Hyperspectral Image Super-Resolution via Non-Negative Structured Sparse Representation

Weisheng Dong, *Member, IEEE*, Fazuo Fu, Guangming Shi, *Senior Member, IEEE*, Xun Cao, *Member, IEEE*, Jinjian Wu, Guangyu Li, and Xin Li, *Senior Member, IEEE*

***Abstract*— Hyperspectral imaging has many applications from agriculture and astronomy to surveillance and mineralogy. However, it is often challenging to obtain high-resolution (HR) hyperspectral images using existing hyperspectral imaging techniques due to various hardware limitations. In this paper, we propose a new hyperspectral image super-resolution method from a low-resolution (LR) image and a HR reference image of the same scene. The estimation of the HR hyperspectral image is formulated as a joint estimation of the hyperspectral dictionary and the sparse codes based on the prior knowledge of the spatial-spectral sparsity of the hyperspectral image. The hyperspectral dictionary representing prototype reflectance spectra vectors of the scene is first learned from the input LR image. Specifically, an efficient non-negative dictionary learning algorithm using the block-coordinate descent optimization technique is proposed. Then, the sparse codes of the desired HR hyperspectral image with respect to learned hyperspectral basis are estimated from the pair of LR and HR reference images. To improve the accuracy of non-negative sparse coding, a clustering-based structured sparse coding method is proposed to exploit the spatial correlation among the learned sparse codes. The experimental results on both public datasets and real LR hypspectral images suggest that the proposed method substantially outperforms several existing HR hyperspectral image recovery techniques in the literature in terms of both objective quality metrics and computational efficiency.**

Manuscript received May 21, 2015; revised September 29, 2015 and January 31, 2016; accepted March 4, 2016. Date of publication March 22, 2016; date of current version April 7, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61227004, Grant 61390512, Grant 61471281, Grant 61472301, and Grant 61372131, in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2013CB329402, in part by the Research Fund for the Doctoral Program of Higher Education under Grant 20130203130001, in part by the International Cooperation Project of Shaanxi Science and Technology Research and Development Program under Grant 2014KW01-02, and in part by the Shenzhen Oversea High Talent Innovation Fund under Grant KQCX20140521161756231. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue M. Lu.

W. Dong is with the State Key Laboratory on Integrated Services Networks (ISN), School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: wsdong@mail.xidian.edu.cn).

F. Fu, G. Shi, and J. Wu are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: ffzbest2008@163.com; gmshi@xidian.edu.cn; jinjian.wu@mail.xidian.edu.cn).

X. Cao is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: caoxun@nju.edu.cn).

G. Li is with Huawei Technologies Company, Ltd., Xi'an, China (e-mail: gyli4512@163.com).

X. Li is with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506-6109 USA (e-mail: xin.li@ieee.org).

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

Digital Object Identifier 10.1109/TIP.2016.2542360

***Index Terms*— Hyperspectral images, high-resolution reconstruction, non-negative dictionary learning, clustering-based sparse representation.**

I. Introduction

HYPERSPECTRAL imaging is an emerging modality that can simultaneously acquire images of the same scene across a number of different wavelengths. Obtaining dense hyperspectral bands is important to remote sensing [1] and computer vision applications including object segmentation, tracking, and recognitions [2], [4], [5]. While hyperspectral imaging can achieve high spectral resolution, it has severe limitations in spatial resolution when compared against regular RGB (a.k.a. multispectral) cameras in visible spectrum. This is due to the fact that hyperspectral imaging systems need a large number of exposures to simultaneously acquire many bands within a narrow spectral window. To ensure sufficient signal-to-noise ratio, long exposures are often necessary, resulting in the sacrifice of spatial resolution.

While high-resolution (HR) hyperspectral images are desirable in real-world applications, it is often challenging to enhance the spatial resolution of those images due to various hardware limitations. Simply increasing the spatial resolution of image sensors would not be effective for hyperspectral imaging because the average amount of photons reaching the sensors would be further reduced leading to even lower signal-to-noise ratio. Consequently, signal processing based approaches have been proposed for obtaining a HR hyperspectral image by combining a low-resolution (LR) hyperspectral image with a HR panchromatic image (covering a large spectral window) [6], [7]. In [8], a multispectral image is first transformed from the RGB color space to the intensity, hue, and saturation (IHS) domain, and then the intensity channel is replaced by the HR panchromatic image. After resampling the hue and saturation channels, one can obtain the reconstructed HR multispectral image by inverse IHS transformation. While this technique does improve the spatial resolution to some extent, it often introduces spectral distortions in the reconstructed multispectral images. To further improve the reconstruction quality, other fusion methods such as improved linear transformations (e.g., principal component analysis, wavelet transform [9]–[11]), unmixing-based [12]–[14] and joint filtering [16] have been developed in the literature. Those approaches - originally developed by the community of remote sensing - have been known as pansharpening and

especially suitable for the case where the spectral-resolution difference between two input images is relatively small.

The class of sparsity promoting techniques [17], [18] have also been proposed for hyperspectral and multispectral image fusion, showing promising results. In [22], a coupled non-negative matrix factorization (CNMF) approach was proposed to estimate the HR hyperspectral image from a pair of multispectral and hyperspectral images. Since non-negative matrix factorization (NMF) is often not unique [23], [24], the results produced in [22] are not always satisfactory. In [15], Huang et al. proposed a sparse matrix factorization method to fuse remote sensing multispectral images at different spatial and spectral resolutions. In [38] the low-resolution and high-resolution dictionary pairs were constructed to fuse the hyperspectral and multispectral images via joint sparse representations. Based on the assumption that the neighboring pixels of a pixel of interest usually share fractions of the same material, a joint sparse model for spectral unmixing has been proposed in [36] for hyperspectral image resolution enhancement. In [37] the fusion of hyperspectral and multispectral images is formulated as an ill-posed inverse problem and the sparsity of hyperspectral images is exploited via subspace learning in the spectral dimension and sparse coding in the spatial dimensions. An overview of recent state-of-the-art hyperspectral and multispectral image fusion methods can be found in [39].

In addition to the development of hyperspectral and multispectral fusion techniques, hybrid HR hyperspectral imaging systems consisting of a LR hyperspectral camera and a HR RGB camera have also been proposed in [19]–[21]. In [20] a sparse matrix factorization technique was proposed to decompose the LR hyperspectral image into a dictionary of basis vectors and a set of sparse coefficients. The HR hyperspectral image was then reconstructed using the learned basis and sparse coefficients computed from the HR RGB image. Wycoff *et al.* [25] proposed a non-negative sparse matrix factorization method to exploit both sparsity and non-negativity constraints of hyperspectral images. The estimation of HR hyperspectral image from a pair of RGB and hyperspectral images is formulated as a joint optimization problem involving non-negative basis and sparse coefficients, which are solved by the alternative direction multiplier method (ADMM) technique [27]. This line of research has culminated in the work done by Akhtar *et al.* [26] where both non-negativity and spatio-spectral sparsity of the scene are jointly exploited. In their recent work, Akhtar et al. proposed a Bayesian dictionary learning and sparse coding algorithm for hyperspectral image super-resolution that has shown improved performance [41]. Most recently, a coupled matrix factorization approach with non-negative and sparsity constraints has also been proposed for hyperspectral image super-resolution [45].

In this paper, we propose a non-negative structured sparse representation (NSSR) approach to recover a HR hyperspectral image from a LR hyperspectral image and a HR RGB image. The estimation of HR hyperspectral image is formulated as a joint estimation of spectral basis and sparse coefficients with the prior knowledge about spatio-spectral sparsity of the hyperspectral image. The contributions of this paper are two-fold. First, an efficient non-negative dictionary learning algorithm using the block coordinate descent optimization algorithm is proposed, which updates an atom per iteration via a closed-form solution. Second, to improve the accuracy of non-negative sparse coding over the learned basis, a structural sparsity constraint is proposed to exploit the clustering-based sparsity of hyperspectral images - namely reconstructed spectral pixels should be similar to those learned centroids. The performance of our proposed method is verified using both simulated low-resolution hyperspectral images from public datasets and real world low-resolution hyperspectral images captured by a hybrid hyperspectral camera. Experimental results seem to suggest that the proposed method substantially outperforms existing state-of-the-art methods in terms of both objective quality metrics and computational efficiency.

II. Problem Formulation

We aim at recovering a HR hyperspectral image $\boldsymbol{Z} \in \mathbb{R}^{L\times N}$ from a LR hyperspectral image $\boldsymbol{X} \in \mathbb{R}^{L\times n}$ and a HR RGB image $\boldsymbol{Y} \in \mathbb{R}^{3\times N}$ of the same scene where $N = W \times H$ and $n = w \times h$ ($w \ll W$, $h \ll H$) denote the number of pixels in the HR hyperspectral image $\boldsymbol{Z}$ and LR hyperspectral image $\boldsymbol{X}$ respectively; and L is the number of spectral bands of $\boldsymbol{Z}$ ($L \gg 3$). Both $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be expressed as linear combinations of the desired hyperspectral image $\boldsymbol{Z}$:

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{H}, \quad \boldsymbol{Y} = \boldsymbol{P}\boldsymbol{Z}, \tag{1}$$

where $\boldsymbol{H} \in \mathbb{R}^{N\times n}$ denotes the degradation operator of blurring and down-sampling associated with LR image $\boldsymbol{X}$; $\boldsymbol{P}$ a transformation matrix mapping the HR hyperspectral image $\boldsymbol{Z}$ to its RGB representation $\boldsymbol{Y}$. As the number of total measurements from $\boldsymbol{X}$ and $\boldsymbol{Y}$ is much smaller than that of the unknowns, i.e., $(nL + 3N) \ll NL$, the estimation of $\boldsymbol{Z}$ from $\boldsymbol{X}$ and $\boldsymbol{Y}$ is an ill-posed inverse problem. For such ill-posed inverse problems, regularization is a popular tool of exploiting the prior knowledge about the unknown ($\boldsymbol{Z}$ in this case).

