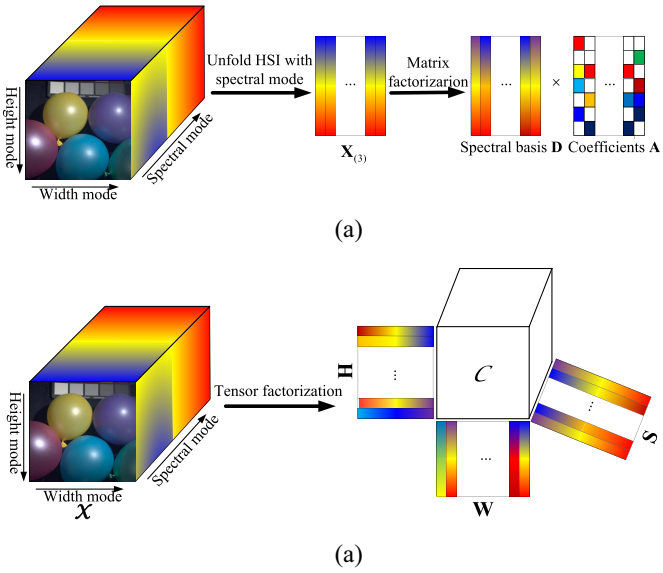


Fig. 1. Illustration of the traditional matrix decomposition and proposed tensor decomposition of HSI. (a) Matrix factorization-based the HR-HSI decomposition. (b) Tensor factorization based the HR-HSI decomposition.

similarities. Furthermore, some nonblind methods [14]–[17] solve the fusion problem from a point of tensor representation. For example, Li *et al.* [14] simultaneously conduct sparse tensor factorization for the LR-HSI and HR-MSI, where dictionaries of three modes and sparse core tensor are alternatively estimated for several iterations. Kanatsoulis *et al.* [15] proposed a hybrid model, which is based on the low-rank matrix and tensor factorization, to settle the fusion problem.

The semiblind HSI and MSI fusion methods assume only that the spectral response of the camera is known and do not rely on the knowledge of PSF, which is often hardly known in practice. The matrix factorization-based semiblind fusion methods have been much actively investigated. These methods first compute a spectral mode HSI unfolding matrix and then factor this matrix into a spectral basis and the coefficients, as Fig. 1(a) schematizes. In this way, the estimation of the HR-HSI is transformed into estimating a spectral basis and the coefficients. These methods, introduced first in [18] and [19], decompose the HR-HSI into a spectral basis, learned from the LR-HSI with the dictionary learning method, and the coefficients computed on the HR-MSI with the learned spectral basis, using a sparsity-inducing prior. A similar idea was proposed by Huang *et al.* [20] for remote sensing HSI, which uses the singular value decomposition to learn the spectral basis. Akhtar *et al.* [21] first estimated a non-negative dictionary, and then compute the coefficients via simultaneous greedy pursuit algorithm, where a prior enforcing similarity of the nearby spectra is adopted. The work in [22] proposes to learn the spectral dictionary with a Beta process from the LR-HSI, and uses the Bayesian sparse coding to infer the coefficients from the HR-MSI.

The HSI is the 3-D data cube and then can be dealt with from a tensor point of view. Recently, tensor factorization has found broad applications in visual tracking [23], [24]; objection detection [25]; HSI denoising [26], [27]; completion [28], [29]; and compressive sensing [30]–[32]. The HSI mainly has two characteristics, that is, high correlations among the spectral bands and nonlocal spatial self-similarities. The high correlations among the spectral bands result in low-rank structure of the HSI, which has been widely used for HSI restoration [27], [30], [32].

Motivated by the aforementioned works, we propose a nonlocal sparse tensor factorization (NLSTF_SMBF) method for the semiblind fusion of an HR-MSI and an LR-HSI. The NLSTF_SMBF approach has two main components: 1) it exploits HSIs local spatial–spectral correlation and nonlocal spatial self-similarities and 2) it is based on the sparse Tucker factorization. Each FBP of the HR-HSI contains its local spatial–spectral information. To model this local information, we approximate each FBP by the dictionaries of three modes multiplied by a core tensor, as shown in Fig. 1(b). In the decomposition, the dictionaries of three modes, respectively, reflect the HSI information of the three modes. Meanwhile, the core tensor encodes the strength of the interaction among the dictionary elements of the three modes. In this way, the HSI information of three modes is incorporated into a unified model and, therefore, we can better model the local spatial–spectral correlations. In addition, images often contain a number of similar patches [33]. To exploit the nonlocal self-similarities in the HR-HSI, the similar FBPs of the HR-HSI are grouped together, and then similar FBPs are assumed to admit sparse representation on the same dictionaries. This article is the longer version of conference paper [34], and we further improve the conference paper for three aspects: 1) we change the dictionary learning algorithm and tensor sparse coding algorithm as the ℓ_1-norm-based ones, which improve both the efficiency and accuracy of the algorithm; 2) we add experiments on the remote sensing HSI, which further demonstrate the effectiveness of our method; and 3) we emphasize that our method semiblind and conduct the experiments on the variant PSF case.

The remainder of this article is organized as follows. The notations on tensors are given in Section II. Section III presents the formulation of the HR-MSI and LR-HSI fusion. We introduce the proposed NLSTF_SMBF method for the HR-MSI and LR-HSI fusion in Section IV. In Section V, experiments and discussions of spatially invariant and variant PSFs are presented. The conclusions are introduced in Section VI.

II. Notations on Tensors

We denote an N-dimensional tensor as $\mathcal{M} = (m_{i_1 \ldots i_N}) \in \mathbb{R}^{I_1 \times I_2 \times, \ldots, \times I_N}$. The mode-$n$ unfolding vector of the tensor $\mathcal{M}$ is defined as I_n-dimensional vector by changing index i_n while keeping all the other indices fixed. The n-mode unfolding matrix of the tensor $\mathcal{M}$ is $\mathbf{M}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2, \ldots, I_{n-1} I_{n+1}, \ldots, I_N}$, whose columns are all n-mode vectors [35].

Tensor Norms: The Frobenius norm and ℓ_1-norm of tensor $\mathcal{M}$ are defined as $||\mathcal{M}||_F = \sqrt{\sum_{i_1,\ldots,i_N} |m_{i_1 \cdots i_N}|^2}$ and $||\mathcal{M}||_1 = \sqrt{\sum_{i_1,\ldots,i_N} |m_{i_1 \cdots i_N}|}$, respectively.

Tensor Product: The n-mode product of the matrix $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$ and tensor $\mathcal{M} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ can be represented as

$$\mathcal{C} = \mathcal{M} \times_n \mathbf{B} \tag{1}$$

where $\mathcal{C} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times J_n \times \cdots \times I_N}$, and its elements are calculated as

$$c_{i_1 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} = \sum_{i_n} m_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} b_{j_n i_n}. \tag{2}$$

Equation (1) is equivalent as matrix multiplication

$$\mathbf{C}_{(n)} = \mathbf{B}\mathbf{M}_{(n)}. \tag{3}$$

The order of the multiplications does not make a difference for distinct modes, which means that

$$\mathcal{M} \times_m \mathbf{A} \times_n \mathbf{B} = \mathcal{M} \times_n \mathbf{B} \times_m \mathbf{A} (n \neq m). \tag{4}$$

For the same mode, (4) is transformed as

$$\mathcal{M} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{M} \times_n (\mathbf{B}\mathbf{A}). \tag{5}$$

Tucker Decomposition: In this article, we use the Tucker model [36] for tensors. Given the n-mode dictionaries $\mathbf{D}_n \in \mathbb{R}^{J_n \times I_n} (n = 1, 2, \ldots, N)$, we define the tensor $\mathcal{C} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$

$$\mathcal{C} = \mathcal{M} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times \cdots \times_N \mathbf{D}_N. \tag{6}$$

Then, we have (see [37])

$$\mathbf{c} = (\mathbf{D}_N \otimes \mathbf{D}_{N-1} \otimes, \ldots, \otimes \mathbf{D}_1)\mathbf{m} \tag{7}$$

where the symbol $\otimes$ represents the Kronecker product, and $\mathbf{c} = \text{vec}(\mathcal{C}) \in \mathbb{R}^J (J = \prod_{n=1}^N J_n)$ and $\mathbf{m} = \text{vec}(\mathcal{M}) \in \mathbb{R}^I (I = \prod_{n=1}^N I_n)$ are vectors obtained by vectorizing the tensors $\mathcal{C}$ and $\mathcal{M}$, respectively. According to (6) and (7), we may say that the tensor $\mathcal{C}$ admits a sparse representation on the n-mode dictionaries, if the vector $\mathbf{c}$ has a sparse representation on the dictionary $\mathbf{D} = \mathbf{D}_N \otimes \mathbf{D}_{N-1} \otimes \cdots \otimes \mathbf{D}_1$.

