Information Fusion 49 (2019) 262–270

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Multispectral and hyperspectral image fusion with spatial-spectral sparse representation

Renwei Dian[a], Shutao Li[a,*], Leyuan Fang[a], Qi Wei[b]

[a] College of Electrical and Information Engineering, Hunan University, Changsha, China
[b] Siemens Corporate Research, Princeton, New Jersey, USA

ARTICLE INFO

Keywords:
Hyperspectral imaging
Hyperspectral image super-resolution
Spatial-spectral sparse representation
Image fusion

ABSTRACT

Fusing a high spatial resolution multispectral image (HR-MSI) with a low spatial resolution hyperspectral image (LR-HSI) of the same scenario to acquire a high spatial resolution hyperspectral image (HR-HSI) has recently attracted more and more attention. We propose a novel spatial-spectral sparse representation (SSSR) based approach for the fusion of an HR-MSI and an LR-HSI of the same scenario in this paper. In the proposed SSSR method, we formulate the fusion problem as the estimation of spectral basis and coefficients from the LR-HSI and HR-MSI. To better model the spatial and spectral characteristics of the HR-HSI, we incorporate the non-local spatial similarities, priors of the spectral unmixing, and a sparse prior to the fusion problem. Meanwhile, instead of keeping the spectral basis fixed, we design the alternative optimization algorithm for the estimation of spectral basis and coefficients, which can achieve the accurate reconstruction. Experimental results on both non-blind fusion and blind fusion cases demonstrate the effectiveness of the SSSR approach.

1. Introduction

Hyperspectral imaging is a promising imaging technique, which can simultaneously acquire images of the same scenario across a number of different wavelengths. Rich spectral characteristics existed in hyperspectral images (HSIs) with a number of spectral bands have shown performance improvement in numerous remote sensing [1,2] and medical imaging [3,4] applications. However, the spreading of sun photons over many spectral bands determines large spectral resolution cells, and thus the hyperspectral imaging sensors can only provide a low spatial resolution for the HSIs, to ensure a high signal-to-noise-ratio (SNR) [5]. Compared with hyperspectral imaging sensors, the existing multispectral imaging sensors can capture an MSI with much higher spatial resolution and SNR given the same exposure time [6]. Therefore, an effective way to reconstruct a high spatial resolution hyperspectral image (HR-HSI) is combining a low spatial resolution hyperspectral image (LR-HSI) with a high spatial resolution multispectral image (HR-MSI) [7]. This combination is called as HSI-MSI fusion or HSI super-resolution [8,9], which belongs to the pixel-level image fusion [10–18]. Pixel-level image fusion is the technique to combine different images into a fused image of better quality [7]. The pixel-level image fusion approaches have shown considerable improvement in medical imaging [19,20], night vision [21], and remote sensing [22] applications.

To acquire an HR-HSI, one class of methods, called as the pan-sharpening [22], fuse a high spatial resolution panchromatic (PAN) image (gray image) with an LR-HSI. The pan-sharpening methods can be categorized two classes, i.e., transform based methods [23–25] and variational methods [26–29], etc. Since a panchromatic image has little spectral resolution, the reconstructed HR-HSIs by these approaches usually contain considerable spectral distortions.

In spatial-spectral image fusion, another class of methods uses Bayesian framework [30–32] to fuse an LR-HSI and an HR-MSI. Based on prior knowledge, these Bayesian fusion methods integrate posterior distribution to achieve accurate estimation. Hardie et al. [30] apply maximum a posteriori (MAP) based framework for fusion of an LR-HSI and an HR-MSI. Recently, Wei et al. [31] formulate the fusion process via the likelihoods of the observations, and achieve fast fusion by solving a Sylvester equation. Akhtar et al. [32] exploit Bayesian sparse representation to solve the fusion problem. Firstly, the probability distributions of spectral basis are inferred with the Beta process, and then the acquired distributions are utilized to calculate sparse coefficients of the HR-HSI.

Recently, utilizing the matrix factorization to fuse a LR-HSI with a HR-MSI has been actively investigated [33–42]. Assuming that the HR-HSI only contains a small number of pure spectral signatures [43], the HR-HSI can be approximated by the spectral basis multiplied by the coefficient, as can be seen from Fig. 1. Kawakami et al. [33] propose to learn the spectral basis from the LR-HSI with a sparse prior, and then conduct sparse coding on the HR-MSI to estimate the coefficients. Then, Huang et al. utilize the sparse prior [34] for the fusion of remote sensed

* Corresponding author.
E-mail address: shutao_li@hnu.edu.cn (S. Li).
https://doi.org/10.1016/j.inffus.2018.11.012
Received 23 December 2017; Received in revised form 29 June 2018; Accepted 25 November 2018
Available online 31 January 2019
1566-2535/© 2019 Elsevier B.V. All rights reserved.

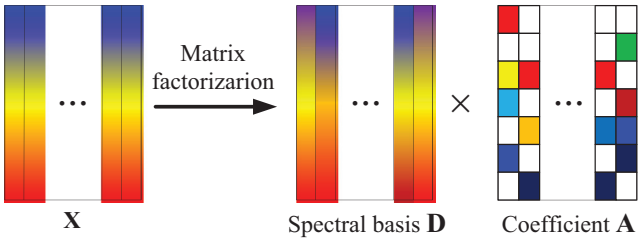

Fig. 1. Spectral unmixing based HR-HSI decomposition.

HSIs, which use the KSVD algorithm [44] to estimate the spectral basis. Furthermore, methods proposed in [36–38] use the priors of spectral unmixing to regularize the fusion problem. Instead of using fixed spectral basis, these methods alternately update spectral basis and coefficients with non-negative and sum-to-one constraints, which yields a more accurate fusion result. To better use the prior information of the HR-HSI, approaches proposed in [39–42] also utilize the spatial structures of the HR-HSI to regularize the fusion problem. For example, Akhtar et al. [39] acquire coefficients with the simultaneous greedy pursuit algorithm for each local patch, which utilizes the similarities of spectral pixels in the local patch in the HR-HSI. Furthermore, Dong et al. [42] propose a nonnegative dictionary-learning algorithm to learn the spectral basis and utilize the structured sparse coding approach to estimate the coefficients. They fully exploit the non-local spatial similarities of the HR-HSI, thus achieving the good fusion results. Although the above matrix factorization based methods achieve state-of-the-art performance, they only utilize sparse, spectral or spatial priors of the HR-HSI, do not fully use these three kinds of priors.

Since deep convolutional neural network (CNN) with deep architecture has demonstrated to be very effective to exploit image characteristics, one class of methods use the deep CNN for HSI-MSI fusion. Work [45] propose a deep residual learning based framework for the fusion, which uses the learned priors by deep CNN to regularize the fusion problem. In addition, Yang et al. [46] proposes the CNN with two branches for HSI-MSI fusion. In order to exploit spectral correlation and fuse the HR-MSI, they extract the features from the spectrum of each pixel in low resolution HSI, and its corresponding spatial neighborhood in MSI, with the two-branches CNN.

In this paper, a spatial-spectral sparse representation method is proposed to reconstruct an HR-HSI by fusing an LR-HSI and an HR-MSI. Specifically, the fusion problem is formulated as the estimation of the spectral basis and sparse coefficients from the LR-HSI and HR-MSI by exploiting the non-local spatial self-similarities, priors knowledge of the spectral unmixing, and a sparse prior. A typical natural scenario is often self-similar, and therefore it usually contains similar pixels from all over the image, which has been proved to be effective for the image restoration [42,47]. Besides, based on the spectral mixture model [43], the spectral basis and coefficients are non-negative, and the coefficients often satisfy the sum-to-one constraint. Furthermore, under appropriate spectral basis, the coefficients can be sparse, which is helpful in many HSI reconstruction problems [48]. These spatial, spectral, and sparse priors of the HSI are used for regularizing the fusion problem to achieve exact reconstruction. Besides, we design alternative optimization algorithm to update the spectral basis and coefficients from the LR-HSI and HR-HSI. Both of the updating for the spectral basis and coefficients can be formulated as convex optimization problems, thus the framework of alternating direction method of multipliers (ADMM) [49] can be used to solve them efficiently.

We organize the remaining parts of this paper as follows. Section 2 formulates the problem of the HSI-MSI fusion as a constrained optimization problem. The proposed SSSR approach for HSI-MSI fusion is presented in Section 3. We give the experimental results on two HSI data sets in Section 4. Section 5 presents the conclusions.

2. Problem formulation

In this paper, the HR-HSI, LR-HSI, and HR-MSI are denoted as the matrices. The first dimension of the matric stands for the number of bands, and the second dimension denotes the number of pixels. The desired HR-HSI is denoted by $\mathbf{X} \in \mathbb{R}^{S\times N}$, where N and S represent the number of pixels and spectral bands, respectively. $\mathbf{Y} \in \mathbb{R}^{S\times n}$ represents the acquired LR-HSI, where n and S denote the number of pixels and bands in the LR-HSI, respectively. $\mathbf{Y}$ is spatially downsampled, i.e., $N > n$, and has the same quantity of spectral bands as $\mathbf{X}$. $\mathbf{Z} \in \mathbb{R}^{s\times N}$ represents the HR-MSI of the same scenario, where s and N are the number of bands and pixels in the HR-MSI, respectively. With respect to $\mathbf{X}$, $\mathbf{Z}$ has the same number of pixels N and is the spectrally downsampled, i.e., $S > s$.

