

# Hyperspectral and Multispectral Image Fusion via Nonlocal Low Rank Tensor Approximation and Sparse Representation

Xuelong Li, *Fellow, IEEE*, Yue Yuan, and Qi Wang, *Senior Member, IEEE*

**Abstract**—The fusion of hyperspectral (HS) and multispectral (MS) images designed to obtain high-resolution HS (HRHS) images is a very challenging work. A series of solutions have been proposed in recent years. However, the similarity structure of HS image has not been fully utilized. In this paper, we present a novel HS and MS image fusion method based on nonlocal low-rank tensor approximation and sparse representation. Specifically, the HS image and MS image are considered to be the spatially and spectrally degraded versions of the HRHS image respectively. Then, the nonlocal low-rank constraint term is adopted in order to form the nonlocal similarity and the spatial-spectral correlation. Meanwhile, we add the sparse constraint term to describe the sparsity of abundance. Thus, the proposed fusion model is established and its optimization is solved by alternative direction method of multipliers (ADMM). The experimental results on three synthetic data sets and one real data set show the advantages of the proposed method over several state-of-the-art competitors.

**Index Terms**—Hyperspectral (HS) Image, multispectral (MS) image, image fusion, low-rank tensor approximation, sparse representation.

## I. INTRODUCTION

**H**YPERSPECTRAL (HS) imaging sensor is able to acquire contiguous images with numerous narrow bandwidth in a wide range of wavelength. For this reason, HS images generally have high spectral resolution, which is significant to a wide range of fields including band selection [1], image classification [2], and change detection [3]. Owing to hardware restrictions and actual signal-to-noise ratio, spatial resolution and spectral resolution cannot achieve the desired effect at the same time and there must be a reasonable trade-off between them. Consequently, HS images have lower spatial resolution compared to multispectral (MS) and panchromatic (PAN) images with much less spectral bands. In order to get high resolution HS (HRHS) images, a popular approach is to fuse low resolution HS (LRHS) image with high resolution MS (HRMS) image, called *HS-MS fusion* [4].

Image fusion is a way to effectively combine the advantages of different images. According to the different images involved in the fusion process, it is divided into HS-PAN fusion and HS-MS fusion.

This work was supported by the National Natural Science Foundation of China under Grant U1864204, U1801262, 61871470, and 61773316.

X. Li, Y. Yuan and Q. Wang are with the School of Computer Science and with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: li@nwpu.edu.cn; yuanyue1996@mail.nwpu.edu.cn; crabwq@gmail.com).

Qi Wang is the corresponding author.

1) HS-PAN fusion [5]. The fusion of LRHS image and high resolution PAN image, also called *HS pansharpening*, has been developed to increase the spatial resolution of HS images. Similar to MS pansharpening [6], [7], it is the fusion of multi-band images with single-band images. The MS pansharpening methods can be directly used for this problem, and have been successfully implemented in [8]. Also, there are some matrix factorization-based methods designed for HS pansharpening [9]. This kind of methods exploits the linear spectral mixture model assuming that the image can be unmixed into a product of endmembers and abundances.

2) HS-MS fusion [4], [10]. Many attempts have been taken to use pansharpening to solve the HS-MS fusion problem. Although it may get a high quantity fused image, it is not entirely suitable for this problem. In recent years, approaches specially designed for this problem have been investigated: matrix factorization-based [11]–[17], Bayesian-based [18]–[20], tensor-based [21]–[25], deep learning-based methods [26]–[28]. Based on matrix factorization, some methods solve the fusion problem using spectral unmixing techniques. Furthermore, Bayesian-based methods are introduced and they regularize the fusion problem by the maximum a posteriori framework. More recently, considering the three-dimension (3D) structure of HS images, tensor is adopted to preserve the original spatial-spectral structure. Additionally, several deep learning models are designed for this problem. Benefiting from its learning ability to high-level features, these methods often get high fusion quality but also high time complexity.

In this paper, we concentrate on tensor-based methods, and present a novel nonlocal low rank tensor approximation and sparse representation (NLRSR) based LRHS image and HRMS image fusion method. The main contributions are summarized as follows.

1) A novel HS-MS fusion model named NLRSR is presented to enhance the spatial resolution of HS image. Different from previous methods, both matrix and tensor are introduced to this model. Experiments show that this model can effectively fuse the spectral and spatial information from HS and MS images.

2) A novel low rank constraint criterion is developed to further exploit the spatial correlation, spectral correlation and nonlocal self-similarity of HS image. The former two physical features are easy to form by 3D low rank, but the nonlocal self-similarity is also significant to HS image. We incorporate all these physical characteristics into the proposed model to inherit structural information better. Different from other ap-

proaches, a minmax concave plus (MCP) penalty is introduced in this low rank constraint criterion.

3) The tensor-train (TT) [29] rank is introduced to improve the low rank representation. It has been proved that the TT rank works better than Tucker rank in some methods related to the color image. It is used in the 4D low rank term and its effectiveness has been shown in experiments.

The remainder of this paper is organized as follows. Section II reviews related HS-MS fusion methods. The proposed NLRSSR method is described in detail in Section III. In Section IV, the optimization algorithm for solving the proposed fusion model is given. The experimental results and analysis of eight comparison algorithms on four HS data sets are shown in Section V. Finally, we conclude this paper in Section VI.

## II. RELATED WORK

In recent years, a series of methods have been developed for the problem of HS-MS fusion. As mentioned in the previous section, there are four typical kind of methods to deal with the problem: matrix factorization-based, Bayesian-based, tensor-based and deep learning-based methods. In this section, we will introduce some popular methods of these four categories.

### A. Matrix Factorization-Based Methods

Methods based on matrix factorization or spectral unmixing [30] rely on a basic hypothesis: the LRHS image is regarded as the spatial degradation result of HRHS image, and the HRMS image is considered to be the degradation of the HRHS image in the spectral domain. These methods aim to reconstruct HRHS image using endmember spectra matrix which is estimated from LRHS image and abundance matrix which is estimated from HRMS image. Different methods propose different ways to estimate the two matrices.

Coupled nonnegative matrix factorization (CNMF) [11] is a classical HS and MS fusion method. Based on the linear mixture model, LRHS and HRMS data are alternately unmixed by nonnegative matrix factorization (NMF) [31], which has been widely used for HS unmixing. In [12], Akhtar *et al.* use online dictionary learning to learn a dictionary from LRHS image. And the corresponding sparse code is learned using a generalization of simultaneous orthogonal matching pursuit (GSOMP+) algorithm. Moreover, total variation regularization is used to preserve the local spatial smoothness in [13], which named HYSURE. Considering the spatial correlation of the HS image, Veganzones *et al.* [14] decompose the HS image with the locally low-rank prior, which conducted on each HS image patch. Utilizing the similarity of adjacent band images and the low rank structure of HS data, a method based on group spectral embedding is presented in [16]. In [17], Wei *et al.* design a sparse regularization for the HS fusion model to take advantage of the self-similarity property of HS images. Taking spatio-spectral sparsity of the HS image into account, a clustering based non-negative structured sparse representation (NSSR) [15] is developed to estimate the abundance matrix from LRHS image. Similar to [12], the endmember spectra matrix in [15] is acquired by dictionary learning algorithm based on block coordinate descent algorithm.

### B. Bayesian-Based Methods

Bayesian-based methods usually regularize the HS-MS fusion problem by the maximum a posteriori framework. It utilizes the priori probability distributions of various characteristics in HS image to compute a posteriori probability which is used to construct HRHS image. This kind of methods solve the problem in a subspace.

Akhtar *et al.* first introduce a Bayesian based method in [18]. They use a non-parametric Bayesian dictionary learning to learn high resolution dictionary from LRHS image, and then use the dictionary with the corresponding coefficient obtained from HRMS image through Bayesian sparse coding so as to construct HRHS image. Different from [18], a method called fast fusion based on Sylvester equation (FUSE) [19] is developed. This method solves the Sylvester equation derived from the HS-MS fusion model and can also be applied to HS-PAN fusion. The Bayesian framework is easily generalized when some prior conditions need to be considered. Subsequently, a more robust algorithm R-FUSE [20] is proposed to enhance the robustness and reduce the computational complexity of FUSE.

### C. Tensor-Based Methods

In the past years, tensor is adopted to represent images to uncover the underlying data structure in image processing [32], [33]. Different from matrix-based methods, tensor-based methods form HS data as a 3D tensor rather than a 2D matrix and add analysis of spatial dimensional structural features. For a 3D HS tensor, the dimensions of the tensor represent its width, height and band number, respectively.

In [21], the HRHS image is reconstructed by dictionary matrices of three modes and a core tensor through sparse coding and dictionary learning. It introduces the tensor factorization to HS-MS fusion. Meanwhile, the HRHS image is spatially partitioned into a number of cubes, and the similar cubes are grouped to take advantage of the non-local spatial similarities. Coupled sparse tensor factorization (CSTF) [22] also utilizes the core tensor and three factor matrices to obtain the HRHS image. Different from [21], it uses Tucker decomposition [34] to estimate them. A low rank tensor decomposition model which is regularized by spatial-spectral graph for HS-MS fusion is proposed in [23]. In this model, two graphs are built to characterize neighboring pixels similarity of HRMS image and adjacent bands correlation of LRHS image, respectively. Method in [24] uses nonlocal patch tensor to preserve spatial and spectral similarities in the HRHS image. Further, a novel tensor rank named low TT rank (LTTR) is introduced in [25]. This method uses the LTTR prior, which has been shown to perform better than Tucker rank in the color image, to improve the resolution of HS image. Different from approaches adopting 3D tensor, it uses 4D tensor to exploit the prior. To be specific, the similar HS cubes are grouped to constitute a 4D tensor.