There is plenty of evidence [37] that the hyperspectral tensors from the real world admit sparse representations with respect to suitable dictionaries. Furthermore, using the Tucker decomposition makes it easy to exploit sparsity in both the spatial and the spectral modes. This is the fundamental reason for the adoption of the Tucker decomposition in this article.

III. Problem Formulation

All of the HR-HSI, HR-MSI, and LR-HSI are denoted as 3-D tensors. The clean HR-HSI is represented by $\mathcal{X} \in \mathbb{R}^{W \times H \times S}$, which has S spectral bands and WH spectral pixels. $\mathcal{Y} \in \mathbb{R}^{w \times h \times S}$ denotes the obtained LR-HSI with S spectral bands, and $w < W$ and $h < H$ pixels, corresponding to a spatial downsampled version of $\mathcal{X}$. $\mathcal{Z} \in \mathbb{R}^{W \times H \times s}$ denotes the acquired HR-MSI of the same scenario with the same spatial resolution as $\mathcal{X}$, and $s < S$, which corresponds to a spectrally downsampled version of $\mathcal{X}$. The fusion aims at estimating $\mathcal{X}$ by combining $\mathcal{Z}$ with $\mathcal{Y}$.

A. Matrix Decomposition-Based HR-MSI and LR-HSI Fusion

The linear unmixing model assumes that each spectral pixel can be linearly represented as a number of spectral signatures [38]. Based on this assumption, the matrix factorization-based fusion approaches first unfold the HR-HSI along the spectral mode as a matrix, and the unfolding matrix is decomposed as coefficients and spectral basis, as illustrated in Fig. 1(a). The decomposition can be formulated as

$$\mathbf{X}_{(3)} = \mathbf{D}\mathbf{A} \tag{8}$$

where $\mathbf{X}_{(3)} \in \mathbb{R}^{S \times WH}$ is the 3-mode unfolding matrix of tensor $\mathcal{X}$ [i.e., each column of $\mathbf{X}_{(3)}$ is a spectral vector of size S], and $\mathbf{A} \in \mathbb{R}^{L \times WH}$ and $\mathbf{D} \in \mathbb{R}^{S \times L}$ are the coefficients and spectral basis, respectively.

Both the HR-MSI and LR-HSI are the downsampled versions of the HR-HSI

$$\begin{aligned} \mathbf{Y}_{(3)} &= \mathbf{X}_{(3)}\mathbf{G} \\ \mathbf{Z}_{(3)} &= \mathbf{P}_3\mathbf{X}_{(3)} \end{aligned} \tag{9}$$

where $\mathbf{Z}_{(3)} \in \mathbb{R}^{s \times WH}$ and $\mathbf{Y}_{(3)} \in \mathbb{R}^{S \times wh}$ are the matrices obtained by unfolding $\mathcal{Z}$ and $\mathcal{Y}$ with the third mode, respectively. Matrix $\mathbf{G}$ models blur and subsampling operations, that is, $\mathbf{G} = \mathbf{G}_b\mathbf{G}_d$, where $\mathbf{G}_b$ is a matrix representing a convolution between the PSF of the sensor and the HR-HSI bands and $\mathbf{G}_d$ is the subsampling matrix, which selects the corresponding spectral pixels. In the nonblind fusion methods, the matrix $\mathbf{G}$ is assumed to be known, however, the semiblind fusion methods do not rely on it. $\mathbf{P}_3 \in \mathbb{R}^{s \times S}$ is the matrix modeling spectral downsampling in the multispectral sensor.

B. Tensor Decomposition-Based HR-MSI and LR-HSI Fusion

As illustrated in Fig. 1(b) and different from the matrix decomposition-based methods, the proposed tensor factorization-based method aims at decomposing the clean HR-HSI $\mathcal{X}$ as dictionaries of three modes and a core tensor, that is,

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S} \tag{10}$$

where the matrices $\mathbf{W} \in \mathbb{R}^{W \times n_w}$, $\mathbf{H} \in \mathbb{R}^{H \times n_h}$, and $\mathbf{S} \in \mathbb{R}^{S \times n_s}$ denote the dictionaries of the width, height, and spectral modes, respectively. The three dictionaries are separable and express the basic information of the HSI respective modes. The tensor $\mathcal{C} \in \mathbb{R}^{n_w \times n_h \times n_s}$ holds the coefficients of $\mathcal{X}$ on the three dictionaries.

The observed LR-HSI $\mathcal{Y}$ can be modeled as the downsampled version of $\mathcal{X}$. If the PSFs and downsampling matrices of the hyperspectral imaging sensor are separable in two spatial modes [39], then the LR-HSI is obtained from the HR-HSI as

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{P_1} \times_2 \mathbf{P_2} \tag{11}$$

where $\mathbf{P_1} \in \mathbb{R}^{w \times W}$ and $\mathbf{P_2} \in \mathbb{R}^{h \times H}$ denote the subsampling matrices of the width and height modes, respectively. The separability assumption is valid, for example, for the boxcar and Gaussian convolution kernels with the major axis aligned with the spatial unit vectors. Under the separable assumption, we have

$$\mathbf{G} = (\mathbf{P_2} \otimes \mathbf{P_1})^T \tag{12}$$

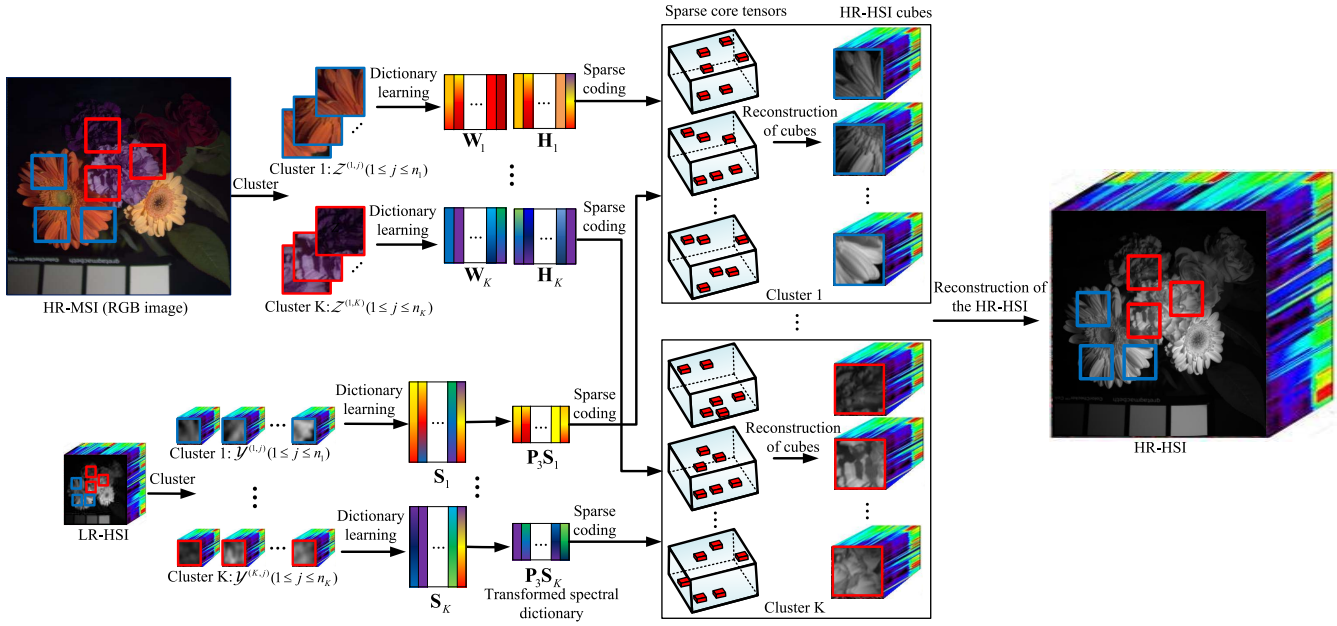


Fig. 2. Schematic view of the proposed NLSTF_SMBF method.

implying that the first equation in (9) and (11) are equivalent. It should be noted the PSF and the downsampling matrices are not necessarily always separable. The separability assumption is a special case, which depends on the hyperspectral imaging sensor. If the PSF or the downsampling matrices are not separable in the width mode and height mode, we can only use the first equation in (9) rather than (11) to describe the spatial downsampling process.

The HR-MSI $\mathcal{Z}$ (i.e., the spectrally subsampled version of the HR-HSI) is given by

$$\mathcal{Z} = \mathcal{X} \times_3 \mathbf{P_3} \tag{13}$$

where $\mathbf{P}_3 \in \mathbb{R}^{s \times S}$ is the spectral subsampling matrix of the multispectral imaging sensor.