In the linear unmixing model, each pixel of the target HR-HSI $\mathbf{x}_i$ is assumed to be represented as the linear combination of distinct spectral signatures [43], which means

$$\mathbf{x}_i = \mathbf{D}\mathbf{a}_i, \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{S\times L}$ represents the spectral basis with L atoms, and $\mathbf{a}_i \in \mathbb{R}^L$ denotes the corresponding coefficient. Accounting for pixels of the whole HSI, the Eq. (1) is equivalent to

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] = \mathbf{D}[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] = \mathbf{D}\mathbf{A}, \tag{2}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] \in \mathbb{R}^{L\times N}$ is the coefficients.

The LR-HSI $\mathbf{Y}$ acquired by hyperspectral imaging sensor, can be formulated as the downsampled version of $\mathbf{X}$ with the spatial mode, i.e.,

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{S}, \tag{3}$$

where $\mathbf{B} \in \mathbb{R}^{N\times N}$ is the blur matrix, which models a convolution between the HR-HSI bands (represented by the rows of $\mathbf{X}$) and point spread function (PSF) of the imaging sensor, and $\mathbf{S} \in \mathbb{R}^{N\times n}$ is the spatial downsampling matrix.

The HR-MSI $\mathbf{Z}$ acquired by multispectral imaging sensor can be formulated as the downsampled version of $\mathbf{X}$ with the spectral mode,

$$\mathbf{Z} = \mathbf{R}\mathbf{X}, \tag{4}$$

where $\mathbf{R} \in \mathbb{R}^{s\times S}$ is the spectral downsampling matrix. The rows of matrix $\mathbf{R}$ represent the response of multispectral imaging sensor.

Combining the linear mixture model (2) with imaging models (3) and (4) leads to

$$\mathbf{Y} = \mathbf{D}\mathbf{A}\mathbf{B}\mathbf{S}, \quad \mathbf{Z} = \mathbf{R}\mathbf{D}\mathbf{A}, \tag{5}$$

In this formulation, the target of fusion is estimating the spectral basis $\mathbf{D}$ and corresponding coefficients $\mathbf{A}$ from the LR-HSI $\mathbf{Y}$ and HR-MSI $\mathbf{Z}$.

3. Proposed SSSR approach

As mentioned above, the fusion problem can be transformed as the estimation of spectral basis $\mathbf{D}$ and coefficients $\mathbf{A}$ from the model (5). Approaches proposed in [33,34,36,37,39] assume the spatial and spectral information mainly exists in the LR-HSI and HR-HSI, respectively. Therefore, they estimate the spectral basis $\mathbf{D}$ and coefficients $\mathbf{A}$ from the LR-HSI and HR-MSI, respectively. However, the HR-MSI and LR-HSI still preserve the spectral information and spatial information, respectively. Therefore, we jointly estimate the spectral basis and coefficients from both LR-HSI and HR-MSI, which can achieve more accurate reconstruction. In this way, the fusion problem can be written as

$$\min_{\mathbf{D},\mathbf{A}} ||\mathbf{Y} - \mathbf{D}\mathbf{A}\mathbf{B}\mathbf{S}||_F^2 + ||\mathbf{Z} - \mathbf{R}\mathbf{D}\mathbf{A}||_F^2. \tag{6}$$

The optimizations for $\mathbf{D}$ and $\mathbf{A}$ are severely ill-posed, and they do not have the unique solution. Therefore, we need to use some prior information of the unknown HR-HSI to regularize it. Multiple image priors have been exploited to constraint the optimization problem, such as a sparse prior [33–35,50], priors of spectral unmixing [36–38], and spatial structures prior [39–42]. In this paper, we incorporate three important priors,

i.e., a sparse prior, non-local spatial similarities, and priors of spectral unmixing, into a unified framework.

3.1. Priors

Sparsity prior has shown to be very effective for dealing with various HSI reconstruction problems [33,42,48], which assume each spectral pixel in the HSI can be expressed as the linear combination of a few distinct spectral signatures. In this way, with appropriate spectral basis, each column of the coefficients $\mathbf{A}$ can be sparse. We can use ℓ_0 norm to describe sparsity. However, ℓ_0 norm constraint optimization problem is NP-hard, and it can be approximated with ℓ_1 norm [51]. To exploit the sparsity prior, the fusion problem can be written as

$$\min_{\mathbf{D},\mathbf{A}} ||\mathbf{Y}-\mathbf{DABS}||_F^2+||\mathbf{Z}-\mathbf{RDA}||_F^2+\eta_1||\mathbf{A}||_1 \tag{7}$$

where $||\mathbf{A}||_1$ stands for the sum of absolute values of all elements of $\mathbf{A}$, and η_1 is the regularization parameter.

A typical natural scene usually contains a collection of similar pixels from all over the image, and these non-local similarities have shown to be very effective for image recovery [42,52]. To exploit the non-local similarities in the HR-HSI, we assume that a pixel $\mathbf{x}_i$ in the HR-HSI can be approximated by a linear combination of pixels, which are similar to $\mathbf{x}_i$. Incorporating this prior, a non-local sparse representation based fusion problem can be formulated as

$$\min_{\mathbf{D},\mathbf{A}} ||\mathbf{Y}-\mathbf{DABS}||_F^2+||\mathbf{Z}-\mathbf{RDA}||_F^2+\eta_1||\mathbf{A}||_1+\eta_2\sum_{i=1}^{N}||\mathbf{Da}_j-\mathbf{c}_i||_F^2 \tag{8}$$

where η_2 is the regularization parameter. $\mathbf{c}_i$ represents the linear combinations of pixels, which are similar to $\mathbf{x}_i$. The vector $\mathbf{c}_i$ can be computed as

$$\mathbf{c}_i=\sum_{j\in S_i} w_{ij}\mathbf{Da}_j \tag{9}$$

where w_{ij} is the weighting coefficients based on the similarity between the pixels x_i and x_j, and S_i denotes the indexes of the similar pixels. The pixels x_i and x_j are not known, and we can compute the weighting coefficients on the HR-MSI, since the HR-MSI preserves the most spatial information of the HR-HSI. More specifically, we perform k-NN clustering on the HR-MSI to search for k-nearest neighbors neighbors for each pixel $\mathbf{z}_i$. The weighting coefficients w_{ij} can be calculated as

$$w_{ij}=\frac{1}{c}\exp(-||\mathbf{z}_i-\mathbf{z}_j||_F^2/h) \tag{10}$$

where c is the normalization constant, and $\mathbf{z}_i$ and $\mathbf{z}_j$ denote the pixels of the HR-MSI. In practice, coefficient vector $\mathbf{a}_i$ is unknown, and we can not compute $\mathbf{c}_i$ directly using the Eq. (9). We overcome this difficulty by iteratively computing $\mathbf{c}_i$ from the previous update of a_i. Taking pixels of the whole image into consideration, the problem (8) can be rewritten as

$$\min_{\mathbf{D},\mathbf{A}} ||\mathbf{Y}-\mathbf{DABS}||_F^2+||\mathbf{Z}-\mathbf{RDA}||_F^2+\eta_1||\mathbf{A}||_1+\eta_2||\mathbf{DA}-\mathbf{C}||_F^2 \tag{11}$$

where $\mathbf{C}=[\mathbf{c}_1,\mathbf{c}_2,\dots,\mathbf{c}_N]$. The Eq. (9) can be equivalently written as the following matrix formulation:

$$\mathbf{C}=\mathbf{WDA} \tag{12}$$

where the element of matrix $\mathbf{W}$ is computed as

$$\mathbf{W}(i,j)=\begin{cases} w_{ij}, & j\in S_i \\ 0, & otherwise \end{cases} \tag{13}$$

Besides, based on the priors of spectral unmixing, the coefficient vectors satisfy the non-negativity and sum-to-one constraints [43]:

$$\mathbf{a}_j\geq 0,\ \mathbf{1}_L^T\mathbf{a}_j=1 \tag{14}$$

where $\mathbf{a}_j$ denotes j^{th} column of $\mathbf{A}$, and $\mathbf{1}_L\in\mathbb{R}^L$ is a vector with all ones. We omit the sum-to-one constraint since it contradicts ℓ_1 regularizer that appears in the cost function of the optimization problem. Taking pixels of the whole image into consideration, the non-negative constraint in (14) can also be written as

$$\mathbf{A}\geq 0 \tag{15}$$

As the spectral basis represents the reflectance of distinct materials, each element of the spectral basis is in the range of [0,1]. Therefore, the constraint for $\mathbf{D}$ can be represented as

$$0\leq\mathbf{D}\leq 1 \tag{16}$$

To incorporate all the priors above, the fusion problem can be written as

$$\begin{aligned}\min_{\mathbf{D},\mathbf{A}}\ & ||\mathbf{Y}-\mathbf{DABS}||_F^2+||\mathbf{Z}-\mathbf{RDA}||_F^2+\eta_1||\mathbf{A}||_1+\eta_2||\mathbf{DA}-\mathbf{C}||_F^2, \\ & \text{s.t. } 0\leq\mathbf{D}\leq 1 \text{ and } \mathbf{A}\geq 0,\end{aligned} \tag{17}$$

In this formulation, the priors of sparsity, spectral unmixing and non-local spatial similarities are incorporated into a unified framework.