### D. Deep Learning-Based Methods

Different from traditional methods, deep learning-based methods aim at learning a mapping between LRHS and HRHS

TABLE I  
NOTATIONS USED IN THIS PAPER.

Notation	Description
$\mathcal{X}$	tensor
$\mathbf{X}$	matrix
$a$	scalar
$\text{Unfold}_n(\mathcal{X}) = \mathbf{X}_{(n)}$	mode- $n$ unfolding of tensor $\mathcal{X}$
$\text{Fold}_n(\mathbf{X}_{(n)}) = \mathcal{X}$	mode- $n$ folding of matrix $\mathbf{X}_{(n)}$
$\mathbf{X}_{(n)}$	mode- $n$ canonical of tensor $\mathcal{X}$
$\ \cdot\ _F$	Frobenius norm

images.

Most of these methods are transformed from computer vision tasks. In recent years, methods for LRHS and HRMS images are proposed. Han *et al.* [26] propose a pre-upsampling super-resolution method and the structures of which is based on SRCNN. In [27], a fusion method with two branches is proposed, which uses two branches to learning spatial and spectral resolution from HRMS and LRHS, respectively. Due to the high dimension of HS images, thus leads to high computational cost. Differently, Dian *et al.* [28] propose a method using traditional optimization as the basic framework and using CNN to learning the priors. It is effective for HS-MS fusion and takes lower time cost.

### III. PROPOSED METHOD

The proposed NLRSR method will be elaborated in this section. First, some preliminaries about tensor are introduced. Then, we describe the problem formulation of HS-MS fusion. Next, the nonlocal low-rank constraint and the sparse constraint criterions are described in detail. Finally, the proposed model is explained.

#### A. Introduction to Tensor

Tensor is known as a  $N$ -dimensional generalization of a 2-dimensional matrix. And tensor representation has been proved to be stable in [32]. In this paper, we present tensors in Euler script letters  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  and matrices using bold Greek script letters  $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ . Mode- $n$  unfolding of tensor  $\mathcal{X}$  is a process of rearranging the tensor mode- $n$  fibers, and it is written as

$$\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}. \quad (1)$$

Besides,  $\text{Unfold}_n(\mathcal{X}) = \mathbf{X}_{(n)}$  and  $\text{Fold}_n(\mathbf{X}_{(n)}) = \mathcal{X}$ . Similarly, mode- $n$  canonical of tensor  $\mathcal{X}$  also rearranges a tensor element to obtain a matrix, and it is written as

$$\mathbf{X}_{(n)} \in \mathbb{R}^{(I_1 \dots I_n) \times (I_{n+1} \dots I_N)}. \quad (2)$$

Using the entry of tensor  $x_{i_1 i_2 \dots i_N}$ , the Frobenius norm

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \dots i_N}^2}. \quad (3)$$

can be computed. Table I summarizes the notations used in this paper.

#### B. Problem Formulation

We use  $\mathcal{Z} \in \mathbb{R}^{W \times H \times L}$  to denote the desired HRHS image, where  $W$  is the width,  $H$  is the height, and  $L$  is the number of spectral bands.  $\mathcal{X} \in \mathbb{R}^{w \times h \times L}$  presents the LRHS image, where  $w < W$  and  $h < H$ .  $\mathcal{Y} \in \mathbb{R}^{W \times H \times l}$  denotes the HRMS image, where  $l < L$ . To solve the image fusion problem, we need to estimate  $\mathcal{Z}$  using  $\mathcal{X}$  and  $\mathcal{Y}$ . Actually, it is reasonable to assume that the HRMS image is the spectrally degraded version of the HRHS image, so  $\mathcal{X}$  can be modeled using  $\mathcal{Z}$  as

$$\mathbf{X}_{(3)} = \mathbf{Z}_{(3)} \mathbf{B} \mathbf{S}, \quad (4)$$

where  $\mathbf{X}_{(3)} \in \mathbb{R}^{L \times wh}$  and  $\mathbf{Z}_{(3)} \in \mathbb{R}^{L \times WH}$  are the mode-3 unfolding of  $\mathcal{X}$  and  $\mathcal{Z}$ , respectively.  $\mathbf{B} \in \mathbb{R}^{WH \times WH}$  presents the spatial blurring operator, and  $\mathbf{S} \in \mathbb{R}^{WH \times wh}$  presents the spatial down-sampling operator. Similarly, the LRHS image can be regarded as the spatially degraded version of the HRHS image and  $\mathcal{Y}$  can be modeled using  $\mathcal{Z}$  as

$$\mathbf{Y}_{(3)} = \mathbf{R} \mathbf{Z}_{(3)}, \quad (5)$$

where  $\mathbf{Y}_{(3)} \in \mathbb{R}^{l \times WH}$ .  $\mathbf{R} \in \mathbb{R}^{l \times L}$  is the spectral response function of  $\mathcal{Y}$ . Therefore, the basic HS-MS fusion model can be formulated as

$$\min_{\mathbf{Z}_{(3)}} \|\mathbf{X}_{(3)} - \mathbf{Z}_{(3)} \mathbf{B} \mathbf{S}\|_F^2 + \lambda \|\mathbf{Y}_{(3)} - \mathbf{R} \mathbf{Z}_{(3)}\|_F^2, \quad (6)$$

where  $\lambda$  is a tradeoff parameter.

#### C. Nonlocal Low-Rank Tensor Approximation

To exploit the local structure information and the nonlocal self-similarity of HS cube, we use nonlocal low rank tensor approximation to regularize the basic fusion model. To do that,  $\mathcal{Z}$  is firstly divided into several overlap cubes. Using the Euclidean distance to measure the similarity of HS image pixels, the similar cubes are grouped together into  $K$  clusters  $\mathcal{C}^k = \{\mathcal{Z}^{(k,i)} \in \mathbb{R}^{b \times b \times L}\}_{i=1}^{N_k}$  ( $k = 1, \dots, K$ ), where  $b$  is the height or width of each cube and  $N_k$  is the quantity of cubes in the  $k$ th cluster. Since  $\mathcal{Z}$  is unknown at first and most of its spatial information is acquired from  $\mathcal{Y}$ ,  $\mathcal{Z}$  is grouped according to the grouping of  $\mathcal{Y}$ . It is worth noting that  $\mathcal{C}^k \in \mathbb{R}^{b \times b \times L \times K}$  is a 4D tensor. We define the operation from  $\mathcal{Z}$  to  $\mathcal{C}^k$  as  $\mathcal{G}_{\mathcal{Z}}^k$ , that is  $\mathcal{G}_{\mathcal{Z}}^k = \mathcal{C}^k$ . Since each cluster  $\mathcal{C}^k$  is composed of  $N_k$  HS cubes, it owns the same structure and features of HS images. First, its pixels are spatial correlation, so  $\mathbf{C}_{(1)}^k$  (the mode-1 unfolding of the tensor  $\mathcal{C}^k$ ) and  $\mathbf{C}_{(2)}^k$  are low-rank. Second, the spectral information of its neighboring bands are redundant, so  $\mathbf{C}_{(3)}^k$  is low-rank. Considering  $N_k$  HS cubes in each cluster is similar, the nonlocal self-similarity means  $\mathbf{C}_{(4)}^k$  is low-rank. Thus, for each cluster, the low-rank structure of  $\mathcal{C}^k$  can be expressed as

$$\sum_{i=1}^4 \alpha_i \text{rank}(\mathbf{C}_{(i)}^k), \quad (7)$$

where  $\alpha_i$  is the weight coefficient and  $\sum_{i=1}^4 \alpha_i = 1$ . It is confirmed that the TT format can be used to effectively solve high dimensional problems [29]. To be specific, TT format

is relatively stable and has a lower variance of ranks, which makes it faster and easier to reach extreme value. The TT low-rank structure of  $\mathcal{C}^k$  can be formulated as

$$\sum_{i=1}^3 \alpha_i \text{rank}(\mathbf{C}_{(i)}^k), \quad (8)$$

where  $\mathbf{C}_{(i)}^k$  is the mode- $i$  canonical of the tensor  $\mathcal{C}^k$  and  $\sum_{i=1}^3 \alpha_i = 1$ .

Since the tensor rank is usually non-convex and its optimization problem is NP-hard, the nuclear norm calculated by summing the singular values of a matrix is often used to replace it in practice. Also, authors in [25] use a logarithm term as a solution to the problem. The special logarithm term shrinks smaller singular value with noise, and this method achieves high fusion performance. However, the logarithm term may not approximate the rank function well when the singular value is large. A folded-concave penalty, minimax concave plus (MCP) penalty, is introduced in [35]. It not only shrinks smaller singular values but also keeps other singular values close to 1. Thus, it gives a better approximation than the logarithm term. This penalty term is written as

$$P_\lambda(t) = \begin{cases} a\lambda^2/2, & \text{if } |t| \geq a\lambda; \\ \lambda|t| - t^2/2a, & \text{otherwise.} \end{cases} \quad (9)$$

As in [36], we define the MCP norm as  $\|\mathbf{M}\|_{P_\lambda} = \sum_{i=1}^r P_\lambda(\sigma_i(\mathbf{M}))$ , where  $\sigma_i(\mathbf{M})$  is the  $i$ th singular value and  $r$  is the rank of matrix  $\mathbf{M}$ . Then, using above MCP norm, the low-rank structure of  $\mathcal{C}^k$  can be rephrased as

$$\|\mathcal{C}^k\|_{P_\lambda} = \sum_{i=1}^3 \alpha_i \|\mathbf{C}_{(i)}^k\|_{P_\lambda}. \quad (10)$$

Finally, the nonlocal low-rank tensor approximation term is written as

$$\|\mathcal{Z}\|_{NLR} = \sum_{k=1}^K \|\mathcal{C}^k\|_{P_\lambda}. \quad (11)$$