According to expression (10), to reconstruct the clean HSI, we need to estimate the dictionaries $\mathbf{W}$, $\mathbf{H}$, $\mathbf{S}$, and core tensor $\mathcal{C}$.

IV. Proposed NLSTF_SMBF Method

As illustrated in Fig. 2, the NLSTF_SMBF method has three main steps: 1) nonlocal clustering of the FBPs; 2) estimating the dictionaries of three modes; and 3) estimating the core tensor. If the entire HR-HSI is directly reconstructed, we have dictionaries of three modes $\mathbf{W} \in \mathbb{R}^{W \times n_w}$, $\mathbf{H} \in \mathbb{R}^{H \times n_h}$, $\mathbf{S} \in \mathbb{R}^{S \times n_s}$, and the Kronecker dictionary $\mathbf{D}_{\text{kron}} = \mathbf{S} \otimes \mathbf{H} \otimes \mathbf{W} \in \mathbb{R}^{WHS \times n_w n_h n_s}$, which, for typical values of W, H, S, n_w, n_h, n_s, is very large. Dealing with such big dictionaries not only has high computational cost but also requires large storage resources. To circumvent these hurdles, the HR-HSI is reconstructed in a patch-by-patch manner. The entire HR-HSI is first partitioned into several FBPs of size $d_W \times d_H \times S$, yielding $\mathbf{W}_1 \in \mathbb{R}^{d_W \times n_w}$, $\mathbf{H}_1 \in \mathbb{R}^{d_H \times n_h}$, and $\mathbf{D}_{\text{kron1}} = \mathbf{S} \otimes \mathbf{H}_1 \otimes \mathbf{W}_1 \in \mathbb{R}^{d_W d_H S \times n_w n_h n_s}$ for each FBP, where $d_W << W$ and $d_H << H$. In this way, the size of dictionaries $\mathbf{W}_1$, $\mathbf{H}_1$, and $\mathbf{D}_{\text{kron1}}$ is considerably reduced. For example, if we have $W = 512, H = 512, d_W = 8$, and $d_H = 8$, the size of the dictionary $\mathbf{D}_{\text{kron1}}$ is reduced by a factor of $1/4096$ compared with the original size. Therefore, we adopt the patch-by-patch reconstruction approach as it largely reduces the computation and storage cost, while yielding a competitive performance, as illustrated in Section V. Another reason to use small FBPs is that it allows to cluster them and built cluster-fitted dictionaries. Based on the above tensor-based HSI factorization, the HR-MSI and LR-HSI fusion is transformed into estimating the core tensor and dictionaries of three modes for each HR-HSI FBP. First of all, to make use of the nonlocal self-similarities, the similar HR-MSI FBPs are grouped together, and then the FBPs of the LR-HSI and desired HR-HSI are also clustered on the basis of their spatial locations. The HR-HSI FBPs in the same cluster are decomposed on the same dictionaries using a sparsity-inducing prior. Next, we learn the dictionaries of three modes for each cluster and then estimate the sparse core tensor for each FBP via a fast tensor sparse coding method. Finally, the learned dictionaries and estimated sparse core tensors are used to reconstruct the HR-HSI FBPs. Each step of the NLSTF_SMBF is introduced in detail in the following text.

A. Nonlocal Clustering of the FBPs

Since the HR-MSI is modeled as the spectrally subsampled version of the HR-HSI, it preserves most of the spatial information of the HR-HSI. Therefore, we can learn the spatial self-similarities from the HR-MSI. Specifically, the HR-MSI $\mathcal{Z} \in \mathbb{R}^{W \times H \times s}$ is spatially partitioned into several overlapping FBPs with the spatial size of $d_W \times d_H$ and full spectral size of HR-MSI s. The total number of FBPs is $N_c = ([W - d_W/[d_W - p] + 1)([H - d_H]/[d_H - p] + 1)$, where p is the size of the overlap. The basic idea is to cluster the HR-MSI FBPs into K groups $Z^{(k)} = \{\mathcal{Z}^{(k,j)}\}_{j=1}^{n_k}, k = 1, 2, \ldots, K$, where

n_k is the number of FBPs in the kth cluster. $\mathcal{Z}^{(k,j)} \in \mathbb{R}^{d_W \times d_H \times s}$ denotes the jth FBP of the kth cluster, where d_W and d_H are the dimensions of the width and height modes, respectively. The Kmeans++ method [40] improves both the speed and accuracy of the K-means approach and, therefore, it is employed to obtain clusters of all HR-MSI FBPs. Based on the learned cluster structure, we build clusters of LR-HSIs and HR-HSIs FBPs with the same spatial structure. Each pixel in $\mathcal{Y}$ is assumed to correspond to a $c \times c$ patch in $\mathcal{Z}$, where c is the downsampling factor. If any pixel in $c \times c$ patch in $\mathcal{Z}$ is clustered into the kth group, the corresponding pixel in $\mathcal{Y}$ is also grouped into the kth group. The pixels in the HR-MSI patch of size $c \times c$ may be clustered into different groups and, therefore, the corresponding LR-HSI pixel is also divided into different groups simultaneously. In this way, we can acquire the clusters $\mathbf{Y}^{(k)} \in \mathbb{R}^{S \times N_k}, k = 1, 2, \ldots, K$ based on the learned cluster structure in $\mathcal{Z}$, where N_k is the number of spectral pixels in the kth cluster group.

B. Tensor Dictionary Learning

This section introduces the process of learning dictionaries of three modes. The HR-HSI FBPs in the same cluster are similar to each other and, therefore, we assume them to admit sparse representation on the same dictionaries. The dictionary learning scheme is the same for all clusters, and we take the dictionary learning of the kth cluster for an example to introduce the dictionary learning scheme without loss of generality.

On the basis of the above tensor decomposition, the FBPs $\mathcal{X}^{(k,j)}$ in the kth cluster are approximately represented as

$$\mathcal{X}^{(k,j)} = \mathcal{C}^{(k,j)} \times_1 \mathbf{W}_k \times_2 \mathbf{H}_k \times_3 \mathbf{S}_k, \quad j = 1, 2, \ldots, n_k \quad (14)$$

where the matrices $\mathbf{W}_k \in \mathbb{R}^{d_W \times l_W}$, $\mathbf{H}_k \in \mathbb{R}^{d_H \times l_H}$, and $\mathbf{S}_k \in \mathbb{R}^{S \times l_S}$ represent the dictionaries of the width mode with l_W atoms, height mode with l_H atoms, and spectral mode with l_S atoms, respectively. The core tensor $\mathcal{C}^{(k,j)} \in \mathbb{R}^{l_W \times l_H \times l_S}$ encodes the strength of the interactions among the columns of the dictionaries $\mathbf{W}_k$, $\mathbf{H}_k$, and $\mathbf{S}_k$. The HR-MSI preserves most of the spatial information of the HR-HSI and, therefore, the dictionaries $\mathbf{W}_k$ and $\mathbf{H}_k$ are learned from $\{\mathcal{Z}^{(k,j)}\}_{j=1}^{n_k}$. According to (13), the HR-MSI FBPs of the kth group, $\mathcal{Z}^{(k,j)}$ can be written as

$$\mathcal{Z}^{(k,j)} = \mathcal{X}^{(k,j)} \times_3 \mathbf{P}_3, \quad j = 1, 2, \ldots, n_k. \quad (15)$$