3.2. Alternating optimization of the fusion problem

The optimization problem (17) is highly non-convex, and the solution is not unique. Approaches proposed in [33,34,36,37,39–42] initialize $\mathbf{D}$, and estimate $\mathbf{A}$ with $\mathbf{D}$ fixed. In this way, $\mathbf{D}$ is not updated, and the result relies on the initializations of $\mathbf{D}$. Realizing the problem (17) is convex with respect to $\mathbf{D}$ and $\mathbf{A}$, respectively, we propose an optimization technique that alternates optimizations with respect to $\mathbf{D}$ and $\mathbf{A}$ to overcome this problem. In specific, we update $\mathbf{A}$ with $\mathbf{D}$ fixed, and then update $\mathbf{D}$ with $\mathbf{A}$ fixed. The above two steps are iterated for convergence. Finally, the desired HR-HSI can be estimated via the Eq. (2). The overall algorithm for the fusion of an LR-HSI and an HR-MSI is summarized in Algorithm 1.

Algorithm 1 SSSR-based HSI-MSI fusion.

Input: $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$
Output: $\mathbf{X}$
Initialization:
Initialize the spectral basis $\mathbf{D}$
learn non-local self-similarities matrix $\mathbf{W}$ via KNN clustering from $\mathbf{Z}$
$\mathbf{A}=0$
For $t=0,1,\cdots,T_1$
(a) Estimate the coefficients $\mathbf{A}$ by solving the problem (18). (see Algorithm 2)
(b) Estimate the spectral basis $\mathbf{D}$ by solving the problem (24). (see Algorithm 3)
End
Compute the HR-HSI $\mathbf{X}$ via the Eq. (2).

3.2.1. Optimization with respect to the coefficients

With the spectral basis $\mathbf{D}$ fixed, the update for coefficients $\mathbf{A}$ can be written as

$$\begin{aligned}\min_{\mathbf{A}}\ & ||\mathbf{Y}-\mathbf{DABS}||_F^2+||\mathbf{Z}-\mathbf{RDA}||_F^2+\eta_1||\mathbf{A}||_1+\eta_2||\mathbf{DA}-\mathbf{C}||_F^2, \\ & \text{s.t. } \mathbf{A}\geq 0,\end{aligned} \tag{18}$$

The optimization problem (18) can be tackled efficiently via the framework of ADMM [49], which can separate the constraint and objective function. In this way, the problem (18) can be decomposed as several sub-problems, and each sub-problem can be solved analytically. We introduce $\mathbf{V}_1=\mathbf{A}$ and $\mathbf{V}_2=\mathbf{DA}$, and obtain the following augmented Lagrangian function:

$$\begin{aligned}L(\mathbf{A},\mathbf{V}_1,\mathbf{V}_2,\mathbf{G}_1,\mathbf{G}_2)=\ & ||\mathbf{Y}-\mathbf{V}_2\mathbf{BS}||_F^2+||\mathbf{Z}-\mathbf{RDA}||_F^2+\eta_1||\mathbf{V}_1||_1 \\ & +\eta_2||\mathbf{DA}-\mathbf{C}||_F^2 \\ & +\mu\left\|\mathbf{V}_1-\mathbf{A}+\frac{\mathbf{G}_1}{2\mu}\right\|_F^2+\mu\left\|\mathbf{V}_2-\mathbf{DA}+\frac{\mathbf{G}_2}{2\mu}\right\|_F^2, \\ & \text{s.t. } \mathbf{V}_1\geq 0,\end{aligned} \tag{19}$$

where $\mathbf{G}_1$ and $\mathbf{G}_2$ are Lagrangian multipliers ($\mu > 0$).

(1): Minimizing the augmented Lagrangian function results in the following iterations:

$$\begin{aligned}
\mathbf{A}^{(t+1)} &= \underset{\mathbf{A}}{\operatorname{argmin}}\, L\left(\mathbf{A}, \mathbf{V}_1^{(t)}, \mathbf{V}_2^{(t)}, \mathbf{G}_1^{(t)}, \mathbf{G}_2^{(t)}\right) \\
\mathbf{V}_2^{(t+1)} &= \underset{\mathbf{V}_2}{\operatorname{argmin}}\, L\left(\mathbf{A}^{(t+1)}, \mathbf{V}_1^{(t)}, \mathbf{V}_2, \mathbf{G}_1^{(t)}, \mathbf{G}_2^{(t)}\right) \\
\mathbf{V}_1^{(t+1)} &= \underset{\mathbf{V}_1}{\operatorname{argmin}}\, L\left(\mathbf{A}^{(t+1)}, \mathbf{V}_1, \mathbf{V}_2^{(t+1)}, \mathbf{G}_1^{(t)}, \mathbf{G}_2^{(t)}\right)
\end{aligned} \tag{20}$$

All sub-problems in (20) can be solved analytically, i.e.

$$\begin{aligned}
\mathbf{A}^{(t+1)} &= [(\mathbf{RD})^T\mathbf{RD} + (\eta_2+\mu)\mathbf{D}^T\mathbf{D} + \mu\mathbf{I}]^{-1}\left[(\mathbf{RD})^T\mathbf{Z} + \eta_2\mathbf{D}^T\mathbf{C}^{(t)}\right. \\
&\quad \left. + \mu\left(\mathbf{V}_1^{(t)} + \frac{\mathbf{G}_1^{(t)}}{2\mu}\right) + \mu\mathbf{D}^T\left(\mathbf{V}_2^{(t)} + \frac{\mathbf{G}_2^{(t)}}{2\mu}\right)\right] \\
\mathbf{V}_2^{(t+1)} &= \left[\mathbf{Y}(\mathbf{BS})^T + \mu\left(\mathbf{DA}^{(t+1)} - \frac{\mathbf{G}_2^{(t)}}{2\mu}\right)\right][\mathbf{BS}(\mathbf{BS})^T + \mu\mathbf{I}]^{-1} \\
\mathbf{V}_1^{(t+1)} &= \max\left(\operatorname{soft}\left(\mathbf{A}^{(t+1)} - \frac{\mathbf{G}_1^{(t)}}{2\mu}, \frac{\eta_1}{2\mu}\right), 0\right)
\end{aligned} \tag{21}$$

where

$$\operatorname{soft}(a, b) = \operatorname{sign}(a) * \max(|a| - b, 0) \tag{22}$$

and $\max(a, b)$ represents the maximum of a and b.

(2): The Lagrangian multipliers $\mathbf{G}_1$ and $\mathbf{G}_2$ are updated by

$$\begin{aligned}
\mathbf{G}_1^{(t+1)} &= \mathbf{G}_1^{(t)} + 2\mu(\mathbf{V}_1^{(t+1)} - \mathbf{A}^{(t+1)}) \\
\mathbf{G}_2^{(t+1)} &= \mathbf{G}_2^{(t)} + 2\mu(\mathbf{V}_2^{(t+1)} - \mathbf{DA}^{(t+1)})
\end{aligned} \tag{23}$$

(3): The matrix $\mathbf{C}^{(t+1)}$ can be updated via the Eq. (12).

The above steps are iterated until convergence. The overall algorithm for the estimation of coefficients $\mathbf{A}$ is summarized in Algorithm 2.

Algorithm 2 Estimate $\mathbf{A}$ with $\mathbf{D}$ fixed.

Input: $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{D}$, $\mathbf{A}$
Output: $\mathbf{A}$
Initialization:
$\mathbf{V}_1 = \mathbf{A}$, $\mathbf{V}_2 = \mathbf{DA}$, $\mathbf{G}_1 = 0$, $\mathbf{G}_2 = 0$, $\mathbf{C} = \mathbf{WDA}$
For $t = 0, 1, \cdots, T_2$
(1) Compute $\mathbf{A}^{(t+1)}$, $\mathbf{V}_1^{(t+1)}$, $\mathbf{V}_2^{(t+1)}$ via the equation (21)
(2) Update the Lagrangian multipliers $\mathbf{G}_1^{(t+1)}$ and $\mathbf{G}_2^{(t+1)}$ via the equation (24)
(3) Update $\mathbf{C}^{(t+1)}$ via the equation (12);
End

3.2.2. Optimization with respect to the spectral basis

With the coefficients $\mathbf{A}$ fixed, the update for spectral basis $\mathbf{D}$ can be written as

$$\begin{aligned}
&\min_{\mathbf{D}} ||\mathbf{Y} - \mathbf{DABS}||_F^2 + ||\mathbf{Z} - \mathbf{RDA}||_F^2, \\
&\quad \text{s.t. } 0 \le \mathbf{D} \le 1,
\end{aligned} \tag{24}$$

Since the non-local spatial correlations of the spectral pixels are mainly reflected by the coefficients $\mathbf{A}$, the term $\eta_2||\mathbf{DA} - \mathbf{C}||_F^2$ is excluded. The problem (24) can also be solved by the ADMM algorithm. Specifically, we introduce the variable $\mathbf{V}_3 = \mathbf{D}$, and acquire the following augmented Lagrangian:

$$\begin{aligned}
L(\mathbf{D}, \mathbf{V}_3, \mathbf{G}_3) = \|\mathbf{Y} - \mathbf{D}\widetilde{\mathbf{A}}\|_F^2 + \|\mathbf{Z} - \mathbf{RDA}\|_F^2 + \mu\left\|\mathbf{V}_3 - \mathbf{D} + \frac{\mathbf{G}_3}{2\mu}\right\|_F^2, \\
\text{s.t. } 0 \le \mathbf{V}_3 \le 1,
\end{aligned} \tag{25}$$

where $\widetilde{\mathbf{A}} = \mathbf{ABS}$ is the dowmsampled coefficient, and $\mathbf{G}_3$ is the Lagrangian multiplier.