#### D. Dictionary Learning and Sparse Coefficient Matrix

It has been proved that sparse prior can effectively solve a variety of ill-posed inverse problems under the background of HS image reconstruction. Before using the sparse prior, the linear spectral mixing model is described. It assumes that the image can be decomposed into the product of the spectral dictionary and the corresponding coefficient. As to  $\mathcal{Z}$ , it can be modeled as

$$\mathbf{Z}_{(3)} = \mathbf{E}\mathbf{A}, \quad (12)$$

where  $\mathbf{E}$  is the spectral dictionary and  $\mathbf{A}$  is the corresponding coefficient. And the coefficient  $\mathbf{A}$  which represents the proportion of each endmember in the image may be sparse. Thus, we use the element-wise  $l_1$  norm regularizer term  $\|\mathbf{A}\|_1$  to describe the columns sparsity of  $\mathbf{A}$ . After substituting equation (12) into equation (4), the following expression is obtained

$$\mathbf{X}_{(3)} = \mathbf{E}\mathbf{A}\mathbf{B}\mathbf{S} = \mathbf{E}\tilde{\mathbf{A}}, \quad (13)$$

where  $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{B}\mathbf{S}$  is the abundance coefficient of  $\mathbf{X}_{(3)}$ . Then, the spectral dictionary  $\mathbf{E}$  can be learned from the minimization problem below

$$\min_{\mathbf{E}} \|\mathbf{X}_{(3)} - \mathbf{E}\tilde{\mathbf{A}}\|_2 + \lambda_l \|\tilde{\mathbf{A}}\|_1, \quad (14)$$

where  $\lambda_l$  is a penalty parameter. We use the online dictionary learning algorithm which is proposed in [37] to solve it.

#### E. Proposed Model

Combined with the above discussion, the HS-MS fusion problem can be expressed as the following minimization problem:

$$\begin{aligned} \min_{\mathbf{Z}_{(3)}} & \|\mathbf{X}_{(3)} - \mathbf{Z}_{(3)}\mathbf{B}\mathbf{S}\|_F^2 + \lambda \|\mathbf{Y}_{(3)} - \mathbf{R}\mathbf{Z}_{(3)}\|_F^2 \\ & + \sum_{k=1}^K \sum_{i=1}^3 \alpha_i \|\mathbf{C}_{(i)}^k\|_{P_{\lambda_1}} + \lambda_2 \|\mathbf{A}\|_1 \\ \text{s.t. } & \mathbf{Z}_{(3)} = \mathbf{E}\mathbf{A}, \end{aligned} \quad (15)$$

where  $\lambda_1$  and  $\lambda_2$  are regularization parameters.

#### IV. OPTIMIZATION ALGORITHM

The optimization algorithm for solving the proposed NLRSR model is presented in this section. Since (15) is a nonconvex optimization problem, we use alternative direction method of multipliers (ADMM) to find the optimal structure. First, we introduce auxiliary variables  $\{\mathcal{M}_i\}_{i=1}^3$ , so (15) can be rewritten as

$$\begin{aligned} \min_{\mathbf{Z}_{(3)}} & \|\mathbf{X}_{(3)} - \mathbf{Z}_{(3)}\mathbf{B}\mathbf{S}\|_F^2 + \lambda \|\mathbf{Y}_{(3)} - \mathbf{R}\mathbf{Z}_{(3)}\|_F^2 \\ & + \sum_{k=1}^K \sum_{i=1}^3 \alpha_i \|\mathcal{G}_{\mathcal{M}_i(i)}^k\|_{P_{\lambda_1}} + \lambda_2 \|\mathbf{A}\|_1 \\ \text{s.t. } & \mathcal{Z} = \mathcal{M}_i, \mathbf{Z}_{(3)} = \mathbf{E}\mathbf{A}, i = 1, \dots, 3. \end{aligned} \quad (16)$$

Then, the augmented Lagrangian expression for (16) is

$$\begin{aligned} L(\mathcal{Z}, \mathcal{M}_i, \mathcal{U}_i, \mathbf{A}, \mathbf{V}) & = \|\mathbf{X}_{(3)} - \mathbf{Z}_{(3)}\mathbf{B}\mathbf{S}\|_F^2 + \lambda \|\mathbf{Y}_{(3)} - \mathbf{R}\mathbf{Z}_{(3)}\|_F^2 \\ & + \sum_{i=1}^3 \sum_{k=1}^K \alpha_i \|\mathcal{G}_{\mathcal{M}_i(i)}^k\|_{P_{\lambda_1}} + \lambda_2 \|\mathbf{A}\|_1 \\ & + \sum_{i=1}^3 (\langle \mathcal{Z} - \mathcal{M}_i, \mathcal{U}_i \rangle + \frac{\mu}{2} \|\mathcal{Z} - \mathcal{M}_i\|_F^2) \\ & + \langle \mathbf{E}\mathbf{A} - \mathbf{Z}_{(3)}, \mathbf{V} \rangle + \frac{\nu}{2} \|\mathbf{Z}_{(3)} - \mathbf{E}\mathbf{A}\|_F^2, \end{aligned} \quad (17)$$

where  $\mu$  and  $\nu$  are penalty parameters.  $\{\mathcal{U}_i\}_{i=1}^3$  and  $\mathbf{V}$  are the Lagrange multipliers. By extracting the terms that is related to a variable, (17) can be split into the following sub-problems.

1) *Optimization of  $\mathcal{Z}$* : By fixing  $\{\mathcal{M}_i\}_{i=1}^3$ ,  $\{\mathcal{U}_i\}_{i=1}^3$ ,  $\mathbf{A}$ ,  $\mathbf{V}$ , the sub-problem of  $\mathcal{Z}$  can be written as

$$\begin{aligned} \min_{\mathbf{Z}_{(3)}} & \|\mathbf{X}_{(3)} - \mathbf{Z}_{(3)}\mathbf{B}\mathbf{S}\|_F^2 + \lambda \|\mathbf{Y}_{(3)} - \mathbf{R}\mathbf{Z}_{(3)}\|_F^2 \\ & + \sum_{i=1}^3 (\langle \mathcal{Z} - \mathcal{M}_i, \mathcal{U}_i \rangle + \frac{\mu}{2} \|\mathcal{Z} - \mathcal{M}_i\|_F^2) \\ & + \langle \mathbf{E}\mathbf{A} - \mathbf{Z}_{(3)}, \mathbf{V} \rangle + \frac{\nu}{2} \|\mathbf{Z}_{(3)} - \mathbf{E}\mathbf{A}\|_F^2. \end{aligned} \quad (18)$$

The minimization problem (18) has a closed-form solution, which is able to obtain by taking its derivative and setting it to zero. The optimal solution of  $\mathbf{Z}$  is equivalent to a solution to the equation as follows

$$\begin{aligned} & \mathbf{R}^T \mathbf{R} \mathbf{Z}_{(3)} + \mathbf{Z}_{(3)} \mathbf{B} \mathbf{S} (\mathbf{B} \mathbf{S})^T + 3\mu \mathbf{Z}_{(3)} + \nu \mathbf{Z}_{(3)} \\ & = \mathbf{X}_{(3)} (\mathbf{B} \mathbf{S})^T + \mathbf{R}^T \mathbf{Y}_{(3)} \\ & + \sum_{i=1}^3 \mu (\mathbf{M}_{i(3)} - \frac{\mathbf{U}_{i(3)}}{\mu}) + \nu (\mathbf{E} \mathbf{A} - \frac{\mathbf{V}}{\nu}). \end{aligned} \quad (19)$$

Equation (19) is a Sylvester matrix equation. And it can be written as

$$\mathbf{H}_1 \mathbf{Z}_{(3)} + \mathbf{Z}_{(3)} \mathbf{H}_2 = \mathbf{H}_3, \quad (20)$$

where

$$\mathbf{H}_1 = \mathbf{R}^T \mathbf{R} + 3\mu \mathbf{I} + \nu \mathbf{I}, \quad (21)$$

$$\mathbf{H}_2 = \mathbf{B} \mathbf{S} (\mathbf{B} \mathbf{S})^T, \quad (22)$$

$$\begin{aligned} \mathbf{H}_3 &= \mathbf{X}_{(3)} (\mathbf{B} \mathbf{S})^T + \mathbf{R}^T \mathbf{Y}_{(3)} \\ &+ \sum_{i=1}^3 \mu (\mathbf{M}_{i(3)} - \frac{\mathbf{U}_{i(3)}}{\mu}) + \nu (\mathbf{E} \mathbf{A} - \frac{\mathbf{V}}{\nu}). \end{aligned} \quad (23)$$

For the Sylvester equation (20), we use R-FUSE [20] to solve it.

2) *Optimization of  $\mathcal{M}_i$  ( $i = 1, \dots, 3$ ):* By fixing  $\mathcal{Z}$ ,  $\{\mathcal{U}_i\}_{i=1}^3$ ,  $\mathbf{A}$ ,  $\mathbf{V}$ , the sub-problem of  $\{\mathcal{M}_i\}_{i=1}^3$  can be written as

$$\min_{\mathcal{M}_i} \sum_{k=1}^K \alpha_i \|\mathcal{G}_{\mathcal{M}_i}^k\|_{P_{\lambda_1}} + \langle \mathcal{Z} - \mathcal{M}_i, \mathcal{U}_i \rangle + \frac{\mu}{2} \|\mathcal{Z} - \mathcal{M}_i\|_F^2. \quad (24)$$

Considering the first term, the linear proximal is first conducted on (24). After that, it can be solved by the weighted singular value thresholding as in [38]. For the singular value decomposition of a matrix  $\mathbf{Z} = \mathbf{U}_Z \Sigma_Z \mathbf{V}_Z^T$ , the corresponding weighted singular value thresholding operator is calculated by

$$\text{SoftTh}(\mathbf{Z}, \tau, w) = \mathbf{U}_Z \text{Sign}(\Sigma_Z) (|\Sigma_Z| - \frac{\tau}{2} w)_+ \mathbf{V}_Z^T, \quad (25)$$

where  $\text{Sign}(\cdot)$  is the sign function and  $w = \text{Diag}((\lambda - (\sigma(\mathbf{Z})/a))_+)$ . Thus, the optimization of  $\mathcal{M}_i$  ( $i = 1, \dots, 3$ ) can be computed by

$$\mathcal{M}_i = \sum_{k=1}^K \text{Fold}_i(\text{SoftTh}(\mathcal{G}_{\mathcal{Z}}^k + \frac{\mathcal{G}_{\mathcal{U}_i}^k}{\mu}, \frac{\lambda_1 \alpha_i}{\mu})), \quad (26)$$

where  $\text{Fold}_i(\cdot)$  is the mode- $i$  folding of a tensor.