Sparsity prior has been shown effective for solving various ill-posed inverse problems under the context of hyperspectral image reconstruction [20], [25], [26], [41], [45]. It assumes that each pixel $\boldsymbol{z}_i \in \mathbb{R}^L$ of the target image $\boldsymbol{Z}$ can be written as the linear combination of a small number of distinct spectral signatures [29], [30] - i.e.,

$$\boldsymbol{z}_i = \boldsymbol{D}\boldsymbol{\alpha}_i + \boldsymbol{e}_i, \tag{2}$$

where $\boldsymbol{D} \in \mathbb{R}_+^{L\times K}$ ($K \geq L$) is the spectral dictionary, $\boldsymbol{\alpha}_i \in \mathbb{R}_+^K$ is the fractional abundance vector assumed to be sparse (i.e., $||\boldsymbol{\alpha}||_0 < T$), and $\boldsymbol{e}_i$ is the approximation error. Each column of $\boldsymbol{D}$ represents a reflectance vector of the underlying material in the scene. As the observed hyperspectral image $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{H}$, each pixel $\boldsymbol{x}_i \in \mathbb{R}^L$ of $\boldsymbol{X}$ can be written as

$$\begin{aligned}\boldsymbol{x}_i &= \sum_{j\in W_i} h_j \boldsymbol{z}_j \\ &= \boldsymbol{D}\sum_{j\in W_i} h_j \boldsymbol{\alpha}_i + \boldsymbol{v}_i = \boldsymbol{D}\boldsymbol{\beta}_i + \boldsymbol{v}_i,\end{aligned} \tag{3}$$

where h_j denotes the weighting coefficients of a window W_i centered at the location i. Here, we assume that at each pixel

location $\boldsymbol{x}_i$ of the LR hyperspectral image $\boldsymbol{X}$, only a few distinct materials can be presented. Therefore, $\boldsymbol{\beta}_i \in \mathbb{R}_+^K$ is a sparse vector. For each pixel $\boldsymbol{y}_i$ of the HR RGB image $\boldsymbol{Y}$, we have

$$\boldsymbol{y}_i = \boldsymbol{P}\boldsymbol{z}_i = \boldsymbol{P}\boldsymbol{D}\boldsymbol{\alpha}_i, \tag{4}$$

from which we can see that for a fixed spectral dictionary $\boldsymbol{D}$ the sparse fractional abundance vectors $\boldsymbol{\alpha}_i$ of the HR hyperspectral image $\boldsymbol{Z}$ can be estimated from the HR RGB image $\boldsymbol{Y}$, as we will elaborate in the next section.

III. Proposed Method

Using the linear mixing model of Eq. (2), we can rewrite the desired HR hyperspectral image $\boldsymbol{Z}$ as

$$\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{A} + \boldsymbol{E}, \tag{5}$$

where $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_N] \in \mathbb{R}_+^{K\times N}$ is the coefficient matrix. However, both spectral dictionary $\boldsymbol{D}$ and coefficients matrix $\boldsymbol{A}$ in Eq. (5) are unknown. The linear mapping relationship between the desired HR hyperspectral image $\boldsymbol{Z}$ and the pair of observed images $\boldsymbol{X}$ and $\boldsymbol{Y}$ motivates us to estimate the spectral dictionary $\boldsymbol{D}$ and the coefficient matrix $\boldsymbol{A}$ simultaneously from the observed pair.

A. Spectral Dictionary Learning

As each pixel in the LR hyperspectral image $\boldsymbol{X}$ can be written as the linear combination of a small number of spectral signatures, we can estimate spectral dictionary $\boldsymbol{D}$ from $\boldsymbol{X}$. In the matrix form, Eq. (3) can be rewritten as

$$\boldsymbol{X} = \boldsymbol{D}\boldsymbol{B} + \boldsymbol{V}, \tag{6}$$

where $\boldsymbol{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_n] \in \mathbb{R}_+^{K\times n}$ is the coefficient matrix and $\boldsymbol{V}$ denotes the approximation error matrix, which is assumed to be additive Gaussian. Both $\boldsymbol{D}$ and $\boldsymbol{B}$ are unknown in Eq. (6). In general, there are infinite possible decompositions of Eq. (6) and it is unlikely that a unique decomposition can be determined. However, for scenes satisfying the sparsity assumption, coefficient vectors $\boldsymbol{\beta}_i$ have to be sparse. It follows that we can solve $\boldsymbol{D}$ and $\boldsymbol{B}$ using sparse non-negative matrix decomposition. Under fairly mild conditions, the sparsest decomposition of Eq. (6) is indeed unique [20]. Therefore spectral dictionary $\boldsymbol{D}$ can be estimated by solving the following sparse non-negative matrix decomposition problem

$$\begin{aligned}(\boldsymbol{D}, \boldsymbol{B}) = \underset{\boldsymbol{D},\boldsymbol{B}}{\operatorname{argmin}}\ & \frac{1}{2}||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{B}||_F^2 + \lambda||\boldsymbol{B}||_1, \\ & s.t.\ \boldsymbol{\beta}_i \geq \boldsymbol{0}, \quad \boldsymbol{d}_k \geq \boldsymbol{0}.\end{aligned} \tag{7}$$

As both sparse codes $\boldsymbol{\beta}_i$ and dictionary $\boldsymbol{D}$ are constrained to be non-negative, existing dictionary learning (DL) algorithms, such as K-SVD algorithm [3] and online dictionary learning (ODL) algorithm [40] cannot be used. In [25], Wycoff et al. adopted ADMM technique to convert a constrained DL problem into an unconstrained version and solved the unconstrained DL problem via alternative optimization. However, in [25] the whole dictionary is updated per iteration, which requires the solution to a large optimization algorithm. This greatly increases the computational complexity. In this paper, inspired by the existing K-SVD [3] and ODL [40] algorithms, we propose a computationally efficient non-negative DL algorithm, which updates each atom per iteration via a closed-form solution. For a fixed $\boldsymbol{D}$, the subproblem with respect to $\boldsymbol{B}$ becomes

$$\boldsymbol{B} = \underset{\boldsymbol{B}}{\operatorname{argmin}} \frac{1}{2}||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{B}||_F^2 + \lambda||\boldsymbol{B}||_1, \quad s.t.\ \boldsymbol{\beta}_i \geq \boldsymbol{0}, \tag{8}$$

which is convex and can be efficiently solved by the iterative shrinkage/thresholding algorithm (ISTA) [42]. However, the converge rate of the ISTA algorithm is known to be slow. For fast convergence rate, we use ADMM technique to solve Eq. (8). To apply ADMM, we reformulate Eq. (8) into

$$\begin{aligned}\boldsymbol{B} = \underset{\boldsymbol{B}}{\operatorname{argmin}}\ & \frac{1}{2}||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}||_F^2 + \lambda||\boldsymbol{B}||_1, \\ & s.t.\ \boldsymbol{B} = \boldsymbol{S}, \quad \boldsymbol{\beta}_i \geq \boldsymbol{0},\end{aligned} \tag{9}$$

Note that we have preserved the non-negative constraint, which is contrary to the ADMM-based ℓ_1-norm non-negative sparse coding in [25]. Applying ADMM [27], we obtain the following augmented Lagrangian function of Eq. (9)

$$\begin{aligned}L_\mu(\boldsymbol{B}, \boldsymbol{S}, \boldsymbol{U}_1) = & \frac{1}{2}||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}||_F^2 + \lambda||\boldsymbol{B}||_1 \\ & + \mu||\boldsymbol{S} - \boldsymbol{B} + \frac{\boldsymbol{U}_1}{2\mu}||_F^2, \quad s.t.\ \boldsymbol{\beta}_i \geq \boldsymbol{0},\end{aligned} \tag{10}$$

where $\boldsymbol{U}_1$ is the Lagrangian multiplier ($\mu > 0$). Then, solving Eq. (9) consists of the following alternative iterations:

$$\begin{aligned}\boldsymbol{S}^{(j+1)} &= \underset{\boldsymbol{S}}{\operatorname{argmin}} L_\mu(\boldsymbol{B}^{(j)}, \boldsymbol{S}, \boldsymbol{U}_1^{(j)}), \\ \boldsymbol{B}^{(j+1)} &= \underset{\boldsymbol{B}}{\operatorname{argmin}} L_\mu(\boldsymbol{B}, \boldsymbol{S}^{(j+1)}, \boldsymbol{U}_1^{(j)}), \quad s.t.\ \boldsymbol{\beta}_i \geq \boldsymbol{0},\end{aligned} \tag{11}$$

where j is the iteration number and Lagrangian multiplier $\boldsymbol{U}_1$ is updated by

$$\boldsymbol{U}_1^{(j+1)} = \boldsymbol{U}_1^{(j)} + \mu(\boldsymbol{S}^{(j+1)} - \boldsymbol{B}^{(j+1)}), \tag{12}$$

Both subproblems in Eq. (11) admit closed-form solutions, namely

$$\begin{aligned}\boldsymbol{S}^{(j+1)} &= (\boldsymbol{D}^\top\boldsymbol{D} + 2\mu\boldsymbol{I})^{-1}(\boldsymbol{D}^\top\boldsymbol{X} + 2\mu(\boldsymbol{B}^{(j)} - \frac{\boldsymbol{U}_1^{(j)}}{2\mu})), \\ \boldsymbol{B}^{(j+1)} &= [\text{Soft}(\boldsymbol{S}^{(j+1)} + \frac{\boldsymbol{U}_1^{(j)}}{2\mu}, \frac{\lambda}{2\mu})]_+,\end{aligned} \tag{13}$$

where Soft$(\cdot)$ denotes a soft-shrinkage operator and $[x]_+ = \max\{x, 0\}$.