The optimization for (25) can be decomposed as solving the following sub-problems:

1) Updating $\mathbf{D}^{(t+1)}$: With $\mathbf{V}_3$ and $\mathbf{G}_3$ fixed, the optimization for $\mathbf{D}^{(t+1)}$ in (25) can be solved analytically, we can acquire the following equation:

$$\mathbf{D}^{(t+1)}\mathbf{H}_1 + \mathbf{H}_2\mathbf{D}^{(t+1)} = \mathbf{H}_3 \tag{26}$$

where

$$\begin{aligned}
\mathbf{H}_1 &= \left(\widetilde{\mathbf{A}}\widetilde{\mathbf{A}}^T + \mu\mathbf{I}\right)\left(\mathbf{AA}^T\right)^{-1} \\
\mathbf{H}_2 &= \mathbf{R}^T\mathbf{R} \\
\mathbf{H}_3 &= \left(\mathbf{Y}\widetilde{\mathbf{A}}^T + \mathbf{R}^T\mathbf{ZA}^T + \mu\left(\mathbf{V}_3^{(t)} + \frac{\mathbf{G}_3^{(t)}}{2\mu}\right)\right)(\mathbf{AA}^T)^{-1}
\end{aligned} \tag{27}$$

Vectorizing $\mathbf{D}^{(t+1)}$ and $\mathbf{H}_3$, the Eq. (26) can also be written as

$$(\mathbf{H}_1^T \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{H}_2)\operatorname{vec}(\mathbf{D}^{(t+1)}) = \operatorname{vec}(\mathbf{H}_4) \tag{28}$$

where $\otimes$ denotes the Kronecker product, and $\operatorname{vec}(\cdot)$ is the vectorization operation that stacks all columns of matrix into a vector. Hence, $\mathbf{D}^{(t+1)}$ can be computed as

$$\operatorname{vec}(\mathbf{D}^{(t+1)}) = (\mathbf{H}_1^T \otimes \mathbf{I} + \mathbf{H}_3^T \otimes \mathbf{H}_2)^{-1}\operatorname{vec}(\mathbf{H}_4) \tag{29}$$

Reference [38] introduces a more efficient way to compute $\mathbf{D}^{(t+1)}$. Since $\mathbf{H}_1$ and $\mathbf{H}_2$ are real symmetric matrices, they can be diagonalized by eigendecomposition, that is, $\mathbf{H}_1 = \mathbf{G}_1\mathbf{\Sigma}_1\mathbf{G}_1^{-1}$ and $\mathbf{H}_2 = \mathbf{G}_2\mathbf{\Sigma}_2\mathbf{G}_2^{-1}$, where $\mathbf{\Sigma}_1 = \operatorname{diag}(a_1, a_2, \ldots, a_L)$ and $\mathbf{\Sigma}_2 = \operatorname{diag}(b_1, b_2, \ldots, b_S)$ are diagonal matrices. In this way, the Eq. (26) is equivalent to

$$\widetilde{\mathbf{D}}^{(t+1)}\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\widetilde{\mathbf{D}}^{(t+1)} = \mathbf{G}_2^{-1}\mathbf{H}_3\mathbf{G}_1 \tag{30}$$

where $\widetilde{\mathbf{D}}^{(t+1)} = \mathbf{G}_2^{-1}\mathbf{D}^{(t+1)}\mathbf{G}_1$. Since $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are diagonal matrices, the problem (30) can be computed as

$$\mathbf{H} \circ \widetilde{\mathbf{D}}^{(t+1)} = \mathbf{G}_2^{-1}\mathbf{H}_3\mathbf{G}_1 \tag{31}$$

where $\bigcirc$ stands for the Hadamard product, and $\mathbf{H}$ is equivalent to

$$\mathbf{H} = \begin{bmatrix} a_1 + b_1 & a_2 + b_1 & \cdots & a_L + b_1 \\ a_1 + b_2 & a_2 + b_2 & \cdots & a_L + b_2 \\ \vdots & & \ddots & \vdots \\ a_1 + b_S & a_2 + b_S & \cdots & a_L + b_S \end{bmatrix} \tag{32}$$

After acquiring $\widetilde{\mathbf{D}}^{(t+1)}$, $\mathbf{D}^{(t+1)}$ can be computed as

$$\mathbf{D}^{(t+1)} = \mathbf{G}_2\widetilde{\mathbf{D}}^{(t+1)}\mathbf{G}_1^{-1}. \tag{33}$$

The step with the highest computational cost is (33), which has the computational complexity $O(SL^2 + LS^2)$, which is lower than $O(S^3L^3)$ of using (29).

2) Updating $\mathbf{V}_3^{(t+1)}$: the optimization for $\mathbf{V}_3^{(t+1)}$ is written as

$$\min_{\mathbf{V}_3^{(t+1)}} \left\|\mathbf{V}_3^{(t+1)} - \mathbf{D}^{(t+1)} + \frac{\mathbf{G}_3^{(t)}}{2\mu}\right\|_F^2, \text{ s.t. } 0 \le \mathbf{V}_3^{(t+1)} \le 1. \tag{34}$$

The solution of (34) is $\mathbf{V}_3^{(t+1)} = \min(\max(\mathbf{D}^{(t+1)} - \frac{\mathbf{G}_3^{(t)}}{2\mu}, 0), 1)$.

3) Updating the Lagrangian multiplier $\mathbf{G}_3^{(t+1)}$:

$$\mathbf{G}_3^{(t+1)} = \mathbf{G}_3^{(t)} + 2\mu\left(\mathbf{V}_3^{(t+1)} - \mathbf{D}^{(t+1)}\right) \tag{35}$$

The overall algorithm for the estimation of spectral basis $\mathbf{D}$ is summarized in Algorithm 3.

Algorithm 3 Estimate **D** with **A** fixed.

Input: **Y**, **Z**, **R**, **B**, **S**, **D**, **A**
Output: **D**
Initialization:
$\mathbf{V}_3 = \mathbf{D}$, $\mathbf{G}_3 = 0$
For $t = 0, 1, \cdots, T_3$
(1) Compute $\mathbf{D}^{(t+1)}$ via the equation (29)
(2) Compute $\mathbf{V}_3^{(t+1)}$ via solving problem (34)
(3) Update the Lagrangian multiplier $\mathbf{G}_3^{(t+1)}$ via the equation (35)
End

3.3. Computation complexity

The proposed SSSR method is summarized in Algorithm 1. The SSSR method is iterated for the estimation of spectral basis and coefficients. In the estimation of coefficients (Algorithm 2), the steps with the highest computational cost in each iteration are computing $\mathbf{A}^{t+1}$ and $\mathbf{V}_2^{(t+1)}$ via (21). Since the term$[(\mathbf{RD})^T\mathbf{RD} + (\eta_2 + \mu)\mathbf{D}^T\mathbf{D} + \mu\mathbf{I}]^{-1}$ can be computed and stored in advance instead of being computed in each iteration, the computation complexity of $\mathbf{A}^{t+1}$ and $\mathbf{V}_2^{(t+1)}$ is $O(L^2N + SLN)$ and $O(SLN + SN^2)$, respectively. Hence, the computation complexity of each iteration in Algorithm 2 is $O(L^2N + SLN + SN^2)$. In the estimation of spectral basis (Algorithm 3), the step with the highest computational cost in each iteration is computing (33), which has the computational complexity $O(SL^2 + LS^2)$. Hence, the computation complexity of each iteration in Algorithm 3 is $O(SL^2 + LS^2)$.

The computation complexity of the SSSR method is $O(T_1T_2(L^2N + SLN + SN^2) + T_1T_3(SL^2 + S^2L))$, where T_1, T_2, and T_3 are the number of iterations in Algorithms 1–3, respectively.

4. Experiments

In the section, we use both remotely sensed and ground-based HSI data sets, to evaluate the effectiveness of the SSSR method.