3) *Optimization of  $\mathbf{A}$ :* By fixing  $\mathcal{Z}$ ,  $\{\mathcal{M}_i\}_{i=1}^3$ ,  $\{\mathcal{U}_i\}_{i=1}^3$ ,  $\mathbf{V}$ , the sub-problem of  $\mathbf{A}$  can be written as

$$\min_{\mathbf{A}} \lambda_2 \|\mathbf{A}\|_1 + \langle \mathbf{E} \mathbf{A} - \mathbf{Z}_{(3)}, \mathbf{V} \rangle + \frac{\nu}{2} \|\mathbf{Z}_{(3)} - \mathbf{E} \mathbf{A}\|_F^2. \quad (27)$$

Problem (27) can be written as the following form

$$\min_{\mathbf{A}} \|\mathbf{E} \mathbf{A} - \mathbf{B}\|_2^2 + \lambda \|\mathbf{A}\|_1. \quad (28)$$

After expanding function (28) and deriving, the distribution of its minimum value is obtained by

$$\text{SoftExp}(\mathbf{B}, \mathbf{E}, \lambda) = (\mathbf{E}^T \mathbf{E})^{-1} \text{Sign}(\mathbf{E}^T \mathbf{B}) (|\mathbf{E}^T \mathbf{B}| - \frac{\lambda}{2})_+. \quad (29)$$

Thus, we can directly update  $\mathbf{A}$  by

$$\mathbf{A} = \text{SoftExp}(\mathbf{Z}_{(3)} - \frac{\mathbf{V}}{\nu}, \mathbf{E}, \frac{2\lambda_2}{\nu}). \quad (30)$$

4) *Updating Multipliers:* The multipliers  $\{\mathcal{U}_i\}_{i=1}^3$  and  $\mathbf{V}$  are updated by equations as follows

$$\mathcal{U}_i = \mathcal{U}_i + \mu (\mathcal{Z} - \mathcal{M}_i), i = 1, \dots, 3, \quad (31)$$

$$\mathbf{V} = \mathbf{V} + \nu (\mathbf{E} \mathbf{A} - \mathbf{Z}_{(3)}). \quad (32)$$

In each iteration,  $\mu$  and  $\nu$  increase with a small value.

$$\mu = \gamma \mu, \quad (33)$$

$$\nu = \gamma \nu, \quad (34)$$

where  $\gamma$  is a little larger than 1. The steps of the proposed NLRSSR are summarized in Algorithm 1.

---

#### Algorithm 1 ADMM for Solving NLRSSR Fusion

---

**Input:** LRHS image  $\mathcal{X}$ , HRMS image  $\mathcal{Y}$ , spatial blurring operator  $\mathbf{B}$ , spatial down-sampling matrix  $\mathbf{S}$ , spectral response matrix  $\mathbf{R}$ , parameters  $K$ ,  $\lambda$ ,  $\lambda_1$ , and  $\lambda_2$ ;

**Initialization:**  $\mathcal{M}_1 = 0$ ,  $\mathcal{M}_2 = 0$ ,  $\mathcal{M}_3 = 0$ ,  $\mathcal{U}_1 = 0$ ,  $\mathcal{U}_2 = 0$ ,  $\mathcal{U}_3 = 0$ ,  $\mathbf{V} = \mathbf{0}$ ,  $\mathbf{A} = \mathbf{1}/m$ ,  $\mu = 10^{-3}$ ,  $\nu = 10^{-6}$ ,  $\gamma = 1.05$ ,  $l = 1$ ,  $maxIter = 100$ ;

- 1: Divide  $\mathcal{Z}$  into several overlapping blocks;
- 2: Group overlapping blocks into  $K$  clusters;
- 3: Learn  $\mathbf{E}$  from  $\mathcal{Y}$  by online dictionary learning;
- 4: **while not converged** and  $l < maxIter$  **do**
- 5:   Update  $\mathcal{Z}$  with R-FUSE on (18);
- 6:   **for**  $i = 1$  to 3 **do**
- 7:     Update  $\mathcal{M}_i$  by (26);
- 8:     Update  $\mathcal{U}_i$  by (31);
- 9:   **end for**
- 10:   Update  $\mathbf{A}$  by (30);
- 11:   Update  $\mathbf{V}$  by (32);
- 12:   Update  $\mu$  by (33) and update  $\nu$  (34);
- 13:   Check the convergence conditions;
- 14:    $l = l + 1$ ;
- 15: **end while**

**Output:** fused image  $\mathcal{Z}$ .

---

## V. EXPERIMENTS

To evaluate the performance of our algorithm, adequate experiments are conducted in this section. Firstly, four data sets and four quality metrics are introduced. Then, we show the experimental results of eight comparison methods. Finally, we analyze the parameter selection of our method.

### A. Experiment Data Sets

We use three synthetic HS remote sensing data sets and one real HS remote sensing data set for experiments, and details of them are described below.

1) *University of Pavia Data Set*<sup>1</sup>: This data set was collected by the Reflective Optics System Imaging Spectrometer

<sup>1</sup>[http://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes)

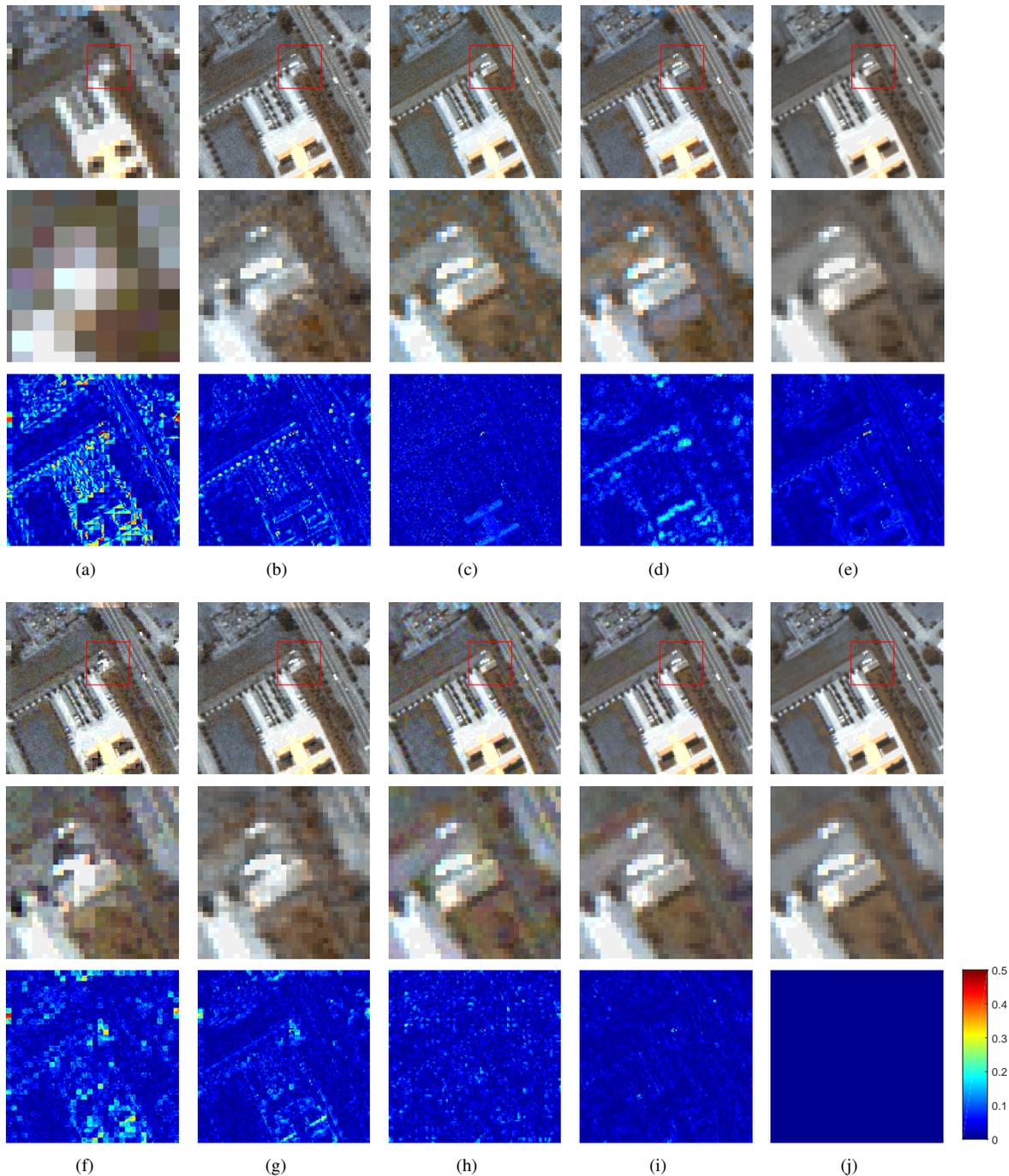


Fig. 1. HS-MS Fusion results among the compared methods on University of Pavia data set. (a) LRHS. (b) CNMF. (c) GSOMP+. (d) HYSURE. (e) SpaBayes. (f) NSSR. (g) CSTF. (h) LTTR. (i) NLRSR. (j) Reference.