Combining (14) and (15), we can also write $\mathcal{Z}^{(k,j)}$ as

$$\mathcal{Z}^{(k,j)} = \mathcal{C}^{(k,j)} \times_1 \mathbf{W}_k \times_2 \mathbf{H}_k \times_3 \mathbf{S}_k^*, \quad j = 1, 2, \ldots, n_k \quad (16)$$

where $\mathbf{S}_k^* = \mathbf{P}_3 \mathbf{S}_k$ is the spectral dictionary for the HR-MSI obtained by downsamping $\mathbf{S}_k$ with the spectral mode. According to (16), the 1-mode (width mode) unfolding matrix of the tensor $\mathcal{Z}^{(k,j)}$ can be represented as

$$\mathbf{Z}_{(1)}^{(k,j)} = \mathbf{W}_k \mathbf{A}_{(1)}^{(k,j)}, \quad j = 1, 2, \ldots, n_k \quad (17)$$

where $\mathbf{A}_{(1)}^{(k,j)}$ and $\mathbf{Z}_{(1)}^{(k,j)}$ are 1-mode unfolding matrices of tensors $\mathcal{A}^{(k,j)} = \mathcal{C}^{(k,j)} \times_2 \mathbf{H}_k \times_3 \mathbf{S}_k^*$ and $\mathcal{Z}^{(k,j)}$, respectively. According to (17), we note that each column of $\mathbf{M}_{w_k} = [\mathbf{Z}_{(1)}^{(k,1)}, \mathbf{Z}_{(1)}^{(k,2)}, \ldots, \mathbf{Z}_{(1)}^{(k,n_k)}]$ can be written as a linear combination of columns of matric $\mathbf{W}_k$. In this article, we do use a sparsity-inducing prior to regularize the corresponding factorization. This prior is used both in the dictionary leaning and in the sparse coding stages. In the first stage, it promotes dictionaries, usually over-complete, with respect to which the respective mode is sparsely represented; in the second case, it yields sparse codes. Herein, our objective is to find a matrix factorization $\mathbf{M}_{w_k} = \mathbf{W}_k \mathbf{B}_{w_k}$ such that the columns of $\mathbf{M}_{w_k}$ are sparsely represented with respect to the dictionary $\mathbf{W}_k$. This matrix factorization problem is a severely ill-posed problem. To obtain such a factorization, $\mathbf{W}_k$ is designed to be over-complete and it is estimated by solving a sparsity-constrained dictionary learning problem, formulated as

$$\min_{\mathbf{W}_k, \mathbf{B}_{W_k}} \left|\left|\mathbf{M}_{w_k} - \mathbf{W}_k \mathbf{B}_{w_k}\right|\right|_F^2 + \lambda_1 \left|\left|\mathbf{B}_{w_k}\right|\right|_1 \quad (18)$$

where $||\cdot||_1$ and $||\cdot||_F$ denote the ℓ_1-norm and Frobenius norm, respectively, and λ_1 is the sparsity regularization parameter. Problem (18) is nonconvex and its solution is often not unique. However, it is convex with regard to the dictionary $\mathbf{W}_k$ and to the coefficients $\mathbf{B}_{w_k}$. Herein, we take the dictionary learning method proposed in [13] to solve the problem (18); it updates alternately the dictionary and the coefficients. The dictionary is updated, keeping the coefficients fixed, using block coordinate descent, which imperatively updates the columns of dictionary. The coefficients are updated, keeping the dictionary fixed, using the alternating direction method of multipliers (ADMMs) method [41].

We now address the 2-mode (height mode) of the tensor $\mathcal{Z}^{(k,j)}$ in a way parallel to mode-1. Equation (16) is equivalent to

$$\mathbf{Z}_{(2)}^{(k,j)} = \mathbf{H}_k \mathbf{B}_{(2)}^{(k,j)}, \quad j = 1, 2, \ldots, n_k \quad (19)$$

where $\mathbf{B}_{(2)}^{(k,j)}$ and $\mathbf{Z}_{(2)}^{(k,j)}$ are 2-mode matrices of tensors $\mathcal{B}^{(k,j)} = \mathcal{C}^{(k,j)} \times_1 \mathbf{W}_k \times_3 \mathbf{S}_k^*$ and $\mathcal{Z}^{(k,j)}$, respectively. On the basis of (19), we can also find that each column of the matrix $\mathbf{M}_{h_k} = [\mathbf{Z}_{(2)}^{(k,1)}, \mathbf{Z}_{(2)}^{(k,2)}, \ldots, \mathbf{Z}_{(2)}^{(k,n_k)}]$ can be linearly represented by columns in the matric $\mathbf{H}_k$. In a way similar to the estimation of $\mathbf{W}_k$, the estimation of $\mathbf{H}_k$ is carried out by solving the sparsity-constrained optimization problem

$$\min_{\mathbf{H}_k, \mathbf{B}_{h_k}} \left|\left|\mathbf{M}_{h_k} - \mathbf{H}_k \mathbf{B}_{h_k}\right|\right|_F^2 + \lambda_2 \left|\left|\mathbf{B}_{h_k}\right|\right|_1 \quad (20)$$

where λ_2 is the sparsity regularization parameter. As in (18), we compute an approximated solution to (20) using the dictionary learning algorithm proposed in [13].

The LR-HSI is modeled as the spatially downsampled version of the HR-HSI and, therefore, it preserves the bulk of the spectral information in the HR-HSI. Therefore, we learn the dictionary for the spectral mode $\mathbf{S}_k$ from $\mathbf{Y}^{(k)} \in \mathbb{R}^{S \times N_k}$, which are made up of LR-HSI pixels in the kth group. Here, we use the vertex component analysis (VCA) [42] to learn the spectral dictionary, which is the state-of-the-art spectral unmixing method.

C. Tensor Sparse Coding

Once we know the dictionaries of the kth cluster, $\mathbf{W}_k$, $\mathbf{H}_k$, and $\mathbf{S}_k$, the core tensor $\mathcal{C}^{(k,j)} \in \mathbb{R}^{l_W \times l_H \times l_S}$ needs to be estimated to recover the HR-HSI FBPs of the kth cluster. The PSF of HSI imaging sensor may be hard to estimate in practice when the PSF is spatially variational. Hence, only the observation model of HR-MSI, the second equation in (9), is used for estimation of the core tensors, which does not rely on the observation model of the LR-HSI, the first equation in (9). Therefore, our method is semiblind. Since the underlying equation related to $\mathcal{C}^{(k,j)}$ is underdetermined, the solution of $\mathcal{C}^{(k,j)}$ is not unique. Hence, the prior information is needed to regularize the estimation of $\mathcal{C}^{(k,j)}$. As in the estimation of the dictionaries, and as a consequence of the HSI self-similarities and local spatial-spectral correlations, we assume that the core tensor $\mathcal{C}^{(k,j)}$ is sparse. The ℓ_1-norm is the convex relaxation of the ℓ_0-norm, and the ℓ_1-norm regularization has been successfully used for tensor completion [28]. Here, we also adopt ℓ_1-norm to promote sparse core tensors, yielding the following optimization problem:

$$\min_{\mathcal{C}^{(k,j)}} \left|\left|\mathcal{Z}^{(k,j)} - \mathcal{C}^{(k,j)} \times_1 \mathbf{W}_k \times_2 \mathbf{H}_k \times_3 \mathbf{S}_k^*\right|\right|_F^2 + \lambda \left|\left|\mathcal{C}^{(k,j)}\right|\right|_1 \tag{21}$$

where λ is the sparsity regularization parameter, and $||\mathcal{C}^{(k,j)}||_1$ represents the ℓ_1-norm of tensor $\mathcal{C}^{(k,j)}$ (defined in Section II). On the basis of the equivalence of (1) and (3), problem in (21) is equivalent to

$$\min_{\mathbf{c}^{(k,j)}} \left|\left|\mathbf{z}^{(k,j)} - \mathbf{D}_k \mathbf{c}^{(k,j)}\right|\right|_F^2 + \lambda \left|\left|\mathbf{c}^{(k,j)}\right|\right|_1 \tag{22}$$

where $\mathbf{z}^{(k,j)} = \text{vec}(\mathcal{Z}^{(k,j)}) \in \mathbb{R}^{sd_W d_H}$ and $\mathbf{c}^{(k,j)} = \text{vec}(\mathcal{C}^{(k,j)}) \in \mathbb{R}^{l_W l_H l_K}$ are the vectors obtained by heaping up all 1-mode vectors of tensors $\mathcal{Z}^{(k,j)}$ and $\mathcal{C}^{(k,j)}$, respectively. The matrix $\mathbf{D}_k$ is computed by

$$\mathbf{D}_k = \mathbf{S}_k^* \otimes \mathbf{H}_k \otimes \mathbf{W}_k. \tag{23}$$