4.1. Experiment datasets

We use one ground-based HSI datasets, Columbia Computer Vision Laboratory (CAVE) database [53], for non-blind fusion, which has been widely used in the HSI-MSI fusion literatures [33,37,42,54]. The CAVE database has 32 HR-HSIs, which are produced by generalized assorted pixel camera with high quality. Each HSI is of size $512 \times 512 \times 31$, where 512×512 is the spatial resolution, and 31 is the dimensionality of the spectral mode. Each HSI has the wavelength range of 400 nm–700 nm. The HSIs from the database are used as ground truth images. The LR-HSI of size $64 \times 64 \times 31$ is stimulated by firstly blurring the HR-HSI with a 8×8 Gaussian blur (standard deviation 2), and then by downsampling every 8 pixels in both two spatial modes for each band of the ground truth image. The HR-MSI **Z** of the same scenario is simulated by downsampling **X** with the spectral model using spectral downsampling matrix $\mathbf{P}_3$ derived from the response of a Nikon D700 camera.

The two remote sensing HSIs Pavia University [55] and Indian Pines [56] are used for blind fusion. The Pavia University is of size $610 \times 340 \times 115$, which has the 115 spectral bands and 610×340 spectral pixels. We crop the top left 256×256 pixels for experiments. We remove water vapor absorption bands, and thus reduce the HSI as 93 bands. To generate the LR-HSI of size $64 \times 64 \times 93$, we firstly blur the HR-HSI by using a 5×5 Gaussian blur (standard deviation 2) and then downsample every 4 pixels in both two spatial modes of the HR-HSI. To generate the HR-MSI with size of $256 \times 256 \times 4$, the IKONOS-like reflectance spectral response filter is used to downsample the HR-HSI with the spectral mode. The second remote sensing HSI is Indian Pines, which is taken by the NASA's Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) [56]. The image is of size $128 \times 128 \times 224$ covering the wavelength range 400nm-2500nm. We have reduced the number of bands to 200 by removing the bands 104–108, 150–163, and 220 of the image because of extremely low SNR and water absorptions in those bands. The LR-HSI of size $32 \times 32 \times 200$ is produced by applying a 5×5 Gaussian blur (standard deviation 2), and then by downsampling every 4 pixels in both two spatial modes for each band of the ground truth image. The HR-MSI with six bands is simulated by Landsat7-like spectral response.

4.2. Compared methods

The SSSR method is compared with three state-of-the-art HSI-MSI fusion approaches, including the coupled spectral unmixing (CSU) [37], coupled non-negative matrix factorization (CNNMF) [39], non-negative structured sparse representation (NSSR) [42], and two two branches CNN (Two-CNN-Fu) [46]. Since the Two-CNN-Fu is blind fusion method, it is only evaluated on blind fusion case for fair comparison. The other compared methods are evaluated for both blind fusion and non-blind fusion cases.

4.3. Quantitative metrics

In our study, we use five indexes to evaluate the quality of reconstructed HSIs. The first index is the peak signal to noise ratio (PSNR), which is defined as the average PSNR of all bands for HSI, e.g.,

$$\mathrm{PSNR}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{1}{S}\sum_{j=1}^{S}\mathrm{PSNR}(\widehat{\mathbf{X}^i}, \mathbf{X}^i), \tag{36}$$

where $\mathbf{X}^i$ and $\widehat{\mathbf{X}}^i$ denote i^{th} band images of ground truth $\mathbf{X} \in \mathbb{R}^{S\times N}$ and estimated HSIs $\widehat{\mathbf{X}} \in \mathbb{R}^{S\times N}$, respectively, and both of them are scaled to a range [0, 255]. The PSNR measures the similarities between the ground truth image and reconstructed image. The higher the PSNR, the better the fusion result.

The second index is the spectral angle mapper (SAM), which is defined as the average angle between the ground truth pixel $\mathbf{x}_j$ and reconstructed pixel $\widehat{\mathbf{x}}_j$, i.e.

$$\mathrm{SAM}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{1}{N}\sum_{j=1}^{N}\arccos\frac{\widehat{\mathbf{x}}_j^T\mathbf{x}_j}{\|\widehat{\mathbf{x}}_j\|_2\|\mathbf{x}_j\|_2}. \tag{37}$$

The SAM is given in degrees. It measures the spectral quality of the reconstructed HR-HSI. The smaller SAM, the better fusion quality.

The third index is the degree of distortion (DD), defined as

$$\mathrm{DD}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{1}{SN}\|\mathrm{vec}(\widehat{\mathbf{X}}) - \mathrm{vec}(\mathbf{X})\|_1, \tag{38}$$

where $\mathrm{vec}(\widehat{\mathbf{X}})$ and $\mathrm{vec}(\mathbf{X})$ are vectorizations of matrixes $\widehat{\mathbf{X}}$ and $\mathbf{X}$, respectively. The smaller the DD, the better the spectral quality.

The fourth index is the relative dimensionless global error in synthesis (ERGAS) [57], which is defined as

$$\mathrm{ERGAS}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{100}{c}\sqrt{\frac{1}{S}\sum_{i=1}^{S}\frac{\mathrm{MSE}(\widehat{\mathbf{X}}^i, \mathbf{X}^i)}{\mu_{\widehat{\mathbf{X}}^i}^2}}, \tag{39}$$

where c is spatial downsampling factor, $\mu_{\widehat{\mathbf{X}}^i}$ is the mean value of $\widehat{\mathbf{X}}^i$, and $\mathrm{MSE}(\widehat{\mathbf{X}}^i, \mathbf{X}^i)$ represents the mean square error of $\mathbf{X}^i$ and $\widehat{\mathbf{X}}^i$. The smaller ERGAS, the better fusion results.

The fifth index is the universal image quality index (UIQI) [58]. In this paper, we compute the UIQI for each window of size 32×32, and then average the UIQI of all windows. Let $\mathbf{X}_j^i$ and $\widehat{\mathbf{X}}_j^i$ denote the j^{th} window of i^{th} band ground truth image and estimated image, respectively. The UIQI between i^{th} band images $\mathbf{X}^i$ and $\widehat{\mathbf{X}}^i$ is given by

$$\mathrm{Q}(\widehat{\mathbf{X}}^i, \mathbf{X}^i) = \frac{1}{H}\sum_{j=1}^{M}\frac{\sigma^2_{\mathbf{X}_j^i\widehat{\mathbf{X}}_j^i}}{\sigma_{\mathbf{X}_j^i}\sigma_{\widehat{\mathbf{X}}_j^i}}\frac{2\mu_{\mathbf{X}_j^i}\mu_{\widehat{\mathbf{X}}_j^i}}{\mu^2_{\mathbf{X}_j^i} + \mu^2_{\widehat{\mathbf{X}}_j^i}}\frac{2\sigma_{\mathbf{X}_j^i}\sigma_{\widehat{\mathbf{X}}_j^i}}{\sigma^2_{\mathbf{X}_j^i} + \sigma^2_{\widehat{\mathbf{X}}_j^i}}, \tag{40}$$

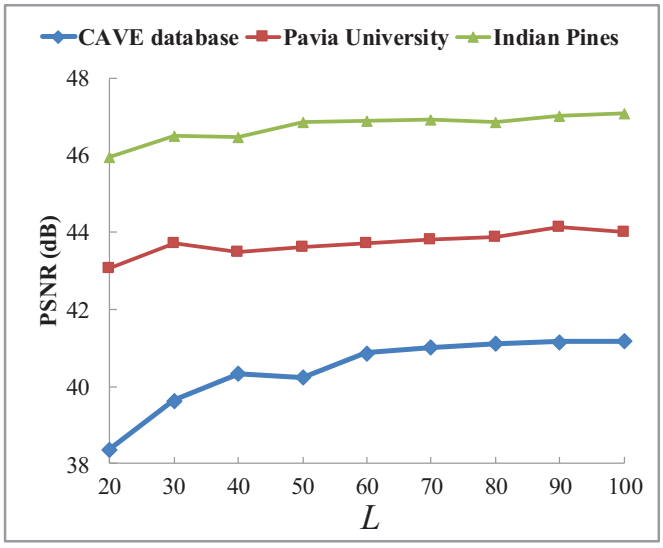

Fig. 2. PSNR curves as functions of the number of atoms L for the *CD* (an HSI in the CAVE database), Pavia University, and Indian Pines.

where H is the number of windows, $\sigma_{\mathbf{X}_j^i\hat{\mathbf{X}}_j^i}$ is the sample covariance between $\mathbf{X}_j^i$ and $\mathbf{X}_j^i$, and $\mu_{\mathbf{X}_j^i}$ and $\sigma_{\mathbf{X}_j^i}$ are the mean value and standard deviation of $\mathbf{X}_j^i$, respectively. The UIQI index for HSI is defined as the average value of all bands, i.e.

$$\mathrm{UIQI}(\hat{\mathbf{X}}, \mathbf{X}) = \frac{1}{S}\sum_{i=1}^{S} \mathrm{Q}(\hat{\mathbf{X}}^i, \mathbf{X}^i). \tag{41}$$

The larger the UIQI, the better the fusion results.

4.4. Parameters discussions

To evaluate the sensitivity of the SSSR to its key parameters, we run the SSSR for different values of the number of atoms L, the sparsity regularization parameter η_1, and the non-local spatial similarities regularization parameter η_2.