(ROSIS) sensor over Pavia, Italy in 2003. After removal of the water vapor absorption and noise bands, the original 224 spectral bands between 0.43 and 0.84  $\mu\text{m}$  remains 103 bands. The full spatial resolution of this image is  $610 \times 610$  pixels with a ground sampling distance (GSD) of 1.3 m and we select a  $128 \times 128$ -pixels-size image in the experiment.

2) *Washington DC Mall Data Set*<sup>2</sup>: The Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor acquired

the data set over the National Mall in Washington, DC in 1995. The HS image consists of  $1280 \times 307$  pixels with 210 bands within a wavelength range of 0.4 to 2.5  $\mu\text{m}$ . Its GSD is 2.5 m. After removal of the water vapor absorption bands between 0.9 and 1.4  $\mu\text{m}$ , the image reduces to 191 bands. We select a  $200 \times 200$ -pixels-size image for the experiment.

3) *Chikusei Data Set*<sup>3</sup>: This image was captured by Headwall Hyperspec-VNIR-C imaging sensor over Chikusei,

<sup>2</sup><https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

<sup>3</sup><https://webpark1587.sakura.ne.jp/hyperdata/>

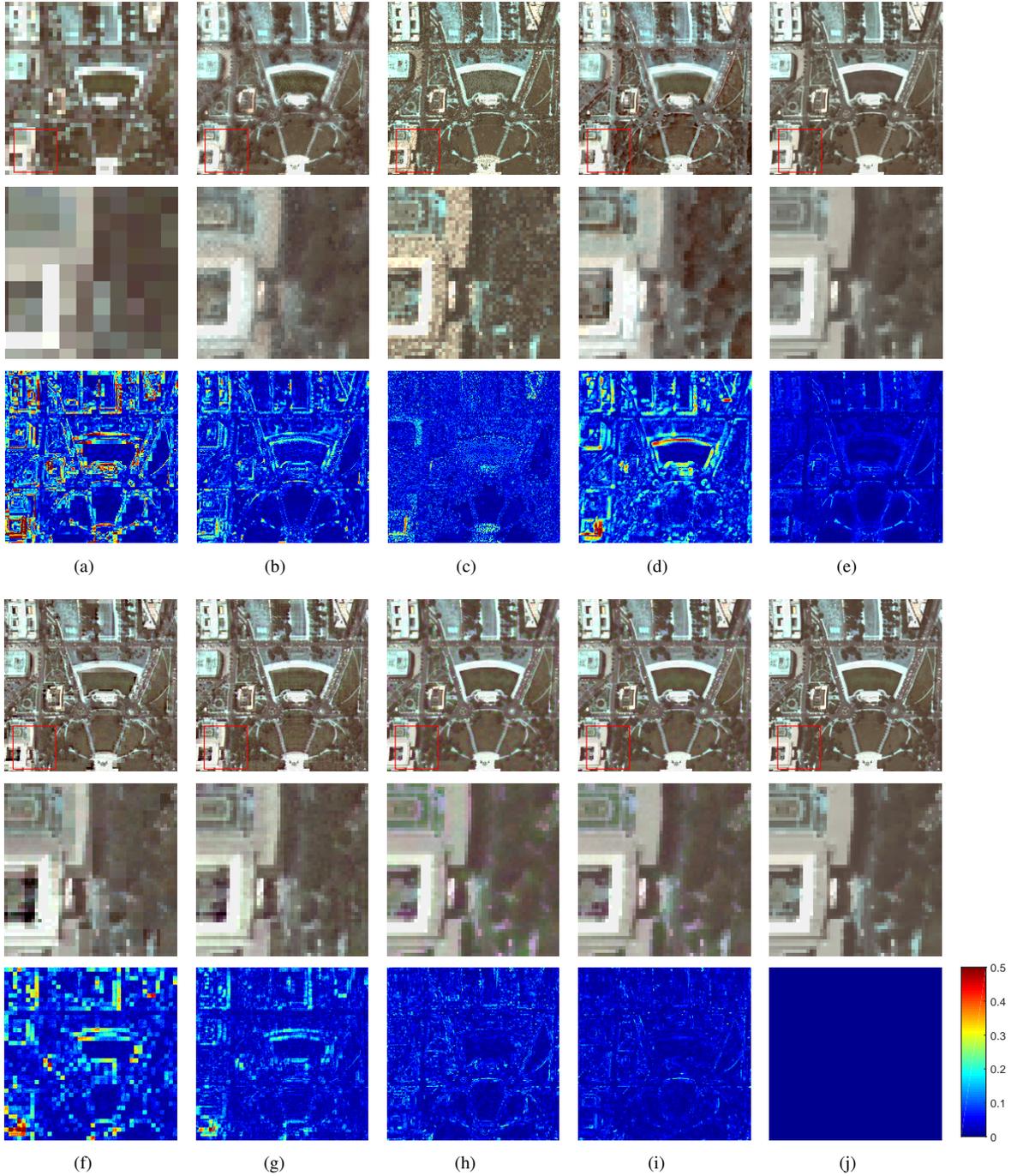


Fig. 2. HS-MS Fusion results among the compared methods on Washington DC Mall data set. (a) LRHS. (b) CNMF. (c) GSOMP+. (d) HYSURE. (e) SpaBayes. (f) NSSR. (g) CSTF. (h) LTTR. (i) NLRSR. (j) Reference.

Ibaraki, Japan in 2014. This HS image has 128 bands between 0.36 and 1.02  $\mu\text{m}$  and the spatial resolution of  $2517 \times 2335$  pixels with a GSD of 2.5 m. In the experiment, we choose a  $240 \times 240$ -pixels-size image.

For each of the first three synthetic HS data sets, the original HS image is treated as a reference image for comparing to the fused image. The LRHS image can be obtained by bluing the original HS image using a  $5 \times 5$  Gaussian filter, and down-sampling in a certain ratio. For three different HS data sets, the ratios are 4, 5 and 6, respectively. Similarly, the HRMS

image is acquired by down-sampling the original HS image using the IKONOS2 spectral response function<sup>4</sup>. In addition, Gaussian noise is added to LRHS bands (SNR = 35 dB) and HRMS bands (SNR = 30 dB).

4) *Real Data Set*<sup>5</sup>: The data set consists of an HS image and an MS image. The HS image was acquired by Hyperion instrument on board the Earth-Observing One (EO-1) satellite over Paris, France on July 29, 2002. It has 242 spectral

<sup>4</sup>[https://fsf.nerc.ac.uk/user\\_group/bands/IKONOS2.xml](https://fsf.nerc.ac.uk/user_group/bands/IKONOS2.xml)

<sup>5</sup><http://eo1.usgs.gov> and <http://eo1.gsfc.nasa.gov>

TABLE II  
EXPERIMENTAL EVALUATION METRICS AMONG EIGHT COMPARED  
METHODS ON UNIVERSITY OF PAVIA DATA SET.

Method	University of Pavia Data Set			
	PSNR	SAM	ERGAS	Q2 <sup>n</sup>
CNMF	31.224	3.225	2.254	0.557
GSOMP	34.175	3.021	1.601	0.700
HYSURE	31.674	3.709	2.253	0.694
SpaBayes	35.114	2.479	1.442	0.480
NSSR	28.837	3.992	2.948	0.519
CSTF	32.640	2.578	1.903	0.652
LTTR	34.673	3.313	1.522	0.651
NLRSR	<b>38.275</b>	<b>1.942</b>	<b>1.009</b>	<b>0.821</b>

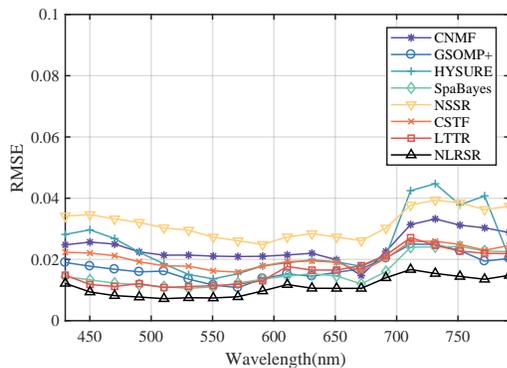


Fig. 3. RMSEs among the compared methods on University of Pavia data set.

bands between 0.36 and 2.58  $\mu\text{m}$  with a GSD of 30 m. After removal of bands with low signal-to-noise ratio, the HS image remains 166 available bands. The MS image was collected by Advanced Land Imager (ALI) instrument on board the EO-1 satellite over Paris, France on the same day of the HS image. It consists of 9 spectral bands (0.4 to 2.5  $\mu\text{m}$ ) with a GSD of 30 m and we choose the 3rd, 6th and 8th bands from it. In the experiment, we select  $150 \times 150$ -pixels-size sub-images for both images.

For the real data set, we treat the HS image as the reference image and the MS image as the HRMS image. To obtain the LRHS image, the HS image is down-sampled at a ratio of 3.

### B. Experiment Setup

In the experiments, a detailed comparison is made between the proposed approach and seven selected state-of-art approaches, which include CNMF [11], GSOMP+ [12], HYSURE [13], SpaBayes [17], NSSR [15], CSTF [22] and LTTR [25]. The parameters of CNMF, GSOMP, HYSURE and SpaBayes are setting according to the original code. In this case, these algorithms are stable and have good performance in all data sets. As to NSSR, the parameter  $K$  which represents the number of atoms in dictionary is set to 40, and the regularization parameters  $\eta_1$  and  $\eta_2$  are both set to  $10^{-4}$ . For CSTF, its key parameters  $n_w$ ,  $n_h$  and  $n_s$  represent the number of atoms of dictionary of spatial modes and spectral mode, respectively. We set  $n_w = 260$ ,  $n_h = 260$  and  $n_s = 15$  for all three data sets. The sparsity regularization parameter

TABLE III  
EXPERIMENTAL EVALUATION METRICS AMONG EIGHT COMPARED  
METHODS ON WASHINGTON DC MALL DATA SET.