Optimization (22) is convex and thus we can solve it efficiently by the ADMM. By introducing the constraint $\mathbf{v} = \mathbf{c}^{(k,j)}$, we can acquire the augmented Lagrangian function

$$L\left(\mathbf{v}, \mathbf{c}^{(k,j)}, \mathbf{g}\right) = \left|\left|\mathbf{z}^{(k,j)} - \mathbf{D}_k \mathbf{v}\right|\right|_F^2 + \lambda \left|\left|\mathbf{c}^{(k,j)}\right|\right|_1 + \mu \left|\left|\mathbf{c}^{(k,j)} - \mathbf{v} + \frac{\mathbf{g}}{2\mu}\right|\right|_F^2 \tag{24}$$

where $\mu > 0$ is the penalty parameter, and $\mathbf{g}$ is a vector holding the scaled Lagrangian multipliers. The saddle points of the augmented Lagrangian function (24) are obtained by iteratively computing the expressions

$$\begin{aligned} \mathbf{v}^{t+1} &= \left(\mathbf{D}_k^T \mathbf{D}_k + \mu \mathbf{I}\right)^{-1} \left(\mathbf{D}_k^T \mathbf{z}^{(k,j)} + \mu \mathbf{c}^{(k,j)^t} + \frac{\mathbf{g}^t}{2}\right) \\ \mathbf{c}^{(k,j)^{t+1}} &= \text{soft}\left(\mathbf{v}^{t+1} - \frac{\mathbf{g}^t}{2\mu}, \frac{\lambda}{2\mu}\right) \\ \mathbf{g}^{t+1} &= \mathbf{g}^t + 2\mu \left(\mathbf{c}^{(k,j)^{t+1}} - \mathbf{v}^{t+1}\right) \end{aligned} \tag{25}$$

where $\text{soft}(a, b) = \text{sign}(a) \max(|a| - b, 0)$, and $\mathbf{I}$ is the identity matrix. Here, the term $(\mathbf{D}_k^T \mathbf{D}_k + \mu \mathbf{I})^{-1}$ in (25) can be computed very efficiently as follows:

$$\begin{aligned} \left(\mathbf{D}_k^T \mathbf{D}_k + \mu \mathbf{I}\right)^{-1} = &\left(\mathbf{P}_3 \otimes \mathbf{P}_2 \otimes \mathbf{P}_1\right) \\ &\left(\mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 + \mu \mathbf{I}\right)^{-1} \\ &\left(\mathbf{P}_3^T \otimes \mathbf{P}_2^T \otimes \mathbf{P}_1^T\right) \end{aligned} \tag{26}$$

where Σ_i and $\mathbf{P}_i$, for $i = 1, 2, 3$, are diagonal matrices and unitary matrices remaining the eigenvalues and eigenvectors of ${\mathbf{W}_k}^T \mathbf{W}_k$, ${\mathbf{H}_k}^T \mathbf{H}_k$, and $\mathbf{S}_k^{*T} \mathbf{S}_k^*$, respectively. Therefore, $(\mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 + \mu \mathbf{I})^{-1}$ is a diagonal matrix, and the multiplication is element wise. In addition, we note that the actions of $\mathbf{P}_i^T$ and $\mathbf{P}_i$ can be made by just i-mode tensor products. Finally, the term $\mathbf{D}_1^T \mathbf{y}$ in (25) can be calculated via the equation

$$\mathbf{D}_k^T \mathbf{z}^{(k,j)} = \text{vec}\left(\mathcal{Z}^{(k,j)} \times_1 \mathbf{W}_k^T \times_2 \mathbf{H}_k^T \times_3 \mathbf{S}_k^{*T}\right) \tag{27}$$

where $\text{vec}(\cdot)$ is the vectorization operation. In this way, all core tensors $\mathcal{C}^{(k,j)}$ for HR-HSI FBPs are estimated.

Once the sparse core tensors $\{\mathcal{C}^{(k,j)}\}_{j=1}^{n_k}$ and dictionaries $\mathbf{W}_k$, $\mathbf{H}_k$, and $\mathbf{S}_k$ are estimated, the HR-HSI FBPs $\{\mathcal{X}^{(k,j)}\}_{j=1}^{n_k}$ of the kth cluster are estimated via (14). Finally, the recovered FBP sets are returned to their locations and aggregated to form the HR-HSI $\mathcal{X}$ by straight averaging.

V. Experiments

A. Datasets

In this section, we evaluate the effectiveness of the NLSTF_SMBF approach by applying it to ground-based and remotely sensed hyperspectral data.

For the ground-based hyperspectral data, we conduct exhaustive experiments on the Columbia Computer Vision Laboratory (CAVE) dataset [43]. The CAVE dataset has 32 indoor HSIs captured by a generalized assorted pixel camera. The HSIs are of size $512 \times 512 \times 31$, which has 31 spectral bands and 512×512 spectral pixels. The bands are acquired with the range 400–700 nm and wavelength interval of 10 nm. We use the HSIs from the CAVE dataset as ground truth. In order to generate LR-HSIs, a Gaussian kernel of size 5×5 of standard variation 2 is first applied to the HR-HSIs, and then subsampled with a factor of 32 in both the width and the height modes, for each band of $\mathcal{X}$. In this way, the size of LR-HSIs is $16 \times 16 \times 31$ for the CAVE dataset. The 3-band HR-MSI $\mathcal{Z}$ of the same scenario is generated by subsampling $\mathcal{X}$ with the spectral model. The spectral subsampling matric is acquired from the response of a Nikon D700 camera.

For the remotely sensed sensing HSIs, we choose Pavia University [44] captured over the urban area of the University of Pavia. The HSI has a spatial resolution of 1.3 m and a spectral range of 0.43 and 0.86 μm. The image has the size of $610 \times 340 \times 115$ which has 115 bands and 610×340 spectral pixels. Some bands contain water vapor absorption, and the HSI is reduced to 93 bands after discarding these bands. The top left 256×256 spectral pixels are selected in the experiment for the convenience of the spatial downsampling process. The LR-HSI of size $64 \times 64 \times 93$ is generated by using a 5×5 Gaussian blur of standard variation 2, and by subsampling every 4 pixels in the width and height modes. We use the IKONOS-like reflectance spectral response filter [12]

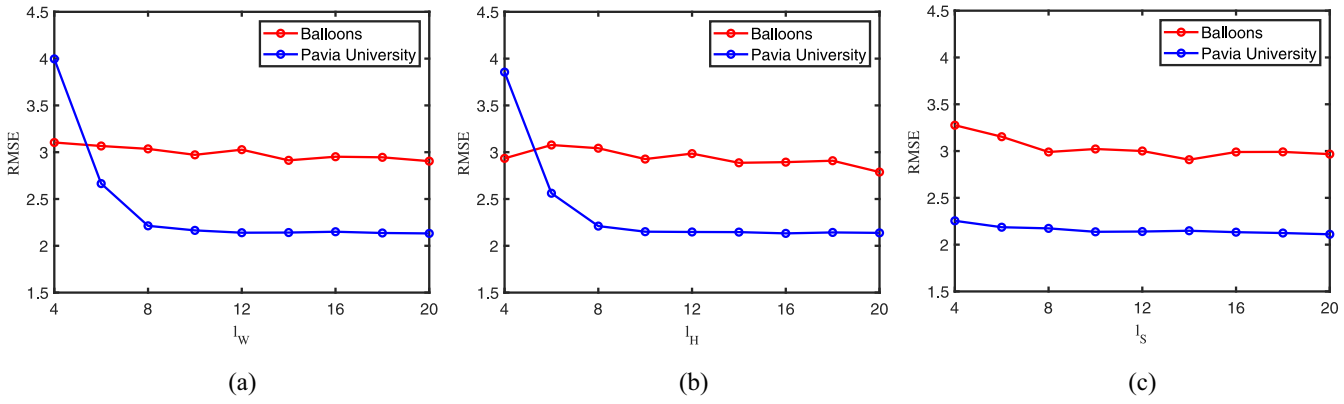


Fig. 3. RMSE curves for the proposed NLSTF_SMBF method as functions of the number atoms (a) l_W, (b) l_H, and (c) l_S.

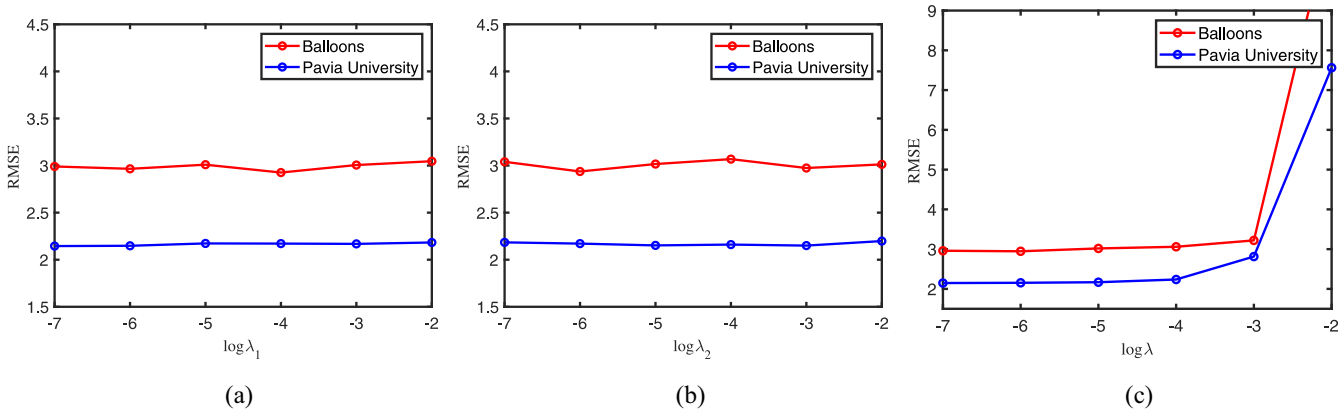


Fig. 4. RMSE curves for the proposed NLSTF_SMBF method as functions of parameters (a) λ_1, (b) λ_2, and (c) λ.

to simulate the HR-MSI with the size of $256 \times 256 \times 4$. The Gaussian noise is added to the HR-MSI (SNR $=$ 35 dB) and LR-HSI (SNR $=$ 30 dB), respectively.