The number of atoms of spectral basis L has an important effect on the fusion process. Fig. 2 plots the PSNR of the image *CD* (a representative HSI in CAVE database), Pavia University, and Indian Pines as functions of the number of atoms L. As Fig. 2 shows, the PSNR has a increase for three datasets when L and varies from 20 to 80. Then, the PSNR reaches a stable level. Therefore, L is set to 80 for three datasets.

The parameter η_1 is highly related to the effect of the sparse prior. The higher η_1, the sparser **A**. Fig. 3(a) plots the PSNR of the image *CD* (a representative HSI in CAVE database), Pavia University, and Indian Pines as functions of $\log\eta_1$ (log is base 10). As can be seen from Fig. 3(a), the PSNR for the three datasets keeps relatively stable as $\log\eta_1$ varies from -7 to -4, and then it decreases as $\log\eta_1$ increases. Therefore, we set $\eta_1 = 1\times10^{-4}$ for three datasets.

The parameter η_2 is highly related to the effect of the non-local spatial similarities. Fig. 3(b) plots the PSNR of the image *CD* ((a representative HSI in CAVE database)), Pavia University, and Indian Pines as functions of $\log\eta_2$ (log is base 10). As we can see from the Fig. 3(b), the PSNR for Pavia University and Indian Pines keeps stable as $\log\eta_2$ varies from -7 to -3, and then it increases rapidly as $\log\eta_2$ further increases. However, the PSNR for CAVE database has the obvious raise when $\log\eta_2$ varies from -7 to -2, and then it decreases rapidly as $\log\eta_2$ is larger than -1. The non-local spatial similarities prior is more effective for the CAVE database, compared with Pavia University and Indian Pines. The reason may be that the spatial resolution of the Pavia University and Indian Pines is much lower than that of the CAVE database, and the spectral pixels in the CAVE database are much similar with each other, which needs larger regularization parameter η_2. We set $\eta_2 = 1\times10^{-4}$ for Pavia University and Indian Pines, and $\eta_2 = 0.015$ for CAVE database.

4.5. Experimental results of non-blind fusion

Non-blind fusion is the case that the spectral response and PSF are assumed to be perfectly known. In this section, we show the non-blind fusion results on the CAVE database.

Table 1 shows the PSNR, SAM, DD, ERGAS, and UIQI of the recovered HSIs for the CAVE database. We mark the best results in bold for clarity. As we can see from Table 2, the SSSR method performs best among the compared methods, and NSSR takes the second place. The CSU and CNNMF exploit the prior of spectral unmixing but do not exploit the non-local spatial similarities and a sparse prior of the HR-HSI. However, NSSR exploits the prior of non-local spatial similarities and

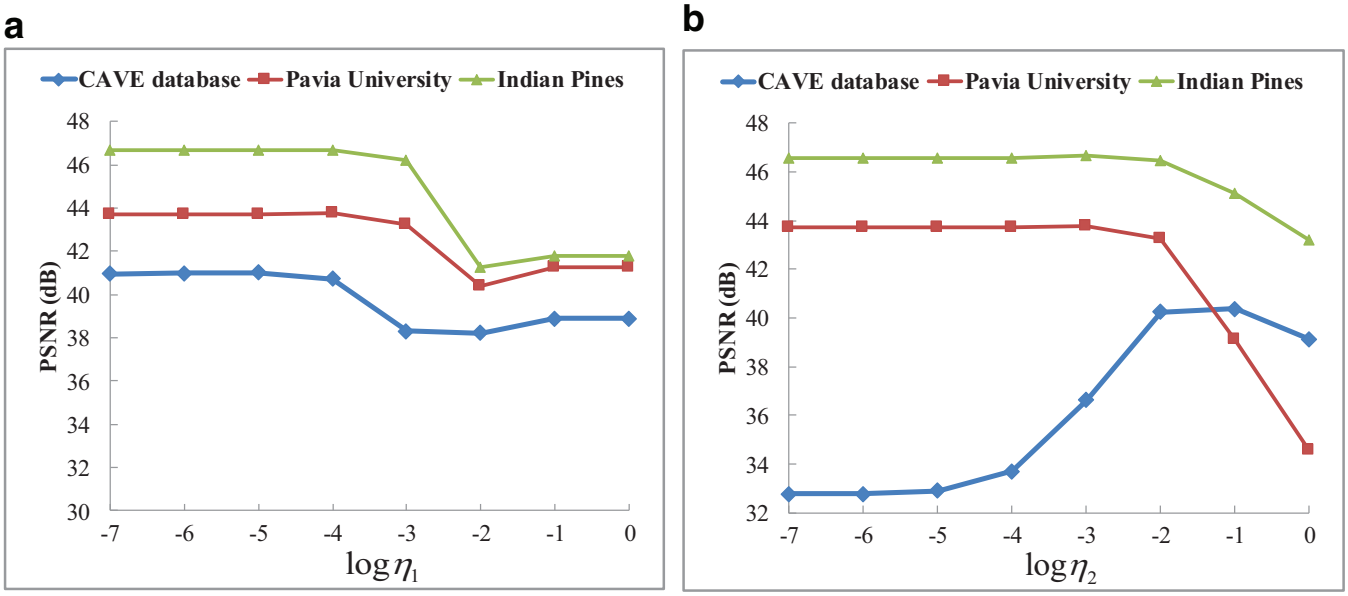

Fig. 3. PSNR curves as functions of $\log\eta_1$ and $\log\eta_2$. (a) $\log\eta_1$. (b) $\log\eta_2$.

Table 1
Average quantitative metrics of non-blind fusion on the CAVE database [53].

Method	CAVE database [53]				
	PSNR	SAM	DD	ERGAS	UIQI
Best Values	+∞	0	0	0	1
CNNMF [36]	43.56	5.49	1.04	1.207	0.8815
CSU [37]	41.68	6.68	1.18	1.504	0.8639
NSSR [42]	45.02	4.69	0.78	0.988	0.8914
SSSR	**45.93**	**4.38**	**0.69**	**0.899**	**0.9022**

Table 2
Quantitative metrics of blind fusion on the Pavia University [55].

Method	Pavia University [55]				
	PSNR	SAM	DD	ERGAS	UIQI
Best Values	+∞	0	0	0	1
CNNMF [36]	42.58	1.96	1.22	1.174	0.9935
CSU [37]	39.65	2.44	1.62	1.694	0.9889
NSSR [42]	42.24	1.89	1.30	1.172	0.9930
Two-CNN-Fu [46]	41.27	2.23	1.37	1.338	0.9923
SSSR	**43.99**	**1.68**	**1.07**	**0.986**	**0.9952**

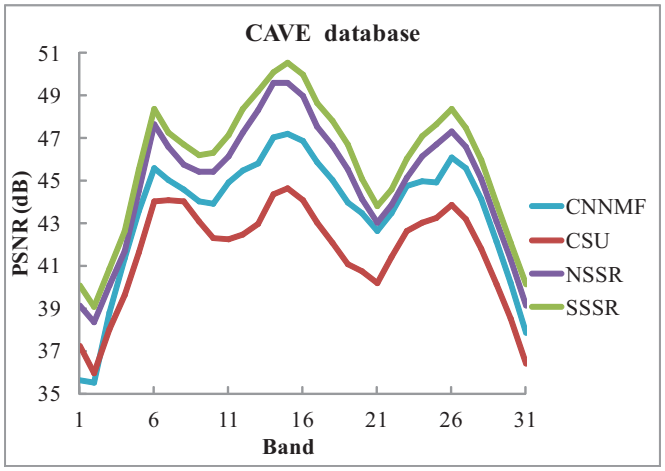

Fig. 4. Average PSNR curves as functions of the spectral bands on CAVE database.

sparse prior of the HR-HSI, but do not fully utilize the priors spectral unmixing. For purpose of the comparisons of the test methods in different spectral bands, Fig. 4 plots the average PSNR curves as functions of the wavelengths over CAVE database for the compared approaches. The SSSR almost performs best in all spectral bands among the test methods, as can be seen from Fig. 4. Besides, all the test methods perform well in the middle spectral bands, and relatively bad in the first and last a few bands. The reason is that a blur exists in the first and last a few bands [59]. To compare the performance in preserving spatial structures of the test methods, the reconstructed HR-HSIs and corresponding error images on *flowers* (an HSI in the CAVE database) of CNNMF, NSSR, and SSSR at 9th and 18th bands are shown in Fig. 5. Since the CNNMF, NSSR, and SSSR performs obviously better than the CSU, we select these three methods for visual comparisons. The error images reflect the differences between the fusion results and ground truths. From Fig. 5, we can see that these three methods can recover most of the spatial details of the