For a fixed $\boldsymbol{B}$, $\boldsymbol{D}$ is updated by solving

$$\boldsymbol{D} = \underset{\boldsymbol{D}}{\operatorname{argmin}} ||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{B}||_F^2, \quad s.t.\ \boldsymbol{d}_k \geq \boldsymbol{0}. \tag{14}$$

In this paper, similar to ODL method [40], we propose to solve Eq. (14) by using block coordinate descent [47], i.e., during each iteration we update one column of $\boldsymbol{D}$ while keeping the others fixed under the non-negative constraint. Let $\boldsymbol{D}^{(t)}$ denote the dictionary obtained after the t-th iteration and let $\boldsymbol{d}_k^{(t+1)} = \boldsymbol{d}_k^{(t)} + \Delta\boldsymbol{d}_k$. Then, $\Delta\boldsymbol{d}_k$ can be obtained by solving

$$\begin{aligned}\Delta\boldsymbol{d}_k = \underset{\Delta\boldsymbol{d}_k}{\operatorname{argmin}}\ & ||\boldsymbol{X} - \boldsymbol{D}^{(t)}\boldsymbol{B} - \Delta\boldsymbol{d}_k\boldsymbol{\beta}_k||_F^2, \\ & s.t.\ (\Delta\boldsymbol{d}_k + \boldsymbol{d}_k^{(t)}) \geq \boldsymbol{0},\end{aligned} \tag{15}$$

Algorithm 1 Non-Negative Spectral Dictionary Learning

• **Initialization**:
(a) Input $\boldsymbol{X}$, λ and μ ;
(b) Initialize $\boldsymbol{D}$ via a randomly selection of the columns of $\boldsymbol{X}$ with normalization;
• **Outer loop**: for $t = 1, 2, ..., T_1$ **do**
(a) **Inner loop** (solving Eq. (8)): for $j = 1, 2, \cdots, J$ **do**
(I) Update $\boldsymbol{B}^{(j+1)}$ and $\boldsymbol{S}^{(j+1)}$ via Eq. (13);
(II) Update $\boldsymbol{U}_1^{(j+1)}$ via Eq.(12);
(III) Update $\mu := \rho\mu$ ($\rho > 1$).
(IV) Output $\boldsymbol{B} = \boldsymbol{B}^{(j+1)}$ if $j = J$.
End for
(b) **Inner loop** (solving Eq. (14)): for $k = 1, 2, \cdots, K$ **do**
(I) Update $\boldsymbol{d}_k^{(t+1)}$ via Eq. (17);
(II) Output $\boldsymbol{D}^{(t+1)}$ if $k = K$.
End for
(c) Output $\boldsymbol{D}^{(t+1)}$ if $t = T_1$.
End for

where $\boldsymbol{\beta}_k \in \mathbb{R}^{1\times n}$ denotes the k-th row of the coefficient matrix $\boldsymbol{B}$. Let $\boldsymbol{R}^{(t)} = \boldsymbol{X} - \boldsymbol{D}^{(t)}\boldsymbol{B}$ denote the residual matrix after the t-th iteration. Then, the above objective function can be rewritten as

$$\begin{aligned}
\Delta\boldsymbol{d}_k &= \underset{\Delta\boldsymbol{d}_k}{\operatorname{argmin}} ||\boldsymbol{R}^{(t)} - \Delta\boldsymbol{d}_k\boldsymbol{\beta}_k||_F^2, \quad s.t.\ (\Delta\boldsymbol{d}_k + \boldsymbol{d}_k^{(t)}) \geq \boldsymbol{0}, \\
&= \underset{\Delta\boldsymbol{d}_k}{\operatorname{argmin}} \sum_l^L \sum_j^n (r_{l,i}^{(t)} - \Delta d_{k,l}\beta_{k,i})^2, \\
&\quad s.t.\ (\Delta\boldsymbol{d}_k + \boldsymbol{d}_k^{(t)}) \geq \boldsymbol{0}, \\
&= \underset{\Delta\boldsymbol{d}_k}{\operatorname{argmin}} \sum_l^L (\Delta d_{k,l} - \frac{\boldsymbol{\beta}_k(\boldsymbol{r}_l^{(t)})^\top}{\sum_i^n \beta_{k,i}^2})^2 + C, \\
&\quad s.t.\ (\Delta\boldsymbol{d}_k + \boldsymbol{d}_k^{(t)}) \geq \boldsymbol{0}, \\
&= \underset{\Delta\boldsymbol{d}_k}{\operatorname{argmin}} ||\Delta\boldsymbol{d}_k - \frac{\boldsymbol{R}^{(t)}\boldsymbol{\beta}_k^\top}{\sum_i^n \beta_{k,i}^2}||_2^2 + C, \\
&\quad s.t.\ (\Delta\boldsymbol{d}_k + \boldsymbol{d}_k^{(t)}) \geq \boldsymbol{0}, \qquad (16)
\end{aligned}$$

where $\Delta d_{k,l}$ denotes the l-th element of $\Delta\boldsymbol{d}_k$, $\boldsymbol{r}_l^{(t)} \in \mathbb{R}^{1\times n}$ denotes l-th row of the residual matrix $\boldsymbol{R}^{(t)}$, and C denotes the constant independent of $\Delta\boldsymbol{d}_k$. Based on Eq. (16), it is easy to show that the solution to Eq. (15) is given by

$$\boldsymbol{d}_k^{(t+1)} = [\boldsymbol{d}_k^{(t)} + \frac{\boldsymbol{R}^{(t)}\boldsymbol{\beta}_k^\top}{b_k}]_+, \qquad (17)$$

where $b_k = \sum_i^n \beta_{k,i}^2$.

The overall algorithm for non-negative dictionary learning is summarized below in **Algorithm 1**.

B. Sparse Codes Estimation via Non-Negative Structured Sparse Coding

Once the spectral dictionary $\boldsymbol{D}$ is estimated, sparse codes $\boldsymbol{\alpha}_i$ for each pixel $\boldsymbol{z}_i$ of the desired HR hyperspectral image $\boldsymbol{Z}$ can be estimated and then $\hat{\boldsymbol{z}}_i$ can be reconstructed as $\boldsymbol{z}_i = \boldsymbol{D}\boldsymbol{\alpha}_i$. Since both observed $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be expressed as linear combination of the desired $\boldsymbol{Z}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ can then be expressed as

$$\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{A} + \boldsymbol{W} = \tilde{\boldsymbol{D}}\boldsymbol{A} + \boldsymbol{W}_1, \quad \boldsymbol{X} = \boldsymbol{D}\boldsymbol{A}\boldsymbol{H} + \boldsymbol{W}_2, \quad (18)$$

where $\tilde{\boldsymbol{D}} = \boldsymbol{P}\boldsymbol{D}$ denotes the transformed spectral dictionary, and $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ denote the approximation error matrix. From the above equation, we see that sparse codes $\boldsymbol{\alpha}_i$ can be estimated from the HR RGB image $\boldsymbol{Y}$ and the low-resolution hyperspectral image $\boldsymbol{X}$. With the sparsity constraint, sparse coefficient matrix $\boldsymbol{A}$ can be estimated by solving the following non-negative sparse coding problem

$$\begin{aligned}
\boldsymbol{A} = \underset{\boldsymbol{A}}{\operatorname{argmin}}\ & ||\boldsymbol{Y} - \tilde{\boldsymbol{D}}\boldsymbol{A}||_F^2 + ||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{A}\boldsymbol{H}||_F^2 + \eta||\boldsymbol{A}||_1, \\
& s.t.\ \boldsymbol{\alpha}_i \geq \boldsymbol{0}. \qquad (19)
\end{aligned}$$

In Eq. (19), sparse codes of each pixel are estimated independently. However, pixels in the HR hyperspectral image have strong spatial correlations with their local and nonlocal similar neighbors (similar pixels are likely to represent similar spectral signatures). The ℓ_1-norm non-negative sparse model of Eq. (19) cannot exploit the spatial correlations among local and nonlocal similar pixels. To address this issue, we propose the following clustering-based non-negative structured sparse representation (NSSR) model

$$\begin{aligned}
\boldsymbol{A} = \underset{\boldsymbol{A}}{\operatorname{argmin}}\ & ||\boldsymbol{Y} - \tilde{\boldsymbol{D}}\boldsymbol{A}||_F^2 + ||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{A}\boldsymbol{H}||_F^2 \\
& + \eta_1 \sum_{q=1}^{Q} \sum_{i\in S_q} ||\boldsymbol{D}\boldsymbol{\alpha}_i - \boldsymbol{\mu}_q||_2^2 + \eta_2||\boldsymbol{A}||_1, \quad s.t.\ \boldsymbol{\alpha}_i \geq \boldsymbol{0}, \qquad (20)
\end{aligned}$$

where $\boldsymbol{\mu}_q$ denotes the centroid of the q-th cluster of the reconstructed spectral pixel $\boldsymbol{z}_i$. In addition to the ℓ_1 sparsity regularization, the proposed NSSR model also exploits a *structural* prior that the reconstructed spectral pixels should be similar to those learned centroids. In our previous work [28], such clustering-based sparse representation (CSR) method have been shown effective for the task of image denoising. Conceptually, the CSR is similar to the block-matching and 3D filtering (BM3D) algorithm [34], which first groups similar image patches and then sparsify the groups via 3D-transformation (2D-DCT plus 1D-wavelet). Different from BM3D, we exploit the structural correlations among the similar patches by unifying dictionary learning and clustering-based regularization into a variational framework. In this paper, we adopt the CSR formulation but introduce the following extensions for the estimation of fractional abundance vectors.

First, the non-negative sparse decomposition of hyperspectral image is exploited - note that material coefficients corresponding to the surface albedo in the physical world can only take non-negative values. Second, instead of using a set of orthogonal PCA basis as in [28], a non-orthogonal dictionary is learned; such relaxation is beneficial to improve the accuracy of sparse reconstruction. Third, instead of using the ℓ_1-norm, we adopt the ℓ_2-norm to exploit the above-mentioned structural prior. The centroid vector $\boldsymbol{\mu}_q$ of the

Algorithm 2 NSSR-Based HR Hyperspectral Image Super-Resolution

- **Initialization**:
  (a) Learn the spectral dictionary $\boldsymbol{D}$ from $\boldsymbol{X}$ using **Algorithm 1**;
  (b) Obtain the sample sets S_q (through k-NN search) using the HR color image $\boldsymbol{Y}$;
  (c) Set parameters η_1 and η_2, and initialize $\boldsymbol{U} = \boldsymbol{0}$;
- **Solving Eq. (24) via ADMM**: for $t = 0, 1, ..., T_2$ **do**
  (a) Compute $\boldsymbol{A}^{(t+1)}$, $\boldsymbol{Z}^{(t+1)}$ and $\boldsymbol{S}^{(t+1)}$ via Eq. (28);
  (b) Update the Lagrangian multipliers $\boldsymbol{V}_1^{(t+1)}$ and $\boldsymbol{V}_2^{(t+1)}$ via Eq. (27);
  (c) Update $\mu := \rho\mu$ $(\rho > 1)$;
  (d) Update $\boldsymbol{U}$ using $\boldsymbol{A}^{(t+1)}$ and Eq. (22);
  (e) Output $\boldsymbol{Z}^{(t+1)}$ if $t = T_2 - 1$.