The proposed NLSTF_SMBF approach is compared with the state-of-the-art HR-MSI and LR-HSI fusion methods, including the generalization of simultaneous orthogonal matching pursuit (GSOMP) [21], the Bayesian sparse representation (BSR) [22], the CSU [10], the nonlocal sparse representation method [13], and the coupled sparse tensor factorization [14]. GSOMP and BSR are the blind fusion methods and CSU, CSTF, and NSSR are the nonblind fusion methods.

B. Quantitative Metrics

Six quantitative indices are used in this article to measure the quality of recovered HSIs, including the root-mean-square error (RMSE), spectral angle mapper (SAM), relative dimensionless global error in synthesis (ERGAS) [45], and universal image quality index (UIQI) [46].

C. Parameter Selection

To evaluate the sensitivity of the proposed method with respect to its key parameters, the NLSTF_SMBF is run for distinct values of the number of atoms of dictionaries of the three modes l_W, l_H, l_S, the sparsity regularization parameters in the dictionary learning process λ_1, λ_2, the sparsity regularization parameter in the sparse coding process λ, and the number of cluster scaling parameter K.

Fig. 3 plots the RMSE of the reconstructed HSIs of *Balloons* (image in the CAVE dataset) and Pavia University as functions of the number of atoms l_W, l_H, and l_S of the dictionaries for the width, height, and spectral modes, respectively. From Fig. 3(a) and (b), we can see that the RMSE for Pavia University has a drop when l_W, l_H, and l_S vary from 4 to 14, and then it reaches the stable level. The RMSE for *Balloons* does not change obviously as l_W, l_H, and l_S vary from 4 to 10. Hence, we set $l_W = 10$, $l_H = 10$, and $l_S = 14$ for both the CAVE dataset and Pavia University.

Parameters λ_1 and λ_2 also have an important effect on the dictionary learning process, and λ controls the sparsity of the core tensor in the tensor sparse coding stage. Fig. 4 plots the RMSE of the recovered HSIs Pavia University and *Balloons* as the functions of $\log\lambda_1$, $\log\lambda_2$, and $\log\lambda$ (log is base 10). As can be seen from Fig. 4, the RMSE for *Balloons* changes little as $\log\lambda_1$, $\log\lambda_2$, and $\log\lambda$ vary from -7 to -3, and then it increases. The RMSE for Pavia University does not change significantly when $\log\lambda_1$, $\log\lambda_2$, and $\log\lambda$ vary from -7 to -2. Therefore, we set $\lambda_1 = 10^{-5}$, $\lambda_2 = 10^{-5}$, and $\lambda = 10^{-6}$ for both the CAVE and Pavia University databases.

The superiority of nonlocal information is mainly influenced by the number of clusters K. Fig. 5 plots the RMSE curves of the recovered HSIs *Balloons* and Pavia University as a function

TABLE I
QUANTITATIVE RESULTS FOR THE INVARIANT PSF CASE ON THE CAVE DATASET [43] AND PAVIA UNIVERSITY [44]

Methods	**CAVE dataset** [43]				**Pavia University** [44]			
	RMSE	SAM	ERGAS	UIQI	RMSE	SAM	ERGAS	UIQI
Best values	0	0	0	1	0	0	1	0
GSOMP [21]	5.30	13.86	0.713	0.808	3.84	3.43	2.267	0.979
BSR [22]	6.15	**12.87**	0.879	0.773	2.33	2.37	1.311	0.991
NLSTF_SMBF	**4.12**	12.99	**0.564**	**0.829**	**2.13**	**2.22**	**1.216**	**0.992**
CSU [10]	3.13	7.85	0.442	0.854	2.26	2.20	1.288	0.991
CSTF [14]	3.20	9.11	0.413	0.811	2.16	2.39	1.230	0.991
NSSR [13]	4.41	13.59	0.559	0.849	2.48	2.76	1.426	0.987

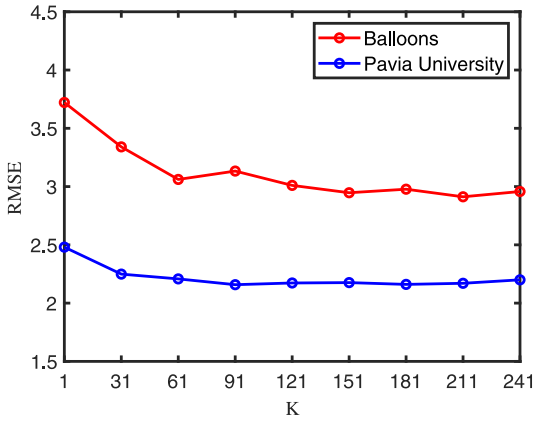


Fig. 5. RMSE curves for the proposed NLSTF_SMBF method as functions of the number of cluster K.

of the parameter K. When $K = 1$, we do not use the nonlocal cluster operation, and all FBPs of the HR-HSI belong to one group and are assumed to share the same dictionaries. It can be seen that the RMSE of all test images drops as K increases, which indicates that the nonlocal cluster process really works. When K reaches the values 151, the RMSE of the two test images reaches a relatively stable level. Therefore, the number of clusters is set to 160 for both two datasets. In addition, the spatial size of HR-HSI FBPs is 8×8 ($d_W = 8, d_H = 8$) with overlap 4 ($p = 4$).

At this point, we note that the proposed NLSTF_SMBF method does not make any assumption about the PSF of the hyperspectral sensor, although it assumes the knowledge of the spectral responses of the multispectral imaging sensor. For this reason, we label our approach as "semiblind." This semiblindness opens the door to solve the fusion problems in which the PSF of the hyperspectral sensor is spatially variant endowing NLSTF_SMBF with a very important form of model robustness. The two ensuing sections address invariant and variant PSF scenarios.

D. Experimental Results With Spatially Invariant PSF

In this section, we assume a spatially invariant PSF, which is assumed to be perfectly known for the CSU and CSTF nonblind fusion methods, and they incorporate the first equation in (9) for reconstruction. However, semiblind fusion methods GSOMP, BSR, and NLSTF_SMBF do not take advantage of the knowledge of the PSF and, therefore, this experimental setting favors the nonblind fusion methods.

Table I shows the values of the considered metrics for the CAVE dataset and the Pavia University dataset. We highlight the best results in bold for clarity in the table. NLSTF_SMBF performs clearly better than GSOMP and BSR, among the semiblind fusion methods. Specifically, the advantage of NLSTF_SMBF is considerable in terms of the RMSE, UIQI, and ERGAS, which means that the recovered HR-HSIs of the NLSTF_SMBF are closer to the ground truth. Since the PSF is assumed to be perfectly known, the nonblind fusion methods CSU, CSTF, and NSSR perform better than the semiblind fusion methods BSR and GSOMP. However, even under unfair conditions, NLSTF_SMBF still has superior performance over that of CSU, CSTF, and NSSR on Pavia University.

E. Experimental Results With Spatially Variant PSF

The PSF of a camera is often spatially variant, which further complicates the already challenging imaging inverse problems. A distinctive feature of NLSTF_SMBF is that it copes with spatially variant PSFs in a blind fashion. To illustrate the ability of NLSTF_SMBF to cope with spatially variant PSFs, we generate an LR-HSI from the HR-HSI using a spatially variant Gaussian filter of size 5×5 and a spatially variant standard deviation; we split the HR-HSI into $(W/4) \times (H/4)$ nonoverlapped FBPs, and set the spatial standard deviation of the Gaussian filter on each FBP to $0.5 + ([i + j]/4)$, where $(i, j) \in \{1, 2, 3, 4\}^2$ indices the 2-D FBPs. The blur is the same across the spectral bands in each FBP. The other settings are the same as for the invariant PSF case. For the nonblind fusion methods CSU and CSTF, the blur is assumed to be 5×5 Gaussian blur of standard variation 1.75 (average standard deviation of all FBPs).