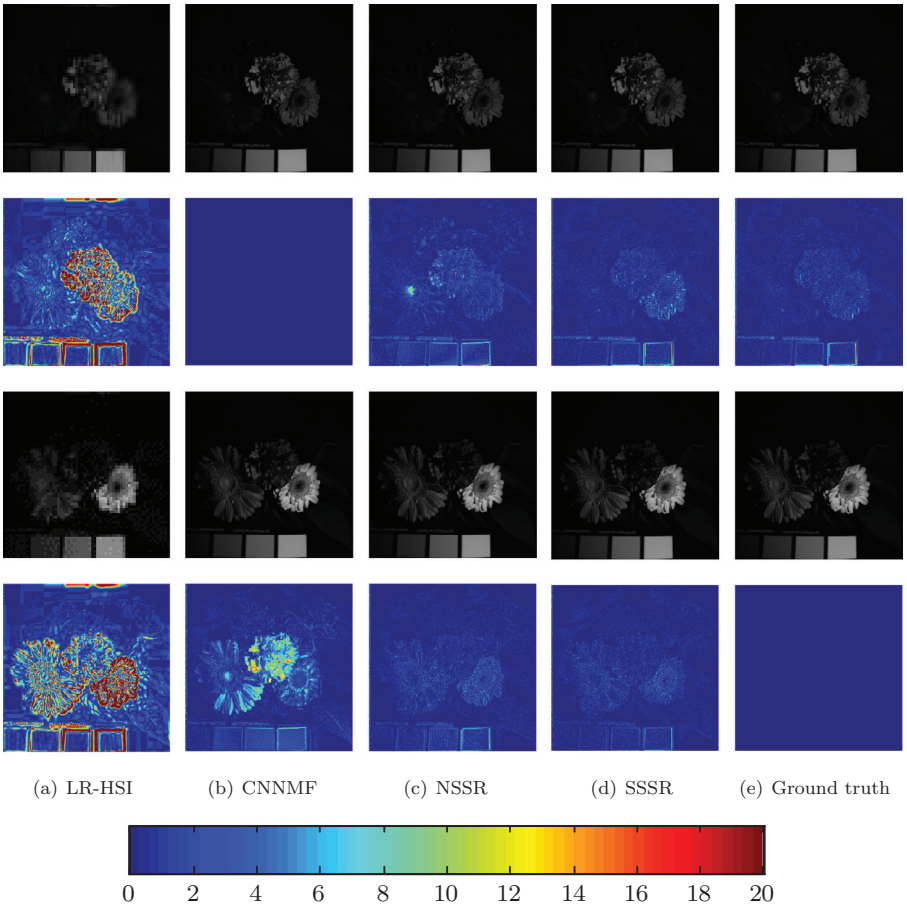

Fig. 5. The first and second rows show the fused and error images of *flowers* at 9th band, respectively. The third and fourth rows show the fused and error images of *flowers* at 18th band, respectively. (a) LR-HSI; (b) CNNMF [36]. (c) NSSR [42]. (d) SSSR. (e) Ground truth.

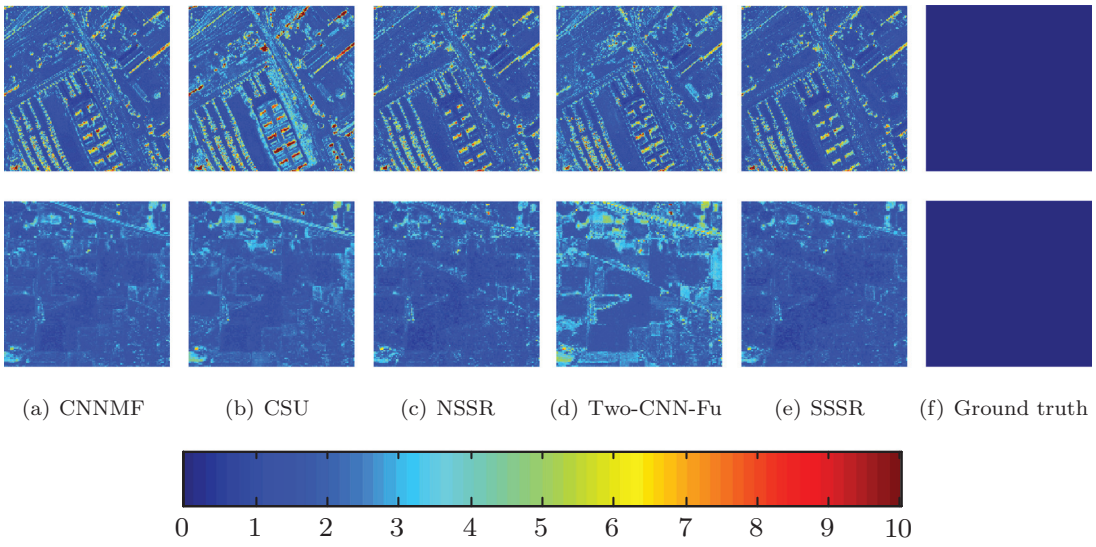

Fig. 6. The first row shows the SAM images of the fused Pavia University. The second row shows the SAM images of the fused Indian Pines. (a) CNNMF [36]. (b) CSU [37]. (c) NSSR [42]. (d) Two-CNN-Fu [46]. (e) SSSR. (f) Ground truth.

Table 3
Quantitative metrics of blind fusion on the Indian Pines [56].

Method	Indian pines [56]				
	PSNR	SAM	DD	ERGAS	UIQI
Best Values	+∞	0	0	0	1
CNNMF [36]	44.68	1.74	1.28	0.804	0.8714
CSU [37]	44.93	1.87	1.38	0.833	0.8818
NSSR [42]	45.55	1.64	1.27	0.735	0.8951
Two-CNN-Fu [46]	41.56	2.27	2.09	1.060	0.8200
SSSR	**46.88**	**1.55**	**1.06**	**0.677**	**0.9085**

HR-HSI, though there are a few distortions in the fusion results. Furthermore, the SSSR method performs best keeping the spatial structures.

4.6. Experimental results of blind fusion

To better simulate the real data fusion, references [41,60] conduct experiments of blind fusion, where the spectral response and PSF are assumed to be unknown, which is similar to real data fusion case. We use the two remote sensing HSIs, i.e., Pavia University and Indian Pines, for the blind fusion experiments. The deep learning method Two-CNN-Fu needs pre-training. The top left 256×256 region of Pavia University is used for testing, and the rest part is used for training. Since the Indian Pines is captured by AVIRIS, the other remote sensing HSIs captured by AVIRIS are used for training when we test on Indian Pines. In the experiments, both of the spectral response and PSF are estimated from the LR-HSI and HR-MSI by the method proposed in [41]. Tables 2 and 3 show the quantitative metrics of blind fusion on the Pavia University and Indian Pines, respectively. As can be seen from the Tables 2 and 3, the proposed SSSR method still performs the best in terms of all quantitative metrics, which means that our method is also effective in the blind fusion cases. The performance of deep learning method Two-CNN-FU is not competitive for the two HSIs. The main reason is that the Two-CNN-FU does not incorporate the observation models (3) and (4), which explore the relationship among the LR-HSI, HR-MSI, and HR-HSI. Fig. 6 shows the SAM images of the test methods on the Pavia University and Indian Pines. The SAM images reflect the SAM between the reconstructed HSIs and ground truth. As can be seen from the SAM images, the SSSR has less spectral errors than other methods on Pavia University. For Indian Pines, the SSSR and CNNMF perform comparably in terms of the SAM images.

5. Conclusions

We present a spatial-spectral sparse representation based method, termed as SSSR, to estimate an HR-HSI, by fusing an LR-HSI with an HR-MSI counterpart. In the proposed SSSR method, we formulate the fusion problem as the estimation of the spectral basis and coefficients by exploiting three important prior information of the HR-HSI, i.e., sparse prior, non-local spatial similarities, priors of spectral unmixing. We design the alternative optimization algorithm to update the spectral basis and coefficients with these priors. Experimental results on both non-blind fusion and and blind fusion cases demonstrate the effectiveness of the SSSR method.

Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their insightful comments and suggestions, which have significantly improved this paper. They would also like to thank Dr. J. Yang for providing the codes. This work was supported by the National Natural Science Fund of China under Grant 61890962 and 61520106001, the Science and Technology Plan Project Fund of Hunan Province under Grant CX2018B171, 2017RS3024, and 2018TP1013, and the Science and Technology Talents Program of Hunan Association for Science and Technology under Grant 2017TJ-Q09.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.inffus.2018.11.012.

References

[1] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, IEEE Geosci. Remote Sens. Mag. 1 (2) (2013) 6–36.
[2] H. Ghassemian, A review of remote sensing image fusion methods, Inf. Fus. 32 (2016) 75–89.
[3] H. Akbari, Y. Kosugi, K. Kojima, N. Tanaka, Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging, IEEE Trans. Biomed. Eng. 57 (8) (2010) 2011–2017.
[4] D. Rav, H. Fabelo, G.M. Callic, G. Yang, Manifold embedding and semantic segmentation for intraoperative guidance with hyperspectral brain imaging, IEEE Trans. Med. Imag. 36 (9) (2017) 1845–1857.
[5] G.A. Shaw, H.K. Burke, Spectral imaging for remote sensing, Lincoln Lab. J. 14 (1) (2003) 3–28.