Method	Washington DC Mall Data Set			
	PSNR	SAM	ERGAS	Q2 <sup>n</sup>
CNMF	27.166	3.304	1.929	0.269
GSOMP	26.081	3.999	2.176	0.488
HYSURE	24.419	5.593	2.722	0.438
SpaBayes	29.956	3.183	<b>1.364</b>	0.347
NSSR	24.288	5.200	2.672	0.290
CSTF	26.534	3.863	1.980	0.438
LTTR	29.741	3.268	1.414	0.626
NLRSR	<b>30.084</b>	<b>2.794</b>	1.403	<b>0.671</b>

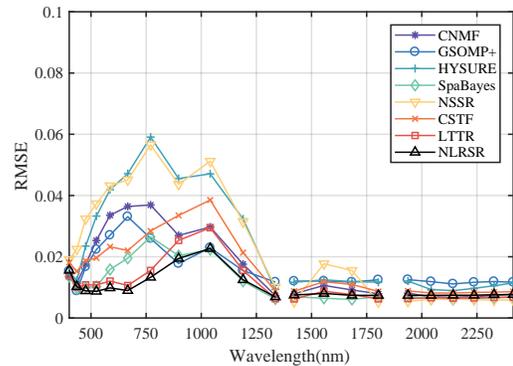


Fig. 4. RMSEs among the compared methods on Washington DC Mall data set.

$\lambda$  is set to  $10^{-5}$ . For LTTR, there are two key parameters. One is the parameter  $K$  which is the number of clusters and is set to 450, the other is the parameter  $\lambda$  which is used to adjust the weight of LTTR regularization term and is set to  $10^{-2}$ . Additionally, the parameter setting of our approach is as follows: the regularization parameters  $\lambda = 1$ ,  $\lambda_1 = 10^{-2}$  and  $\lambda_2 = 10^{-4}$ , the spatial size of HS cube is  $8 \times 8$ , and the overlap between neighboring cubes is 2.

To assess the difference in performance between the proposed method and the competitors, the following four quantitative metrics are adopt: peak signal to noise ratio (PSNR), spectral angle mapping (SAM), erreur relative globale adimensionnelle desynthese (ERGAS), and Q2<sup>n</sup>. To define PSNR, we use  $\mathbf{Z}^i$  ( $i = 1, \dots, L$ ) to represent the  $i$ th band fused image  $\mathcal{Z}$  and use  $\mathbf{R}^i$  to represent the  $i$ th band reference image  $\mathcal{R}$ . Then, the first metric PSNR is defined as

$$\text{PSNR}(\mathbf{Z}^i, \mathbf{R}^i) = 20 \log_{10} \left( \frac{\max(\mathbf{Z}^i)}{\sqrt{P} \cdot \text{RMSE}(\mathbf{Z}^i, \mathbf{R}^i)} \right), \quad (35)$$

where root-mean-square error (RMSE) is defined as

$$\text{RMSE}(\mathbf{Z}^i, \mathbf{R}^i) = \frac{1}{P} \sqrt{\|\mathbf{Z}^i - \mathbf{R}^i\|_2^2}. \quad (36)$$

Let  $\mathbf{V}_{\mathbf{Z}}^j$  ( $j = 1, \dots, P$ ,  $P = W \times H$ ) denote the  $j$ th pixel spectral signature in  $\mathbf{Z}$  and  $\mathbf{V}_{\mathbf{R}}^j$  denote the  $j$ th pixel spectral signature in  $\mathbf{R}$ . The second metric SAM, which is used to evaluate the spectral distortion, is defined as

$$\text{SAM}(\mathbf{V}_{\mathbf{Z}}^j, \mathbf{V}_{\mathbf{R}}^j) = \arccos \left( \frac{(\mathbf{V}_{\mathbf{Z}}^j)^T \mathbf{V}_{\mathbf{R}}^j}{\|\mathbf{V}_{\mathbf{Z}}^j\|_2 \cdot \|\mathbf{V}_{\mathbf{R}}^j\|_2} \right), \quad (37)$$

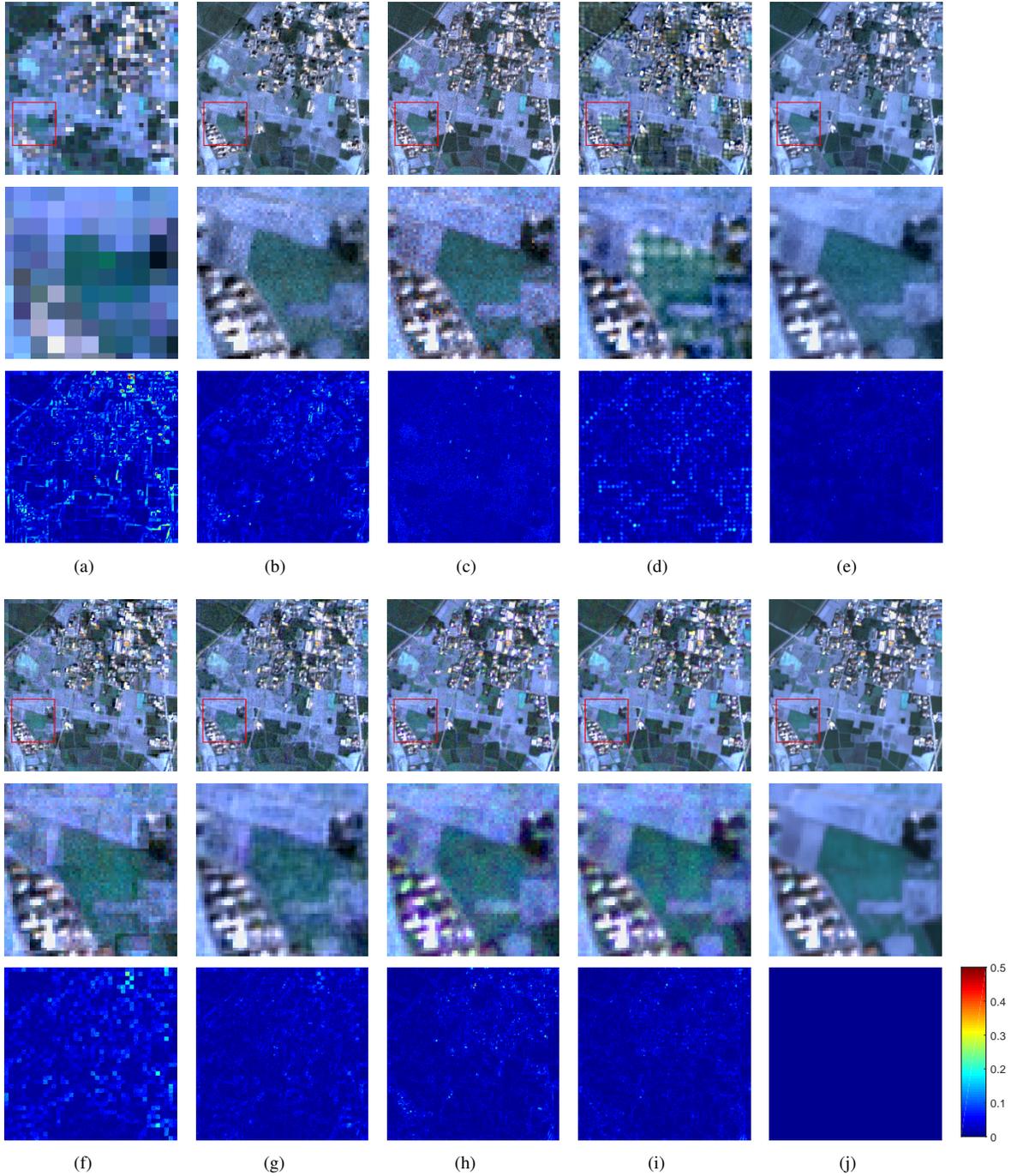


Fig. 5. HS-MS Fusion results among the compared methods on Chikusei data set. (a) LRHS. (b) CNMF. (c) GSOMP+. (d) HYSURE. (e) SpaBayes. (f) NSSR. (g) CSTF. (h) LTTR. (i) NLRSSR. (j) Reference.

ERGAS is adopt to estimate the image quality in spatial domain, which is defined as

$$\text{ERGAS}(\mathbf{Z}^i, \mathbf{R}^i) = \frac{100}{r} \sqrt{\frac{1}{B} \sum_{i=1}^B \left( \frac{\|\mathbf{Z}^i - \mathbf{R}^i\|_2^2}{\frac{1}{P} \sum_{i=1}^P \mathbf{R}^i} \right)^2}, \quad (38)$$

where  $r$  means the down-sampling rate between HRMS and LRHS images in the spatial domain.  $Q2^n$  is the general definition of universal image quality index (UIQI) and its

calculation formula is

$$Q(\mathbf{Z}^i, \mathbf{R}^i) = \frac{4\sigma_{\mathbf{Z}^i, \mathbf{R}^i} \overline{\mathbf{Z}^i} \cdot \overline{\mathbf{R}^i}}{(\sigma_{\mathbf{Z}^i}^2 + \sigma_{\mathbf{R}^i}^2)(\overline{\mathbf{Z}^i}^2 + \overline{\mathbf{R}^i}^2)}, \quad (39)$$

where  $\sigma_{\mathbf{Z}^i, \mathbf{R}^i}$  is the covariance of  $\mathbf{Z}^i$  and  $\mathbf{R}^i$ .  $\sigma_{\mathbf{Z}^i}$  and  $\sigma_{\mathbf{R}^i}$  are their standard deviations, and  $\overline{\mathbf{Z}^i}$  and  $\overline{\mathbf{R}^i}$  are their mean value, respectively.