Table II shows the quantitative results of the spatially variant PSF. The performance of the nonblind fusion methods is clearly reduced in the spatially variant PSF case; on the contrary, the semiblind fusion methods BSR and NLSTF_SMBF have similar (to the invariant scenario) performance in this case. The reason is that the semiblind fusion methods do not rely on the knowledge of the PSF. The proposed NLSTF_SMBF method performs the best in this case. Fig. 6 shows the false color images and SAM images computed from the estimated HSIs produced by the tested methods on the spatially invariant PSF case and spatially variant PSF case. The SAM images reflect the spectral errors associated with the fusion results. As can be seen from that figure, the nonblind fusion methods CSTF, CSU, and NSSR perform well on the

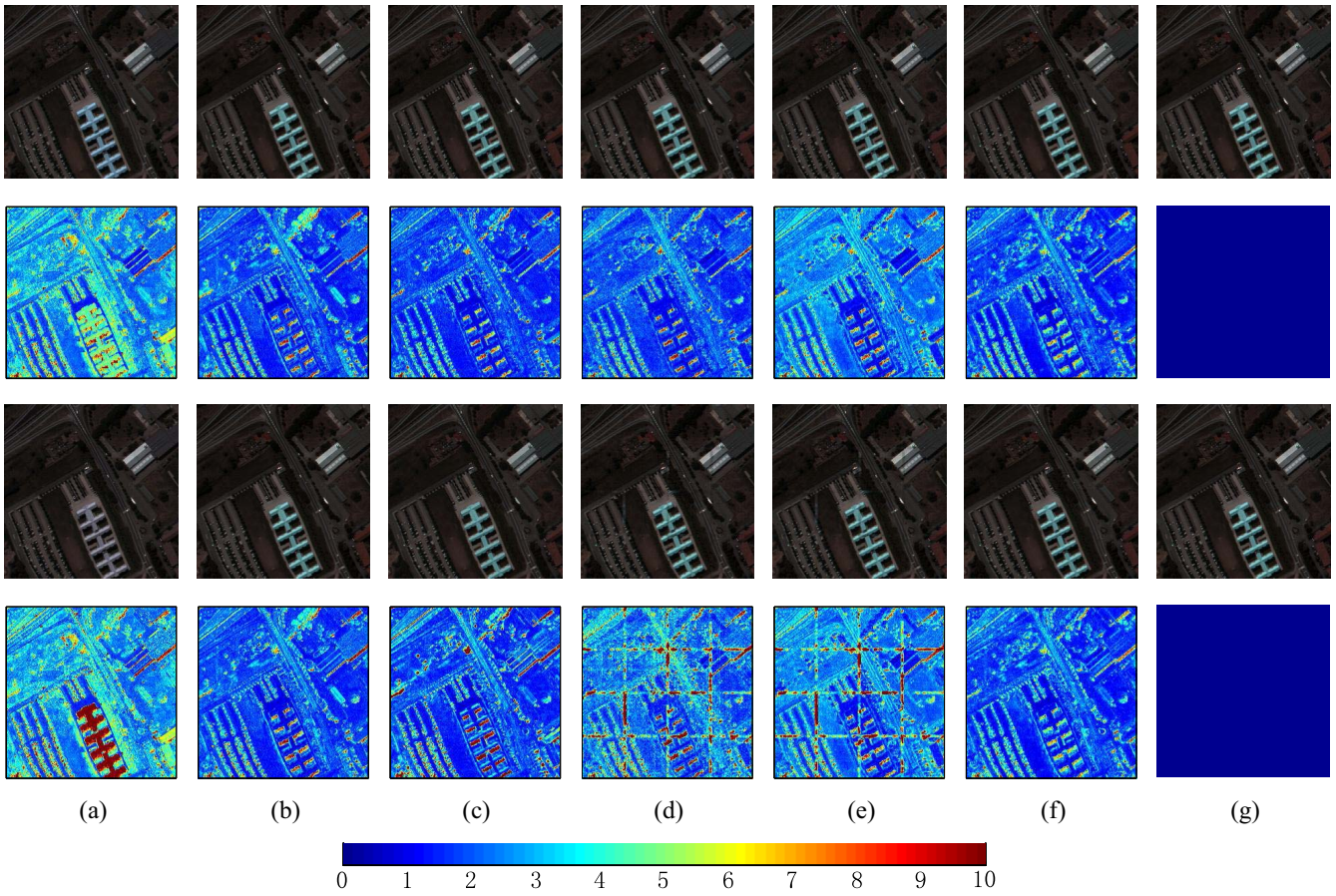


Fig. 6. First and second rows show the false color images (composed by bands 32, 13, 8) and SAM images of the testing methods on a spatially invariant PSF case. The third and fourth rows show the false color images (composed by bands 32, 13, 8) and SAM images of the testing methods on a spatially variant PSF case.

TABLE II
QUANTITATIVE RESULTS FOR THE VARIANT PSF CASE ON THE CAVE DATASET [43] AND PAVIA UNIVERSITY [44]

Methods	**CAVE dataset** [43]				**Pavia University** [44]			
	RMSE	SAM	ERGAS	UIQI	RMSE	SAM	ERGAS	UIQI
Best values	0	0	0	1	0	0	0	1
GSOMP [21]	5.52	14.24	0.731	0.807	5.52	3.85	3.439	0.973
BSR [22]	5.55	12.88	0.756	0.775	2.34	2.37	1.309	0.991
NLSTF_SMBF	**4.13**	14.48	**0.562**	0.826	**2.13**	**2.22**	**1.212**	**0.992**
CSU [10]	9.98	13.85	1.442	0.681	3.52	2.62	2.089	0.983
CSTF [14]	5.17	**11.54**	0.662	0.736	4.14	3.43	2.443	0.972
NSSR [13]	4.55	14.13	0.578	**0.840**	4.03	3.26	2.355	0.974

spatially invariant PSF case and have obvious flaws on the spatially variant PSF case. However, our method performs well on both the cases.

F. Experimental Results on Real Data

The real LR-HSI is captured by the Hyperion sensor onboard of Earth Observing-1 satellite. The LR-HSI has a spatial resolution of 30 m and 220 spectral bands in the spectral range of 400–2500 nm. After removing the bands of low SNR, 89 bands are preserved. An area of spatial size 80×80 is used in this experiment. The real HR-MSI is acquired by the Sentinel-2A satellite. It has 13 spectral bands, and we use the four bands with 10-m spatial resolution for the fusion. The central wavelengths of the four bands are 490, 560, 665, and 842 nm. The spatial size of the HR-MSI is 240×240. We estimate the spectral response $\mathbf{R}$ and convolution blur $\mathbf{B}$ via the method proposed in [11]. Fig. 7 shows the fused HR-HSI at the sixth band. As shown in the figure, all shown methods can visibly improve the spatial resolution of the observable LR-HSI. The fusion results of BSR and NLSTF_SMBF look better than the others. NLSTF_SMBF is, however, faster than BSR.

G. Effectiveness of Nonlocal Clustering

To verify the effectiveness of this nonlocal method, Table III reports the average quantitative metrics with/without the clustering strategy for the invariant PSF case on the CAVE dataset and Pavia University. When $K = 1$, we do not use the nonlocal cluster operation, and all FBPs of the HR-HSI belong to one group and are assumed to share the same dictionaries.

TABLE III
QUANTITATIVE RESULTS OF WITH/WITHOUT THE CLUSTERING STRATEGY FOR THE INVARIANT PSF CASE ON THE CAVE DATASET AND PAVIA UNIVERSITY

Methods	CAVE dataset				Pavia University			
	RMSE	SAM	ERGAS	UIQI	RMSE	SAM	ERGAS	UIQI
Best Values	0	0	0	1	0	0	0	1
NLSTF_SMBF (K=1)	5.54	21.52	0.679	0.825	2.44	2.62	1.369	0.990
NLSTF_SMBF ($K = 160$)	**4.12**	**12.99**	**0.564**	**0.829**	**2.13**	**2.22**	**1.216**	**0.992**

TABLE IV
COMPUTATIONAL TIME IN SECONDS OF THE COMPARED APPROACHES

Dataset	Methods					
	GSOMP	BSR	CSU	CSTF	NSSR	NLSTF_SMBF
CAVE dataset	676	8174	1111	649	**215**	261
Pavia University	371	1453	486	262	128	**74**
Hyperion dataset	1347	61048	**24**	216	92	55

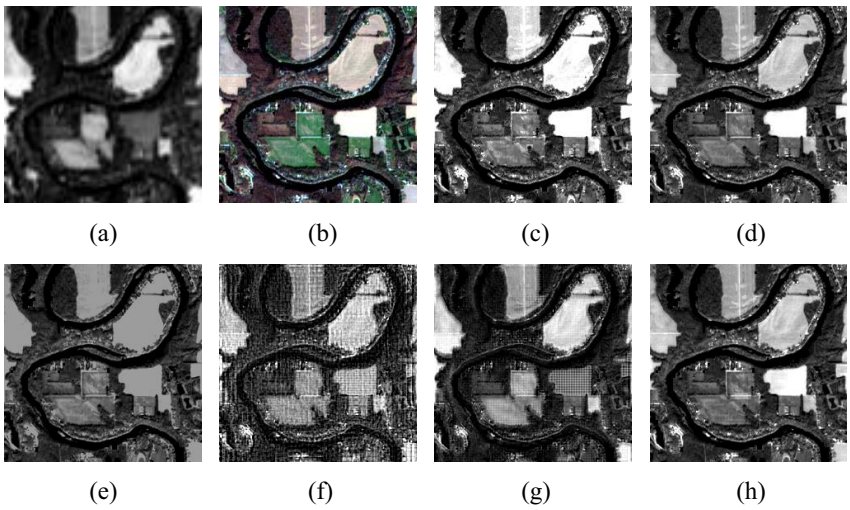


Fig. 7. Real data results for the sixth band of a HR-HSI. (a) Hyperion LR-HSI, (b) Sentinel-2A HR-MSI, (c) GSOMP [21], (d) BSR [22], (e) CSU [10], (f) CSTF [14], (g) NSSR [13], (h) NLSTF_SMBF.