End for

q-th cluster in Eq. (20) is then computed as

$$\boldsymbol{\mu}_q = \sum_{i \in S_q} w_i (\boldsymbol{D}\boldsymbol{\alpha}_i), \tag{21}$$

where $w_i = \frac{1}{c}\exp(-||\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{y}}_q)||_2^2/h$ is the weighting coefficients based on the similarity of RGB image patches, c is the normalization constant, $\tilde{\boldsymbol{y}}_i$ and $\tilde{\boldsymbol{y}}_q$ denote the RGB image patches centered at positions i and q respectively. In practice, fractional abundance vectors $\boldsymbol{\alpha}_i$ are unknown, and we cannot compute $\boldsymbol{\mu}_q$ directly using Eq. (21). We overcome this difficulty by iteratively estimating $\boldsymbol{\mu}_q$ from the current estimates of $\boldsymbol{\alpha}_i$. Let $\boldsymbol{\alpha}_i = \hat{\boldsymbol{\alpha}}_i + \boldsymbol{e}_i$, wherein $\boldsymbol{e}_i$ denotes estimation errors observing a Gaussian distribution with zero-mean. Then, Eq. (21) can be rewritten into

$$\boldsymbol{\mu}_q = \sum_{i \in S_q} w_i (\boldsymbol{D}\hat{\boldsymbol{\alpha}}_i) + \sum_{i \in S_q} w_i (\boldsymbol{D}\boldsymbol{e}_i) = \hat{\boldsymbol{\mu}}_q + \boldsymbol{n}_q, \tag{22}$$

where $\boldsymbol{n}_q$ denotes the estimation error of $\boldsymbol{\mu}_q$. As $\boldsymbol{e}_i$ is assumed to be zero-mean and Gaussian, $\boldsymbol{n}_q$ would be small. Therefore, $\boldsymbol{\mu}_q$ can be readily estimated from the current estimate of $\boldsymbol{\alpha}_i$. In our implementation, we recursively compute $\boldsymbol{\mu}_q$ using the previous estimate of $\boldsymbol{\alpha}_i$ after each iteration. In our implementation, we use the k-Nearest Neighbor (k-NN) clustering method to group similar spectral pixels for each spectral pixel. Due to the structural similarity between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, we perform k-NN clustering on the HR RGB image patches to search for similar neighbors of $\boldsymbol{\alpha}_q$, i.e.,

$$S_q = \{i \mid ||\tilde{\boldsymbol{y}}_q - \tilde{\boldsymbol{y}}_i|| < T\}. \tag{23}$$

Alternatively, we can also form the cluster S_q by selecting the $m = 20$ patches that are closest to the exemplar one $\tilde{\boldsymbol{y}}_q$.

After estimating $\boldsymbol{\mu}_q$ for each $\boldsymbol{\alpha}_i$, Eq. (20) can be rewritten as

$$\begin{aligned}\boldsymbol{A} = \underset{\boldsymbol{A}}{\operatorname{argmin}} &||\boldsymbol{Y} - \tilde{\boldsymbol{D}}\boldsymbol{A}||_F^2 + ||\boldsymbol{X} - \boldsymbol{D}\boldsymbol{A}\boldsymbol{H}||_F^2 \\ &+ \eta_1||\boldsymbol{D}\boldsymbol{A} - \boldsymbol{U}||_2^2 + \eta_2||\boldsymbol{A}||_1 \quad s.t.\ \boldsymbol{\alpha}_i \geq \boldsymbol{0},\end{aligned} \tag{24}$$

where $\boldsymbol{U} = [\hat{\boldsymbol{\mu}}_1, \cdots, \hat{\boldsymbol{\mu}}_N]$. Similar to Eq. (8), Eq. (24) is convex and can be solved by an ISTA algorithm. However, the convergence rate of ISTA algorithms is known to be slow. For fast convergence, we use ADMM technique to solve Eq. (24) instead. More specifically, we obtain the following augmented Lagrangian function:

$$\begin{aligned}&L_\mu(\boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{V}_1, \boldsymbol{V}_2) \\ &= ||\boldsymbol{Y} - \tilde{\boldsymbol{D}}\boldsymbol{S}||_F^2 + ||\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{H}||_F^2 \\ &\quad + \eta_1||\boldsymbol{D}\boldsymbol{S} - \boldsymbol{U}||_2^2 + \eta_2||\boldsymbol{A}||_1 + \mu||\boldsymbol{D}\boldsymbol{S} - \boldsymbol{Z} + \frac{\boldsymbol{V}_1}{2\mu}||_F^2 \\ &\quad + \mu||\boldsymbol{S} - \boldsymbol{A} + \frac{\boldsymbol{V}_2}{2\mu}||_F^2, \quad s.t.\ \boldsymbol{\alpha}_i \geq \boldsymbol{0},\end{aligned} \tag{25}$$

where $\boldsymbol{V}_1, \boldsymbol{V}_2$ are Lagrangian multipliers ($\mu > 0$). Note that the non-negative constraint is also preserved in the above augmented Lagrangian function. Minimizing the augmented Lagrangian function leads to the following iterations:

$$\begin{aligned}\boldsymbol{A}^{(t+1)} &= \underset{\boldsymbol{A}}{\operatorname{argmin}}\, L_\mu(\boldsymbol{A}, \boldsymbol{Z}^{(t)}, \boldsymbol{S}^{(t)}, \boldsymbol{V}_1^{(t)}, \boldsymbol{V}_2^{(t)}) \\ \boldsymbol{Z}^{(t+1)} &= \underset{\boldsymbol{Z}}{\operatorname{argmin}}\, L_\mu(\boldsymbol{A}^{(t+1)}, \boldsymbol{Z}, \boldsymbol{S}^{(t)}, \boldsymbol{V}_1^{(t)}, \boldsymbol{V}_2^{(t)}) \\ \boldsymbol{S}^{(t+1)} &= \underset{\boldsymbol{S}}{\operatorname{argmin}}\, L_\mu(\boldsymbol{A}^{(t+1)}, \boldsymbol{Z}^{(t+1)}, \boldsymbol{S}, \boldsymbol{V}_1^{(t)}, \boldsymbol{V}_2^{(t)}),\end{aligned} \tag{26}$$

where the Lagrangian multipliers are updated by

$$\begin{aligned}\boldsymbol{V}_1^{(t+1)} &= \boldsymbol{V}_t^{(t)} + \mu(\boldsymbol{D}\boldsymbol{S}^{(t+1)} - \boldsymbol{Z}^{(t+1)}), \\ \boldsymbol{V}_2^{(t+1)} &= \boldsymbol{V}_2^{(t)} + \mu(\boldsymbol{S}^{(t+1)} - \boldsymbol{A}^{(t+1)}).\end{aligned} \tag{27}$$

All sub-problems in Eq. (26) can be solved analytically, i.e.

$$\begin{aligned}\boldsymbol{A} &= [\text{Soft}(\boldsymbol{S}^{(t)} + \frac{\boldsymbol{V}_2^{(t)}}{2\mu}, \frac{\eta_2}{2\mu})]_+, \\ \boldsymbol{Z} &= [\boldsymbol{X}\boldsymbol{H}^\top + \mu(\boldsymbol{D}\boldsymbol{S}^{(t)} + \frac{\boldsymbol{V}_1}{2\mu})](\boldsymbol{H}\boldsymbol{H}^\top + \mu\boldsymbol{I})^{-1}, \\ \boldsymbol{S} &= [\tilde{\boldsymbol{D}}^\top\tilde{\boldsymbol{D}} + (\eta_1 + \mu)\boldsymbol{D}^\top\boldsymbol{D} + \mu\boldsymbol{I}]^{-1}[\tilde{\boldsymbol{D}}^\top\boldsymbol{Y} + \eta_1\boldsymbol{D}^\top\boldsymbol{U} \\ &\quad + \mu\boldsymbol{D}^\top(\boldsymbol{Z}^{(t)} + \frac{\boldsymbol{V}_1^{(t)}}{2\mu}) + \mu(\boldsymbol{A}^{(t)} + \frac{\boldsymbol{V}_2^{(t)}}{2\mu})],\end{aligned} \tag{28}$$

As the matrix to be inverted in the equation of updating $\boldsymbol{Z}$ are large, we use conjugate gradient algorithm to compute the matrix inverse. The overall algorithm for estimating the HR hyperspectral image is summarized below in **Algorithm 2**.

IV. Experimental Results

To verify the performance of our proposed method, we have conducted extensive experiments on both simulated LR hyperspectral images and real-world LR hyperspectral images. The basic parameters of the proposed NSSR

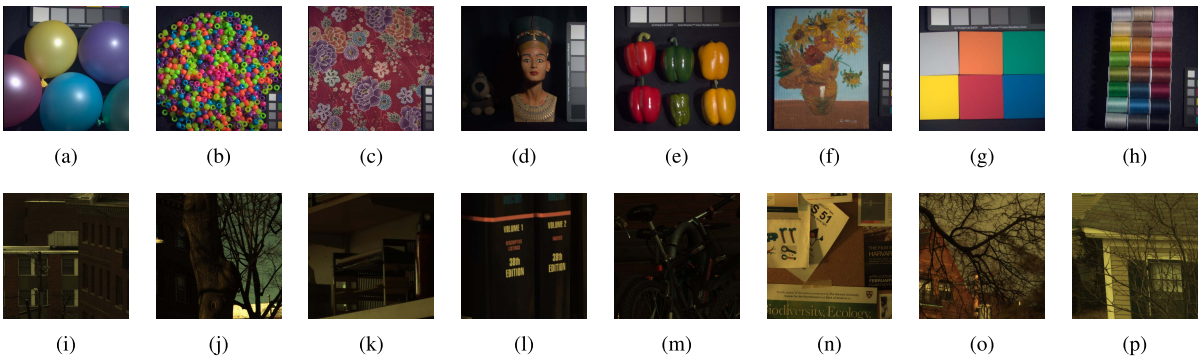


Fig. 1. The HR RGB images from the CAVE (the first row) [31] and Harvard (the second row) [32] datasets. (a) *Balloons*. (b) *Beads*. (c) *Cloth*. (d) *Statue*. (e) *Peppers*. (f) *Painting*. (g) *Sponges*. (h) *Spools*. (i) *Image 1*. (j) *Image 2*. (k) *Image 3*. (l) *Image 4*. (m) *Image 5*. (n) *Image 6*. (o) *Image 7*. (p) *Image 8*.