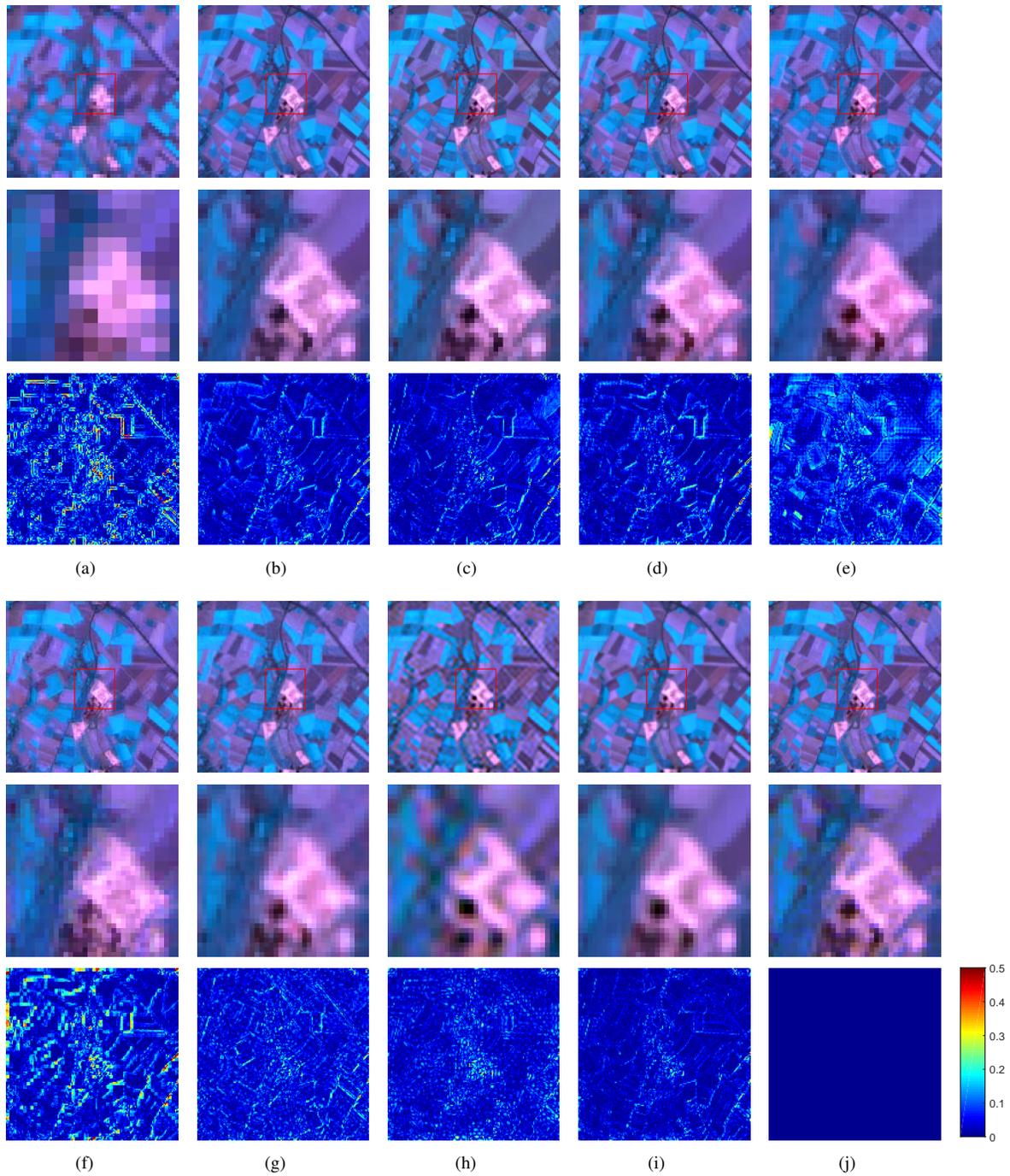


Fig. 6. HS-MS Fusion results among the compared methods on real data set. (a) LRHS. (b) CNMF. (c) GSOMP+. (d) HYSURE. (e) SpaBayes. (f) NCSR. (g) CSTF. (h) LTTR. (i) NLRSR. (j) Reference.

### C. Experiment Results

1) *University of Pavia Data Set*: Table II gives the fusion performance of compared methods on University of Pavia data set. The best values are marked in bold. Among all four quantitative indicators, the NLRSR achieves the best results. Fig. 1 shows a further visual comparison of these eight methods. And the difference image obtained by subtracting the reference image from the fused image is introduced to make the result more intuitive. In Fig. 1, pseudocolor images constructed by the 20th, 31st and 48th bands of fused images

are displayed in the first row. The sub-images marked with red boxes in the first row are magnified in the second row. The third row displays the difference images corresponding to the 48th band of fused images. In the fused images of CNMF and HYSURE, trees are fuzzy. There are some differences in the color of the grass in the color image of SpaBayes. After dividing the HS image into several disjoint blocks, the fusion operation of CSTF is conducted on HS blocks, thus there are some block errors in the difference image of Fig. 1(g). The difference image of our proposed method is the darkest,

TABLE IV  
EXPERIMENTAL EVALUATION METRICS AMONG EIGHT COMPARED  
METHODS ON CHIKUSEI DATA SET.

Method	Chikusei Data Set			
	PSNR	SAM	ERGAS	Q2 <sup>n</sup>
CNMF	28.676	3.531	1.601	0.591
GSOMP	29.919	3.644	1.412	0.646
HYSURE	28.675	3.604	1.618	0.628
SpaBayes	32.877	<b>2.381</b>	1.015	0.629
NSSR	27.490	4.204	1.813	0.530
CSTF	31.386	2.728	1.203	0.628
LTTR	33.629	2.809	0.930	0.743
NLRSR	<b>34.163</b>	2.640	<b>0.877</b>	<b>0.755</b>

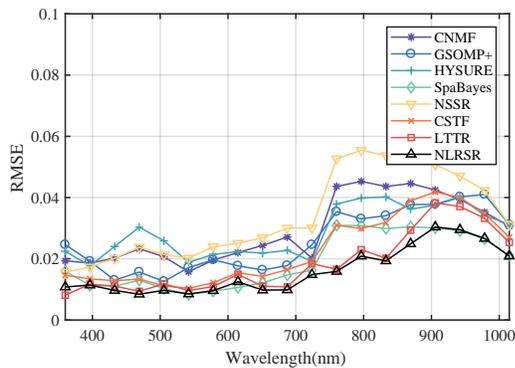


Fig. 7. RMSEs among the compared methods on Chikusei data set.

which demonstrates that the fused image obtained by NLRSR is closest to the reference. Fig. 3 shows RMSEs overall bands on University of Pavia data set. As can be observed in Fig. 3, the proposed method obtains the smallest RMSE values overall bands compared to seven competitors.

2) *Washington DC Mall Data Set*: Table III reports the average of the PSNR, SAM, ERGAS and Q2<sup>n</sup> results of compared methods on Washington DC Mall data set. For the NLRSR, only ERGAS is just above the minimum, and all other indicators achieve the best results. In general, NLRSR achieves the most satisfying results. The pseudocolor images composed of fused HS images at the 29th, 44th and 51st bands and difference images at the 44th band of all the competing methods are displayed in Fig. 2. There is an obvious color error in the fused image of GSOMP+ compared to the reference image. We can see that the proposed method achieves the smallest difference between the reference image and the reconstructed HS image. In Fig. 4, a comparison of RMSEs among these eight methods for Washington DC Mall data set is shown. The proposed NLRSR method gets the smallest RMSE values at the visible spectrum and slightly higher than NSSR and SpaBayes at the near-infrared spectrum (0.78-2.53  $\mu\text{m}$ ).

3) *Chikusei Data Set*: Table IV shows the four quantitative assessment results of the fused HS images on Chikusei data set. Overall, our method still gets the best fusion results compared to other methods. For visual comparison, pseudocolor images composed of fused images at 18th, 32nd, and 49th bands and difference images at the chosen 49th band are shown

TABLE V  
EXPERIMENTAL EVALUATION METRICS AMONG EIGHT COMPARED  
METHODS ON REAL DATA SET.

Method	Real Data Set			
	PSNR	SAM	ERGAS	Q2 <sup>n</sup>
CNMF	32.969	2.260	3.707	0.759
GSOMP	30.903	3.383	4.375	0.693
HYSURE	31.987	2.710	4.137	0.733
SpaBayes	32.502	2.956	3.613	0.680
NSSR	30.164	2.627	4.610	0.623
CSTF	32.114	2.735	4.084	0.706
LTTR	30.723	2.659	4.470	0.610
NLRSR	<b>33.836</b>	<b>2.004</b>	<b>3.401</b>	<b>0.762</b>

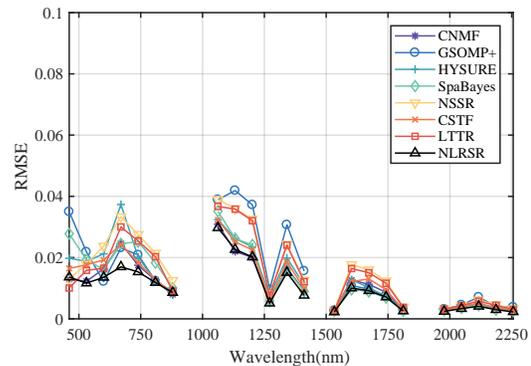


Fig. 8. RMSEs among the compared methods on real data set.

in Fig. 5. For the marked sub-images, there are obvious noise points in Fig. 5(b) and 5(c). Besides, several white spots appear in Fig. 5(d). Though SpaBayes gets the lowest SAM value, the fusion quality in the village is worse than our method. In the difference images, LTTR and NLRSR get the smallest errors. But there are fewer bright spots in the proposed NLRSR than that in LTTR. Fig. 7 shows the RMSEs plot overall HS bands on Chikusei data set. It is observed that NLRSR achieves optimal values on almost all bands.