We can see from the table that the clustering strategy clearly improves the performance on the CAVE and Pavia University datasets.

H. Running Time

All experiments are coded in MATLAB R2018b and run in a computer with an Intel Core-E5-2603 CPU with 1.6-GHz and 96-GB random access memory. Table IV shows the average running time on the CAVE dataset and Pavia University, respectively. In all the compared methods, the proposed NLSTF_SMBF has the speed advantage. Besides, since the NLSTF_SMBF method can be implemented for each cluster separately, the proposed NLSTF_SMBF can be further accelerated via parallel computing. To ensure fair time comparison, we have not considered parallel computing in calculating the running time of NLSTF_SMBF.

VI. CONCLUSION

In this article, we presented a novel NLSTF_SMBF-based framework to estimate an HR-HSI, by fusing an LR-HSI with an HR-MSI counterpart. Unlike recent matrix factorization-based HSI and MSI fusion methods, the proposed NLSTF_SMBF method considers each FBP of the HSI as a tensor with three modes and factorizes it as a sparse core tensor multiplication by dictionaries of the three modes. In addition, nonlocal spatial self-similarities are incorporated into the sparse tensor factorization. With the proposed framework, the HSI spatial–spectral information is fully exploited. Two distinctive features of NLSTF_SMBF is that it is blind with respect to the PSF of the hyperspectral sensor and copes with spatially variant PSFs. Our approach is compared with the state-of-the-art methods on ground-based and remotely sensed-based HSIs. The obtained results systematically outperformed the competitors, giving experimental evidence of the effectiveness of the proposed NLSTF_SMBF method.

ACKNOWLEDGMENT

The authors would like to thank the editors and reviewers for their outstanding comments and suggestions, which significantly improved this article.

REFERENCES

[1] Y. Zhou and Y. Wei, "Learning hierarchical spectral—Spatial features for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1667–1678, Jul. 2016.

[2] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003.

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

[4] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.

[5] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[6] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.

[7] R. Dian, S. Li, L. Fang, and Q. Wei, "Multispectral and hyperspectral image fusion with spatial-spectral sparse representation," *Inf. Fusion*, vol. 49, pp. 262–270, Sep. 2019.

[8] Y. Tang and Y. Yuan, "Learning from errors in super-resolution," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2143–2154, Nov. 2014.

[9] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[10] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 3586–3594.

[11] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

[12] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[13] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[14] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[15] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: Combining low rank tensor and matrix structure," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, 2018, pp. 3318–3322.

[16] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[17] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019, doi: 10.1109/TNNLS.2018.2885616.

[18] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2011, pp. 2329–2336.

[19] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 963–978, Sep. 2011.

[20] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1693–1704, Mar. 2014.

[21] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 63–78.

[22] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3631–3640.

[23] J. Chen, B. Jia, and K. Zhang, "Trifocal tensor-based adaptive visual trajectory tracking control of mobile robots," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3784–3798, Nov. 2017.

[24] B. Ma, L. Huang, J. Shen, and L. Shao, "Discriminative tracking using tensor pooling," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2411–2422, Nov. 2016.

[25] W. K. Wong, Z. Lai, Y. Xu, J. Wen, and C. P. Ho, "Joint tensor feature analysis for visual object recognition," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2425–2436, Nov. 2015.

[26] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 2949–2956.

[27] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[28] Y. Wu, H. Tan, Y. Li, J. Zhang, and X. Chen, "A fused CP factorization method for incomplete tensors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 751–764, Mar. 2019.

[29] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari, "Bayesian robust tensor factorization for incomplete multiway data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 736–748, Apr. 2016.

[30] L. Feng, H. Sun, Q. Sun, and G. Xia, "Compressive sensing via nonlocal low-rank tensor regularization," *Neurocomputing*, vol. 216, pp. 45–60, Dec. 2016.

[31] S. Yang, M. Wang, P. Li, L. Jin, B. Wu, and L. Jiao, "Compressive hyperspectral imaging via sparse tensor and nonlinear compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5943–5957, Nov. 2015.

[32] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal tensor sparse representation and low-rank regularization for hyperspectral image compressive sensing reconstruction," *Remote Sens.*, vol. 11, no. 2, p. 193, Jan. 2019.

[33] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep./Oct. 2009, pp. 2272–2279.

[34] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 3862–3871.

[35] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[36] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1996.

[37] C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using Kronecker bases," *Neural Comput.*, vol. 25, no. 1, pp. 186–220, Jan. 2013.

[38] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[39] Y. Rivenson and A. Stern, "Compressed imaging with a separable sensing operator," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 449–452, Jun. 2009.

[40] D. Arthur and S. Vassilvitskii, "*k*-means++: The advantages of careful seeding," in *Proc. Annu. ACM-SIAM Symp. Discr. Algorithm*, New Orleans, LA, USA, 2007, pp. 1027–1035.

[41] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Feb. 2011.

[42] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[43] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

[44] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.

[45] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Jan. 2000, pp. 99–103.

[46] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

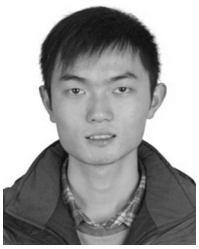
Renwei Dian (S'16) received the B.S. degree from the Wuhan University of Science and Technology, Wuhan, China, in 2015. He is currently pursing the Ph.D. degree with the Laboratory of Vision and Image Processing, Hunan University, Changsha, China.

From November 2017 to November 2018, he is a visiting Ph.D. student with the University of Lisbon, Lisbon, Portugal, supported by the China Scholarship Council. His research interests include hyperspectral image super-resolution, image fusion, tensor decomposition, and deep learning. More information can be found in his homepage https://sites.google.com/view/renweidian/.

Shutao Li (M'07–SM'15–F'19) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, He was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong. From November 2002 to November 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, London, U.K. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored over 200 refereed papers. His current research interests include image processing, pattern recognition, and artificial intelligence.

Prof. Li gained two second-Grade State Scientific and Technological Progress Awards of China in 2004 and 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. He is an Editorial Board Member of the Information Fusion and the Sensing and Imaging.



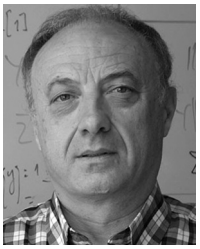
Leyuan Fang (S'10–M'14–SM'17) received the B.S. and Ph.D. degrees from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2008 and 2015, respectively.

From September 2011 to September 2012, he was a visiting Ph.D. student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. Since January 2017, he has been an Associate Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang has won the Scholarship Award for Excellent Doctoral Student granted by Chinese Ministry of Education in 2011.

Ting Lu (S'16–M'17) received the B.S. and Ph.D. degrees from Hunan University, Changsha, China, in 2011 and 2017, respectively.

From 2014 to 2015, she was a visiting Ph.D. student with the Department of Information Engineering and Computer Science, University of Trento, Trento, Italy, supported by the China Scholarship Council. Since 2017, she has been an Assistant Professor with the College of Electrical and Information Engineering, Hunan University. Her research interests include sparse representation, image fusion, and remote sensing image processing.



José M. Bioucas-Dias (S'87–M'95–SM'15–F'17) received the E.E., M.Sc., Ph.D., and Habilitation degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa (currently, Universidade de Lisboa), Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he is currently a Professor and teaches inverse problems in imaging and electric communications, and also a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing. He has introduced scientific contributions in the areas of imaging inverse problems, statistical image processing, optimization, phase estimation, phase unwrapping, and in various imaging applications, such as hyperspectral and radar imaging.

Prof. Bioucas-Dias was included in Thomson Reuters Highly Cited Researchers 2015 list and was a recipient of the IEEE GRSS David Landgrebe Award for 2017.