TABLE I

THE AVERAGE AND STANDARD DEVIATION OF PSNR, RMSE, SAM, AND ERGAS RESULTS OF THE TEST METHODS FOR DIFFERENT SCALING FACTORS ON THE CAVE DATASET [31]

Method	GSOMP	MF[20]	SNNMF[25]	CNMF[22]	BSR [37]	**Proposed NSSR**
s=8						
PSNR	33.64±**3.24**	41.83±4.09	43.53±3.83	43.11±4.13	44.15±3.79	**45.91**±3.66
RMSE	5.69±2.22	2.34±1.38	1.89±0.94	2.01±1.03	1.75±0.88	**1.42±0.67**
SAM	11.86±5.43	3.88±1.29	3.42±1.00	3.45±1.02	3.31±1.03	**2.84±0.87**
ERGAS	2.99±1.26	1.26±0.66	1.03±0.54	1.07±0.57	0.97±0.51	**0.81±0.40**
s=16						
PSNR	32.96±**3.02**	40.43±3.96	41.21±3.76	41.40±3.76	41.57±3.76	**44.17**±3.03
RMSE	6.08±2.06	2.71±1.41	2.45±1.18	2.39±1.13	2.36±1.18	**1.75±0.88**
SAM	12.60±4.34	4.82±1.67	4.61±1.64	4.31±1.23	4.57±1.55	**3.43±1.02**
ERGAS	1.43±0.69	0.73±0.41	0.66±0.35	0.63±0.35	0.58±0.31	**0.48±0.27**
s=32						
PSNR	32.48±**3.08**	39.37±3.76	38.73±3.79	39.53±3.55	39.16±3.91	**42.26**±4.01
RMSE	6.47±2.53	3.03±1.44	3.26±1.57	2.93±1.30	3.13±1.57	**2.21±1.19**
SAM	14.19±5.42	6.12±2.17	6.50±2.32	5.48±1.62	6.75±2.37	**4.33±1.37**
ERGAS	0.77±0.32	0.40±0.22	0.44±0.23	0.39±0.21	0.37±0.22	**0.30±0.18**

method are set as follows: the number of atoms in dictionary $\boldsymbol{D}$ is $K = 80$; the maximal iteration numbers in **Algorithm 1** and **2** are $T = 10$, $J = 70$ and $T_2 = 25$, $\eta_1 = 0.015$ and $\eta_2 = 0.1 \times 10^{-3}$. We have compared the proposed method with several leading hyperspectral image super-resolution methods, including Matrix Factorization method (MF) method [20], coupled non-negative matrix factorization (CNMF) method [22], Sparse Non-negative Matrix Factorization (SNNMF) method [25], Generalization of Simultaneous Orthogonal Matching Pursuit (G-SOMP+) method [26], and Bayesian sparse representation (BSR) method [37].[1] The source code accompanying this paper can be downloaded from the following website: http://see.xidian.edu.cn/faculty/wsdong/HSI_SR_project.htm.

A. Experiments on Simulated LR Hyperspectral Images

Two different public datasets of hyperspectral images, i.e., the CAVE [31] and the Harvard [32] datasets are used to assess the performance of our proposed method. The CAVE dataset consists of 512 × 512 hyperspectral images of everyday objects, which are captured using 31 spectral bands ranging from 400nm to 700nm at an interval of 10nm. The Harvard dataset contains hyperspectral images of real-world indoor and outdoor scenes, which are acquired using 31 spectral bands ranging from 420nm to 720nm with an incremental of 10nm. Some test images used in this paper from the two datasets are shown in Fig. 1. The hyperspectral images from the two datasets served as ground-truth images are used to generate simulated LR hyperspectral images and HR RGB images. As in [20], [25], and [26], the original HR hyperspectral images $\boldsymbol{Z}$ are downsampled by averaging over disjoint $s \times s$ blocks to simulate the LR hyperspectral images $\boldsymbol{X}$, where s is the scaling factor (e.g., $s = 8$, 16, 32). Similar to [25] and [26], HR RGB images $\boldsymbol{Y}$ are generated by downsampling the hyperspectral images $\boldsymbol{Z}$ along the spectral dimension using the spectral transform matrix $\boldsymbol{F}$ derived from the response of a Nikon D700 camera.[2] To evaluate the quality of reconstructed hyperspectral images, four objective quality metrics - namely peak-signal-to-noise ration (PSNR), root-mean-square error (RMSE), relative dimensionless global error in synthesis (ERGAS) [46], and spectral angle mapper (SAM) [37], [43] - are used in our study.

[1] We thank the authors of [20], [22], [25], [26], and [37] for providing their codes.

[2] Available at: https://www.maxmax.com/spectral_response.htm.

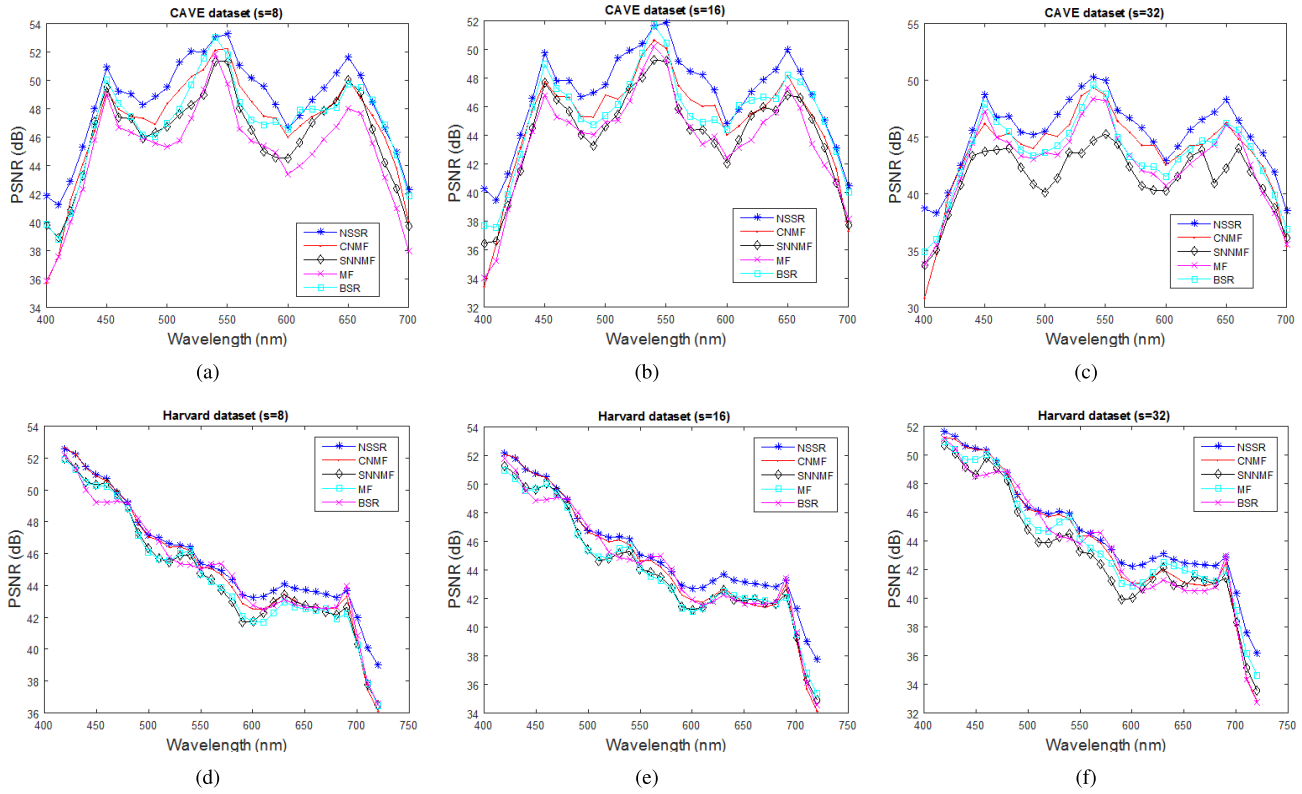


Fig. 2. The average PSNR curves as functions of the wavelengths of the spectral bands for the test methods on (a)-(c) CAVE dataset with scaling factors 8, 16, and 32, respectively; (d)-(f) Harvard dataset with scaling factors 8, 16, and 32, respectively.

The average and standard deviations of the PSNR, RMSE, ERGAS and SAM results of competing methods for different scaling factors on the CAVE dataset are reported in Table I.[3] From Table I, it can be seen that the G-SOMP+ method [26] performs worse than other methods. This may be due to the fact that it doesn't exploit the observation constraint $\boldsymbol{X} = \boldsymbol{ZH}$, and thus doesn't require the downsampling matrix $\boldsymbol{H}$ as a prior knowledge, which is generally unknown and has to be estimated in practical applications. Clearly, the proposed **NSSR** method outperforms all other competing methods. The proposed **NSSR** method performs much better than the SNNMF method [25], which only exploits the sparsity prior. On average, the PSNR and RMSE values of the proposed **NSSR** method for scaling factors $s = 32$ and 16 are comparable or even better than those of other methods [20], [22], [25] for scaling factors $s = 16$ and 8, respectively. Fig. 2 (a)-(c) shows the average PSNR curves as functions of the wavelengths of the spectral bands over the CAVE dataset for the test methods. It can be seen that the proposed **NSSR** method consistently outperforms other methods at each spectral bands for all scaling factors. In Fig. 3, we show the reconstructed HR hyperspectral images at 480nm, 550nm and 650nm by the competing methods for test image *Pepper* of the CAVE dataset. From Fig. 3, we can see that all the test methods can well reconstruct the HR spatial structures of the hyperspectral images. Obviously, the proposed **NSSR** method performs best in recovering the details of the original hyperspectral images.