4) *Real Data Set*: Table V shows the quality measures on real data set. It can be observed that NLRSR achieves the most satisfying result. Fig. 6 displays the fusion results for eight methods. In this figure, pseudocolor images consist of fused HS images at the 28th, 40th and 51st bands and difference images are calculated by the 40th band. The fused image of NSSR has obvious errors at the boundary of different fields in the farmland. Several block errors appear on the fused images of CSTF and LTTR. In Fig. 8, the RMSEs overall bands are plotted and our method gets the lowest RMSE in almost all bands. Thus, our NLRSR has better fusion performance than the other seven methods.

#### D. Parameter Selection

To assess the proposed method sensitivity of three key parameters, i.e. the number of clusters  $K$ , the regularization parameters  $\lambda_1$  and  $\lambda_2$ , we change their values to observe the effect on fusion results. In our algorithm,  $K$  uses to determine the number of clusters of self-similarity cubes. And the larger it is, the more clusters of HS cubes are

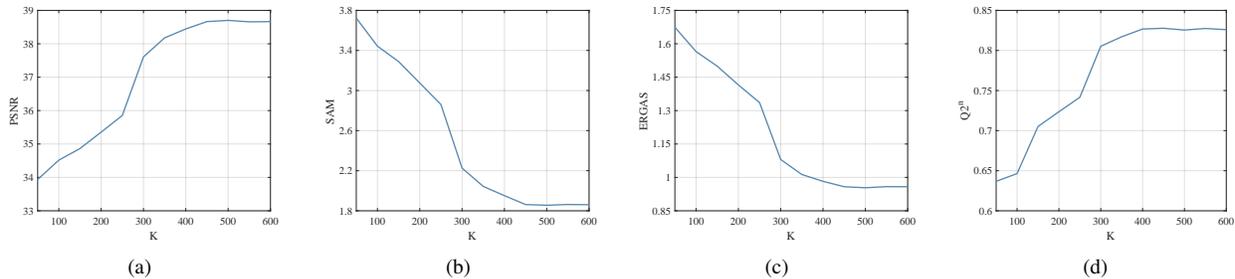


Fig. 9. Performance curves as a function of parameter K. (a) PSNR. (b) SAM. (c) ERGAS. (d)  $Q2^n$ .

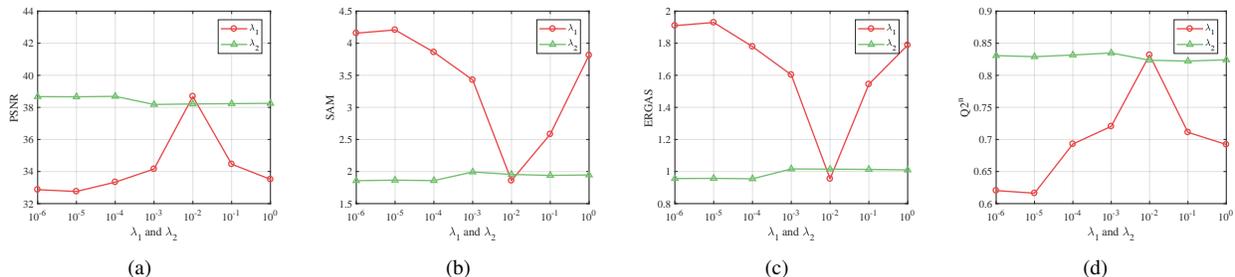


Fig. 10. Performance curves as a function of parameters  $\lambda_1$  and  $\lambda_2$ . (a) PSNR. (b) SAM. (c) ERGAS. (d)  $Q2^n$ .

grouped, which implies similar scenes with more classes in the HS image. Fig. 9 plots four quality measures of fused images on University of Pavia data set as a function of  $K$ , which varies from 50 to 600 with step 50. According to the experimental results shown, the performance of the proposed method is relatively high when  $K > 450$ . Thus, we set  $K = 550$  for this data set considering the values of all metrics. There are two parameters  $\lambda_1$  and  $\lambda_2$  in (15). The fusion performance as a function of  $\lambda_1$  and  $\lambda_2$  are plotted in Fig. 10.  $\lambda_1$  is chosen from  $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0\}$ . As can be observed from this figure, it achieves optimal overall performance at  $\lambda_1 = 10^{-2}$ .  $\lambda_2$  is chosen from  $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0\}$ . When  $\lambda_2$  is no bigger than  $10^{-4}$ , the overall performance is relatively high. Thus, we set  $\lambda_1 = 10^{-2}$  and  $\lambda_2 = 10^{-4}$  for University of Pavia data set. Besides, we analyze the parameter settings of the other three data sets. Finally, we set  $K = 450, \lambda_1 = 10^{-2}, \lambda_2 = 10^{-4}$  for Washington DC Mall data set,  $K = 500, \lambda_1 = 10^{-2}, \lambda_2 = 10^{-3}$  for Chikusei data set and  $K = 550, \lambda_1 = 5 \times 10^{-2}, \lambda_2 = 10^{-4}$  for real data set.

## VI. CONCLUSION

A novel NLRSSR is proposed for HS and MS fusion in this paper. For better preservation of spatial and spectral information, the nonlocal low-rank tensor approximation is designed to regularize the HS image. In the prior, MCP penalty is used to formulate the low-rank structure of four-dimensional tensor which is composed of two spatial dimensions, one spectral dimension, and one non-local self-similarity dimension. Meanwhile, dictionary learning and sparse representation are used to describe the sparsity of spectral coefficient. Finally, an optimization algorithm based on ADMM for solving the NLRSSR fusion model is presented. Experiments are performed

on three HS data sets and one real data set, and the fusion results indicate our method is superior to several state-of-art methods. Since the similar cubes clustering leads to high time complexity, we will try to accelerate the algorithm with as little loss of accuracy as possible in the future work.

## REFERENCES

- [1] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5910–5922, Oct 2018.
- [2] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 911–923, Feb 2019.
- [3] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 3–13, Jan 2019.
- [4] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, June 2017.
- [5] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [6] G. Vivone, L. Alparone, J. Chanussot, M. D. Mura, A. Garzelli, G. Licciardi, R. Restaino, and L. Wald, "A critical comparison of pansharpening algorithms," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 191–194.
- [7] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, Oct 2007.
- [8] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3008–3024, June 2015.

- [9] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *CVPR 2011*, June 2011, pp. 2329–2336.
- [10] X. Li, Y. Yuan, and Q. Wang, "Hyperspectral and multispectral image fusion based on band simulation," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 479–483, 2020.
- [11] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, Feb 2012.
- [12] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *European Conference on Computer Vision*. Springer, 2014, pp. 63–78.
- [13] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image super-resolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, June 2015.
- [14] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, Jan 2016.
- [15] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [16] K. Zhang, M. Wang, and S. Yang, "Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1363–1371, March 2017.
- [17] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, July 2015.
- [18] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 3631–3640.
- [19] Q. Wei, N. Dobigeon, and J. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4109–4121, Nov 2015.
- [20] Q. Wei, N. Dobigeon, J. Tourneret, J. Bioucas-Dias, and S. Godsill, "R-FUSE: Robust fast fusion of multiband images based on solving a Sylvester equation," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1632–1636, Nov 2016.
- [21] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 3862–3871.
- [22] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, Aug 2018.
- [23] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1030–1040, April 2018.
- [24] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3034–3047, June 2019.
- [25] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.
- [26] X. Han, B. Shi, and Y. Zheng, "SSF-CNN: Spatial and spectral fusion with cnn for hyperspectral image super-resolution," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 2506–2510.
- [27] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sensing*, vol. 10, no. 5, 2018. [Online]. Available: <https://www.mdpi.com/2072-4292/10/5/800>
- [28] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5345–5355, Nov 2018.
- [29] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [30] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE signal processing magazine*, vol. 19, no. 1, pp. 44–57, 2002.
- [31] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 757–772, Feb 2011.
- [32] D. Tao, S. Maybank, W. Hu, and X. Li, "Stable third-order tensor representation for colour image classification," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*. IEEE, 2005, pp. 641–644.
- [33] S. Aja-Fernández, R. de Luis Garcia, D. Tao, and X. Li, *Tensors in image processing and computer vision*. Springer Science & Business Media, 2009.
- [34] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [35] J. Fan, L. Xue, and H. Zou, "Strong oracle optimality of folded concave penalized estimation," *Annals of statistics*, vol. 42, no. 3, pp. 819–849, 2014.
- [36] W. Cao, Y. Wang, C. Yang, X. Chang, Z. Han, and Z. Xu, "Folded-concave penalization approaches to tensor completion," *Neurocomputing*, vol. 152, pp. 261–273, 2015.
- [37] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 689–696.
- [38] Y. Wang, X. Chen, Z. Han, and S. He, "Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization," *Remote Sensing*, vol. 9, no. 12, 2017. [Online]. Available: <https://www.mdpi.com/2072-4292/9/12/1286>

**Xuelong Li** (M'02-SM'07-F'12) is a full professor with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

**Yue Yuan** received the B.E. degree in software engineering from the Northwestern Polytechnical University, Xi'an, China, in 2018. She is currently working toward the M.S. degree in computer science in the Center for OPTical IMagery Analysis and Learning (OPTIMAL), School of Computer Science, Northwestern Polytechnical University, Xi'an, China. Her research interests include hyperspectral image fusion and computer vision.



**Qi Wang** (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science and with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.