[6] H. Kwon, Y.-W. Tai, RGB-guided hyperspectral image upsampling, in: Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 307–315.
[7] S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: a survey of the state of the art, Inf. Fus. 33 (2017) 100–112.
[8] S. Li, R. Dian, L. Fang, J.M. Bioucas-Dias, Fusing hyperspectral and multispectral images via coupled sparse tensor factorization, IEEE Trans. Image Process. 27 (8) (2018) 4118–4130, doi:10.1109/TIP.2018.2836307.
[9] R. Dian, S. Li, L. Fang, Learning a low tensor-train rank representation for hyperspectral image super-resolution, IEEE Trans. Neural Netw. Learn. Syst. (2019, accepted for publication).
[10] A.A. Goshtasby, S. Nikolov, Image fusion: advances in the state of the art, Inf. Fus. 8 (2) (2007) 114–118.
[11] J.J. Lewis, R.J.O. Callaghan, S.G. Nikolov, D.R. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, Inf. Fus. 8 (2) (2007) 119–130.
[12] S. Yang, M. Wang, L. Jiao, R. Wu, Z. Wang, Image fusion based on a new Contourlet packet, Inf. Fus. 11 (2) (2010) 78–84.
[13] Z. Liu, E. Blasch, V. John, Statistical comparison of image fusion algorithms: recommendations, Inf. Fus. 36 (2017) 251–260.
[14] Z. Liu, E. Blasch, G. Bhatnagar, V. John, W. Wu, R.S. Blum, Fusing synergistic information from multi-sensor images: an overview from implementation to performance assessment, Inf. Fus. 42 (2018) 127–145.
[15] Y. Liu, X. Chen, Z. Wang, Z.J. Wang, R.K. Ward, X. Wang, Deep learning for pixel-level image fusion: recent advances and future prospects, Inf. Fus. 42 (2018) 158–173.
[16] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inf. Fus. 24 (2015) 147–164.
[17] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fus. 36 (2017) 191–207.
[18] Y. Liu, S. Liu, Z. Wang, Multi-focus image fusion with dense SIFT, Inf. Fusion 23 (2015) 139–155.
[19] A.P. James, B.V. Dasarathy, Medical image fusion: a survey of the state of the art, Inf. Fus. 19 (1) (2014) 4–19.
[20] L. Wang, B. Li, L. Tian, Multi-modal medical image fusion using the inter-scale and intra-scale dependencies between image shift-invariant Shearlet coefficients, Inf. Fus. 19 (1) (2014) 20–28.
[21] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, W. Wu, Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study, IEEE Trans. Pattern Anal. Mach. Intell. 34 (1) (2012) 94–109.
[22] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, L.M. Bruce, Comparison of pan-sharpening algorithms: outcome of the 2006 GRSS data-fusion contest, IEEE Trans. Geosci. Remote Sens. 45 (10) (2007) 3012–3021.
[23] S. Li, J.T. Kwok, Y. Wang, Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images, Inf. Fus. 3 (1) (2002) 17–23.
[24] F. Nencini, A. Garzelli, S. Baronti, L. Alparone, Remote sensing image fusion using the Curveletcurvelet transform, Inf. Fus. 8 (2) (2007) 143–156.
[25] S. Yang, M. Wang, L. Jiao, Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis, Inf. Fus. 13 (3) (2012) 177–184.
[26] S. Li, B. Yang, A new pan-sharpening method using a compressed sensing technique, IEEE Trans. Geosci. Remote Sens. 49 (2) (2011) 738–746.
[27] W. Wang, L. Jiao, S. Yang, Fusion of multispectral and panchromatic images via sparse representation and local autoregressive model, Inf. Fus. 20 (2014) 73–87.
[28] Q. Zhang, Y. Liu, R.S. Blum, J. Han, D. Tao, Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review, Inf. Fus. 40 (2018) 57–75.
[29] S. Yang, K. Zhang, M. Wang, Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets, IEEE Trans. Neural Netw. Learn. Syst. (2018), doi:10.1109/TNNLS.2017.2736011.
[30] R.C. Hardie, M.T. Eismann, G.L. Wilson, Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor, IEEE Trans. Image Process. 13 (9) (2004) 1174–1184.
[31] Q. Wei, N. Dobigeon, J.-Y. Tourneret, Fast fusion of multi-band images based on solving a Sylvester equation, IEEE Trans. Image Process. 24 (11) (2015) 4109–4121.
[32] N. Akhtar, F. Shafait, A. Mian, Bayesian sparse representation for hyperspectral image super resolution, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 3631–3640.
[33] R. Kawakami, J. Wright, Y.W. Tai, Y. Matsushita, M. Ben-Ezra, K. Ikeuchi, High-resolution hyperspectral imaging via matrix factorization, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2011, pp. 2329–2336.
[34] B. Huang, H. Song, H. Cui, J. Peng, Z. Xu, Spatial and spectral image fusion using sparse matrix factorization, IEEE Trans. Geosci. Remote Sens. 52 (3) (2014) 1693–1704.
[35] H. Song, B. Huang, K. Zhang, H. Zhang, Spatio-spectral fusion of satellite images based on dictionary-pair learning, Inf. Fus. 18 (2014) 148–160.
[36] N. Yokoya, T. Yairi, A. Iwasaki, Coupled non-negative matrix factorization unmixing for hyperspectral and multispectral data fusion, IEEE Trans. Geosci. Remote Sens. 50 (2) (2012) 528–537.
[37] C. Lanaras, E. Baltsavias, K. Schindler, Hyperspectral super-resolution by coupled spectral unmixing, in: Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 3586–3594.
[38] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, S. Godsill, Multi--band image fusion based on spectral unmixing, IEEE Trans. Geosci. Remote Sens. 54 (12) (2016) 7236–7249.
[39] N. Akhtar, F. Shafait, A. Mian, Sparse spatio-spectral representation for hyperspectral image super-resolution, in: Proc. Euro. Conf. Comput. Vis., 2014, pp. 63–78.
[40] M.A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J.M. Bioucas-Dias, J. Chanussot, Hyperspectral super-resolution of locally low rank images from complementary multisource data, IEEE Trans. Geosci. Remote Sens. 25 (1) (2016) 274–288.
[41] M. Simoes, J. Bioucas-Dias, L. Almeida, J. Chanussot, A convex formulation for hyperspectral image superresolution via subspace-based regularization, IEEE Trans. Geosci. Remote Sens. 53 (6) (2015) 3373–3388.
[42] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, X. Li, Hyperspectral image super-resolution via non-negative structured sparse representation, IEEE Trans. Image Process. 25 (5) (2016) 2337–2352.
[43] M.D. Iordache, J. Bioucas-Dias, A. Plaza, Sparse unmixing of hyperspectral data, IEEE Trans. Geosci. Remote Sens. 49 (6) (2011) 2014–2039.
[44] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[45] R. Dian, S. Li, A. Guo, L. Fang, Deep hyperspectral image sharpening, IEEE Trans. Neural Netw. Learn. Syst. 29 (11) (2018) 5345–5355, doi:10.1109/TNNLS.2018.2798162.
[46] J. Yang, Y. Zhao, J. Chan, Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network, Remote Sens. 10 (5) (2018) 800.
[47] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Non-local sparse models for image restoration, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 2272–2279.
[48] T. Lu, S. Li, L. Fang, Y. Ma, J.A. Benediktsson, Spectral-spatial adaptive sparse representation for hyperspectral image denoising, IEEE Trans. Geosci. Remote Sens. 54 (1) (2016) 373–385.
[49] 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. 3 (1) (2011) 1–122.
[50] S. Li, R. Dian, L. Fang, J.M. Bioucas-Dias, Fusing hyperspectral and multispectral images via coupled sparse tensor factorization, IEEE Trans. Image Process. 27 (8) (2018) 4118–4130, doi:10.1109/TIP.2018.2836307.
[51] D.L. Donoho, M. Elad, Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization, Proc. Natl. Acad. Sci. 100 (5) (2003) 2197–2202.
[52] W. Dong, X. Li, L. Zhang, G. Shi, Sparsity-based image denoising via dictionary learning and structural clustering, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2011, pp. 457–464.
[53] F. Yasuma, T. Mitsunaga, D. Iso, S.K. Nayar, Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum, IEEE Trans. Image Process. 19 (9) (2010) 2241–2253.
[54] R. Dian, L. Fang, S. Li, Hyperspectral image super-resolution via non-local sparse tensor factorization, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 5344–5353.
[55] F. Dell'Acqua, P. Gamba, A. Ferrari, J.A. Palmason, J.A. Benediktsson, Exploiting spectral and spatial information in hyperspectral urban data with high resolution, IEEE Geosci. Remote Sens. Lett. 1 (4) (2004) 322–326.
[56] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis, Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS), Remote Sens. Environ. 65 (3) (1998) 227–248.
[57] L. Wald, Quality of high resolution synthesised images: is there a simple criterion? in: Proc. Int. Conf. Fusion Earth Data, 2000, pp. 99–103.
[58] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.
[59] N. Akhtar, F. Shafait, A. Mian, Hierarchical beta process with Gaussian process prior for hyperspectral image super resolution, in: Proc. Euro. Conf. Comput. Vis., 2016, pp. 103–120.
[60] N. Yokoya, C. Grohnfeldt, J. Chanussot, Hyperspectral and multispectral data fusion: a comparative review of the recent literature, IEEE Geosci. Remote Sens. Mag. 5 (2) (2017) 29–56.