The average RMSE and PSNR results of the recovered HR hyperspectral images of the Harvard dataset [32] are shown in Table II. It can be observed that the proposed **NSSR** method also outperforms other competing methods. Fig. 2 (d)-(f) shows the average PSNR curves as functions of the wavelengths of the spectral bands over Harvard dataset for the test methods. It can be seen that all the test methods can well reconstruction the spectral bands when the wavelengths are shorter than 500nm, and the performance of the proposed method and the CNMF method [22] are comparable for these short spectral bands. This is because that the pixel values of the spectral bands of the wavelengths shorter than 500nm are small and smooth. Hence, all the methods can well reconstruct these spectral bands. It can be seen that the PSNR gains of the proposed method over other methods increase for the spectral bands corresponding to longer wavelengths. For visual comparison, parts of the reconstructed spectral images at 620nm, 640nm and 680nm by the test methods are shown in Fig. 4. We can seen that all the competing methods can well recover the HR spatial structures of the scene, but the proposed method achieves the smallest reconstruction errors.

In the above experiments, the uniform blur kernel of size $s \times s$ is applied to $\boldsymbol{Z}$ before downsampling. In practice, the optics blur may be generated, which can be modeled by a Gaussian function. To verify the robustness of the proposed method to Gaussian blur, we also simulate $\boldsymbol{X}$ by first applying

[3]For detailed quality metrics, please refer to the project website: http://see.xidian.edu.cn/faculty/wsdong/HSI_SR_Project.htm.

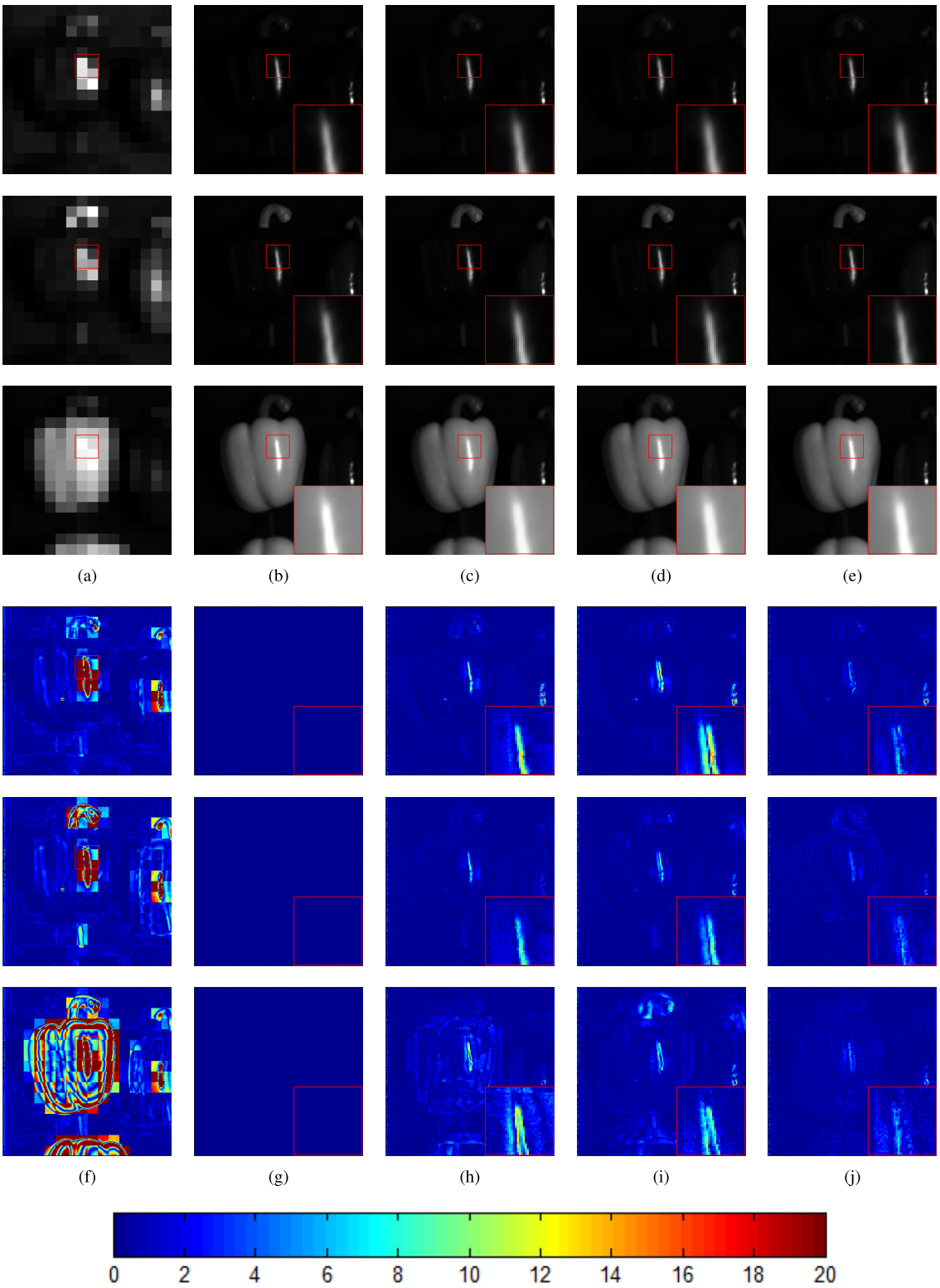


Fig. 3. Reconstructed images of *Pepper* in CAVE dataset at 480nm, 550nm and 650nm with uniform blur kernel and scaling factor $s = 16$. The first 3 rows show the reconstructed images for 580nm, 620nm and 680nm bands, respectively; the last 3 rows show the error images of the competing methods. (a) the LR images $\boldsymbol{X}$; (b) the original images $\boldsymbol{Z}$; (c) the CNMF method [22] (PSNR = 39.52 dB, RMSE = 2.69, ERGAS = 0.95, SAM = 4.61); (d) the BSR method [37] (PSNR = 44.30 dB, RMSE = 1.55, ERGAS = 0.67, SAM = 4.04); (e) the **proposed NSSR** method (PSNR = **46.84** dB, RMSE = **1.16**, ERGAS = **0.47**, SAM = **3.08**). (f) Error of $\boldsymbol{X}$. (g) Error of $\boldsymbol{Z}$. (h) CNMF [22]. (i) BSR [37]. (j) **Proposed NSSR**.

a 8×8 Gaussian blur function of standard deviation 3 to $\boldsymbol{Z}$ before downsampling along both horizontal and vertical directions with scaling factor 8. Table III shows the quality metric values of the test methods. We can see that proposed **NSSR** method still outperforms other competing methods. Fig. 5 shows the parts of the reconstructed *Cloth* images at

TABLE II

THE AVERAGE AND STANDARD DEVIATION OF PSNR, RMSE, SAM, AND ERGAS RESULTS OF THE TEST METHODS FOR DIFFERENT SCALING FACTORS ON THE HARVARD DATASET [32]

Method	GSOMP	MF[20]	SNNMF[25]	CNMF[22]	BSR [37]	**Proposed NSSR**
s=8						
PSNR	38.89±5.94	43.74±3.79	43.86±3.67	44.35±3.81	44.51±4.07	**45.03±3.57**
RMSE	3.79±3.39	1.83±0.90	1.79±0.78	1.70±0.76	1.71±0.96	**1.56±0.67**
SAM	4.00±2.14	2.66±0.92	2.63±0.87	2.51±0.79	2.48±0.85	**2.37±0.71**
ERGAS	1.65±1.15	0.87±0.45	0.85±0.44	0.78±0.32	0.84±0.47	**0.76±0.31**
s=16						
PSNR	38.56±5.72	43.30±3.99	43.31±3.87	43.56±4.26	43.80±4.37	**44.51±3.59**
RMSE	3.83±3.10	1.94±0.95	1.93±0.90	1.93±1.18	1.91±1.33	**1.65±0.72**
SAM	4.16±2.23	2.85±1.02	2.82±1.01	2.74±1.05	2.69±1.09	**2.48±0.78**
ERGAS	0.77±0.43	0.47±0.29	0.45±0.24	0.42±**0.18**	0.45±0.290	**0.41**±0.20
s=32						
PSNR	38.02±5.71	43.19±3.87	42.03±**3.61**	43.00±4.44	43.11±4.59	**44.00**±3.63
RMSE	4.08±3.55	1.96±0.97	2.20±0.94	2.08±1.34	2.10±1.60	**1.76±0.79**
SAM	4.79±2.99	2.93±1.06	3.17±1.07	2.91±1.18	2.93±1.33	**2.64±0.86**
ERGAS	0.41±0.24	0.23±0.14	0.26±0.27	0.23±**0.11**	0.24±0.15	**0.21**±0.12

TABLE III

THE AVERAGE AND STANDARD DEVIATION OF PSNR, RMSE, SAM AND ERGAS RESULTS OF THE TEST METHODS ON THE CAVE [31] AND HARVARD DATASETS [32] (GAUSSIAN BLUR KERNEL, SCALING FACTOR 8)

Method	GSOMP[26]	MF[20]	SNNMF[25]	CNMF[22]	BSR[37]	**Proposed NSSR**
CAVE dataset						
PSNR	33.69±**3.07**	36.35±3.73	36.13±3.10	40.39±3.17	43.96±3.76	**45.41**±3.55
RMSE	5.62±2.11	4.31±2.37	4.24±1.55	2.61±1.06	1.79±0.88	**1.50±0.68**
SAM	11.99±5.53	11.04±3.54	14.01±5.01	8.01±2.83	3.33±1.04	**2.95±0.84**
ERGAS	2.96±1.36	2.19±0.95	2.21±0.92	1.37±0.68	0.99±0.51	**0.82±0.41**
Harvard dataset						
PSNR	38.77±5.41	40.40±4.98	40.48±4.85	43.42±4.72	44.55±4.00	**44.95±3.55**
RMSE	3.60±2.53	2.91±1.92	2.86±2.00	2.05±1.61	1.69±0.90	**1.57±0.67**
SAM	3.92±1.98	4.03±1.99	4.14±2.24	2.83±1.22	2.47±0.83	**2.38±0.73**
ERGAS	1.56±0.93	1.23±0.78	1.17±0.54	0.84±0.39	0.83±0.47	**0.76±0.32**

TABLE IV

PERFORMANCE OF DIFFERENT HSI SUPER-RESOLUTION METHODS ON *Pavia* IMAGE WITH SCALING FACTOR 4. RMSE (IN 10^{-2}), SAM (IN DEGREES) AND ERGAS

Method	CNMF[22]	Wavelet MAP[43]	HMC[44]	BSR[37]	**Proposed NSSR**
RMSE	1.119	1.099	1.011	0.947	**0.936**
SAM	2.039	1.849	1.653	1.492	**1.474**
ERGAS	1.089	0.994	0.911	0.850	**0.833**

450nm, 550nm and 650 nm by the test methods. It can be seen that the proposed **NSSR** method outperforms the CNMF [22] and BSR [37] methods in recovering fine image details.

We have also conducted an experiment using the remotely sensed hyperspectral image. The *Pavia* hyperspectral image containing infrared bands used in [37] is adopted. For a fair comparison, the experimental setting on the *Pavia* test image is same as that of [37]. The LR hyperspectral image $\boldsymbol{X}$ is simulated by first applying a 5×5 Gaussian kernel with standard deviation 2.5 and then downsampling along both horizontal and vertical directions with scaling factor 4. The Gaussian white noise is added to the LR hyperspectral image such that the SNR is 35 dB for the first 43 bands and 30 dB for the remaining 50 bands. The IKONOS-like reflectance spectral response filter is used to generate the HR multispectral image $\boldsymbol{Y}$. Table IV shows the performance of different test methods on the *Pavia* test image. For fair comparison, the results of other competing methods are directly obtained from [37]. It can be seen that the both the Bayesian sparse representation (BSR) method of [37] and the proposed **NSSR** method are very effective in reconstructing the HR hyperspectral image. The proposed **NSSR** method slightly outperforms the current state-of-the-art BSR method [37]. Parts of the reconstructed HR HSI images are shown in Fig. 6.

B. Experiments on Real-World LR Hyperspectral Images

We have also evaluated the performance of the proposed method on real-world data captured by hybrid spectral and

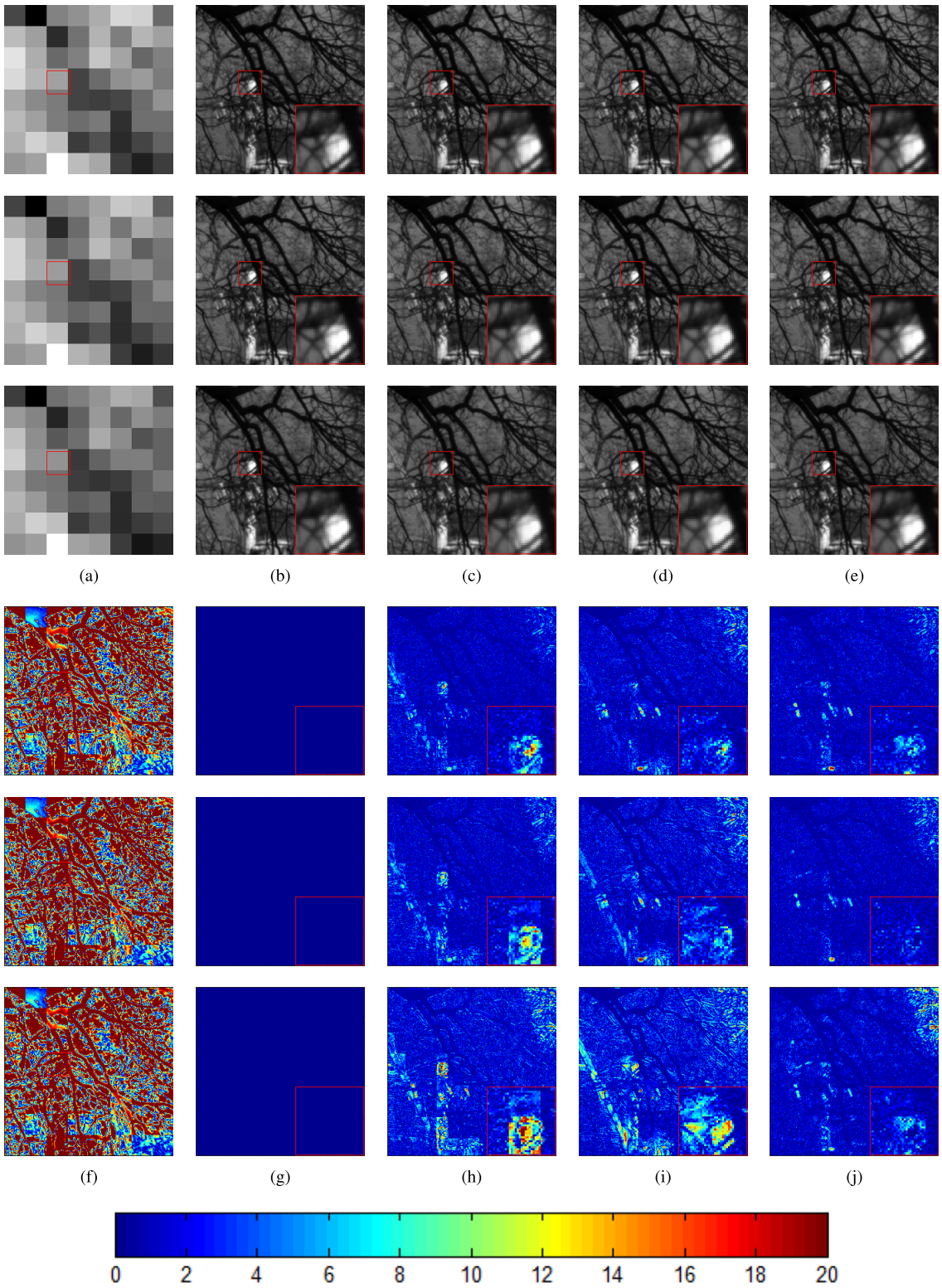


Fig. 4. Reconstructed images of *Image 8* at 620nm, 640nm and 680nm with uniform blur kernel and scaling factor $s = 32$. The first 3 rows show the reconstructed images for 460nm, 540nm and 620nm bands, respectively; the last 3 rows show the error images of the competing methods. (a) the LR images $\boldsymbol{X}$; (b) the original images $\boldsymbol{Z}$; (c) the CNMF method [22] (PSNR = 36.88 dB, RMSE=3.65, ERGAS = 0.18, SAM = 3.41); (d) the BSR method [37] (PSNR = 35.41 dB, RMSE = 4.32, ERGAS = 0.21, SAM = 3.66); (e) the **proposed NSSR** method (PSNR = **41.45** dB, RMSE = **2.16**, ERGAS = **0.13**, SAM = **2.15**). (f) Error of $\boldsymbol{X}$. (g) Error of $\boldsymbol{Z}$. (h) CNMF [22]. (i) BSR [37]. (j) **Proposed NSSR**

RGB cameras as developed in [21]. The spectral camera captures dense spectra information but with significant loss of spatial resolution, while the RGB camera captures the information of the same scene with high spatial resolution. The spectral data captured with the hybrid spectral camera of [21] are shown in Fig. 7 (the second column), where each spectral

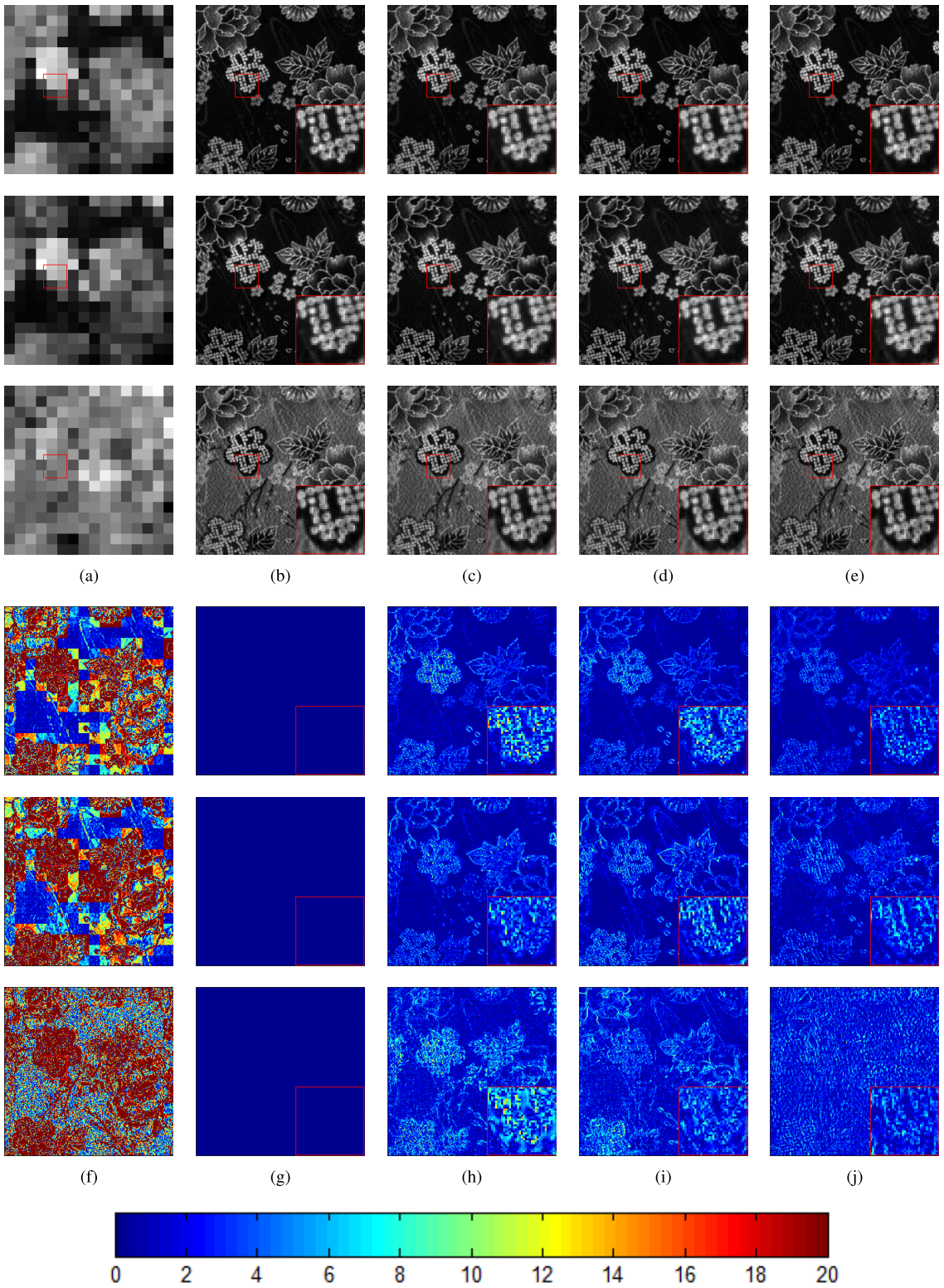


Fig. 5. Reconstructed images of *Cloth* at 450nm, 550nm and 650nm with Gaussian blur kernel and scaling factor $s = 32$. The first 3 rows show the reconstructed images for 460nm, 540nm and 620nm bands, respectively; the last 3 rows show the error images of the competing methods. (a) the LR images $\boldsymbol{X}$; (b) the original images $\boldsymbol{Z}$; (c) the CNMF method [22] (PSNR = 36.00 dB, RMSE = 4.04, ERGAS = 1.10, SAM = 3.87); (d) the BSR method [37] (PSNR = 36.68 dB, RMSE = 3.74, ERGAS = 1.03, SAM = 3.40); (e) the **proposed NSSR** method (PSNR = **39.04** dB, RMSE = **2.85**, ERGAS = **0.84**, SAM = **2.79**). (f) Error of $\boldsymbol{X}$. (g) Error of $\boldsymbol{Z}$. (h) CNMF [22]. (i) BSR [37]. (j) **Proposed NSSR**

sample consists of 61 bands over the range of $400 - 700$ nm. The total number of sampled spectral samples is *351*. The spatial resolution of the RGB image is 1420×1280. The transformation matrix $\boldsymbol{P}$ for transforming the hyperspectral images into RGB images are computed according to the method described in [35]. For hyperspectral images

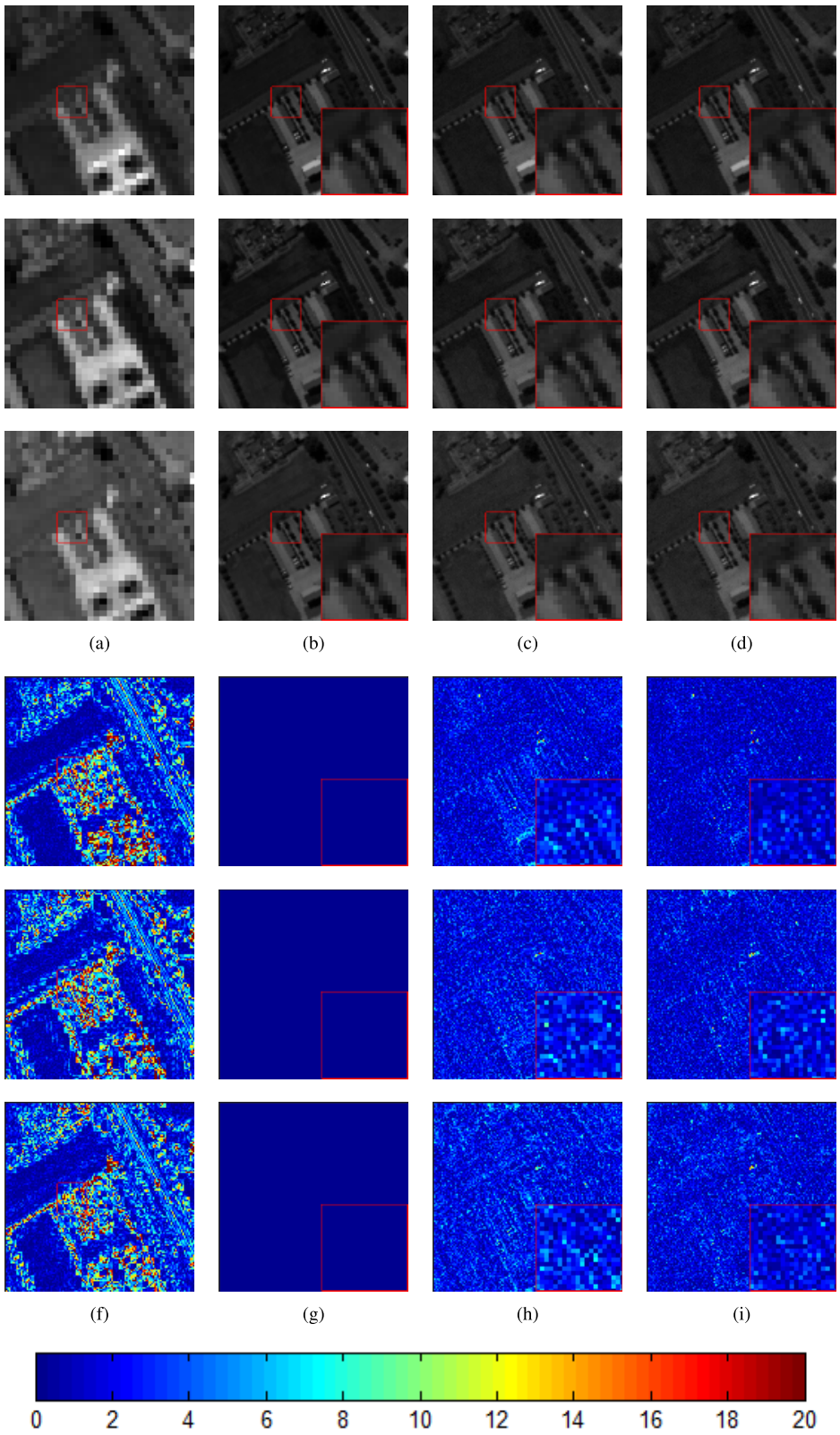


Fig. 6. Reconstructed images of *Pavia* image. The first 3 rows show the reconstructed images for 25-th, 45-th, and 60-th bands, respectively; the last 3 rows show the error images of the competing methods. (a) the LR images $\boldsymbol{X}$; (b) the Original image $\mathbf{Z}$; the BSR method [37] (RMSE=0.00947, ERGAS=0.850, SAM=1.492); (d) the proposed **NSSR** method (RMSE=0.00936, ERGAS=0.833, SAM=1.472); (f) Error of X; (g) Error of $\mathbf{Z}$; (h) Error of BSR [37]; (i) Error of **Proposed NSSR**.

containing more than L ($L > 33$) channels, we use the cubic interpolator to interpolate the matrix $\boldsymbol{P}$. The matrix $\boldsymbol{H}$ is set as a downsampling matrix that directly downsamples the original HR hyperspectral images according to the predefined mask [21]. The two types of data are integrated to generate the HR hyperspectral images. Fig. 7 shows the reproduced

(a) (b) (c) (d) (e) (f)

Fig. 7. Results on the real data captured with the hybrid camera developed in [21]. From left to right: (a) the HR RGB images captured in [21]; (b) the captured spectral samples in [21]; (c) reproduced RGB images from the reconstructed hyperspectral images by the method in [22]; (d) reproduced RGB images from the reconstructed hyperspectral images by the method in [21]; (e) reproduced RGB images from the reconstructed hyperspectral images by the proposed **NSSR** method; (f) Reconstructed hyperspectral spectral bands at 560nm by the proposed **NSSR** method.

RGB images from the reconstructed hyperspectral images by the competing methods. It can be seen that the RGB images reproduced by both the proposed method and the CNFM method are more consistent with the captured RGB images than that of [21], verifying that the proposed method and CNMF method can reconstruct more accurate hyperspectral images than the method of [21], which uses the bilateral upsampling method [16]. Obviously, the proposed **NSSR** method outperforms the CNMF method in recovering HR hyperspectral images from spectral samples and RGB images.

C. Parameters Selection

To evaluate the sensitivity of key parameters to the performance of the proposed method, we have varied them, i.e., the number of atoms of dictionary K, the number of iterations of **Algorithm 1** and **2**, and the regularization parameters η_1 and η_2. Fig. 8 plots the average PSNRs of the reconstructed images of the CAVE dataset as a function of the number of atoms K. We can see that the proposed method performs best when the number of atoms is in the range of 80$\sim$100 and is insensitive to the variation in the values of K in this range. In our implementation, we set $K = 80$.

Fig. 9 (a) plots the curve of the objective function values of Eq. (7), from which we can see that **Algorithm 1** typically converges after $J = 10$ iterations. We have also compared to the ADMM-based non-negative dictionary learning method that has been adopted in [25]. We can see that the proposed non-negative dictionary learning converges faster and can also converge to a smaller objective function value. Fig. 9 (b) plots the curve of the objective function value of Eq. (24).

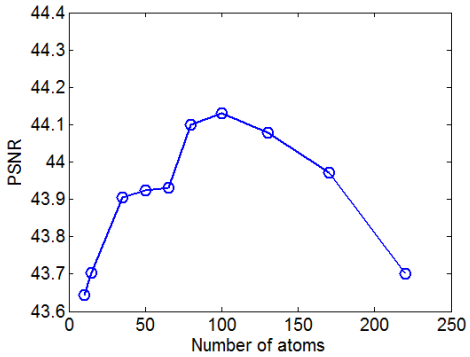


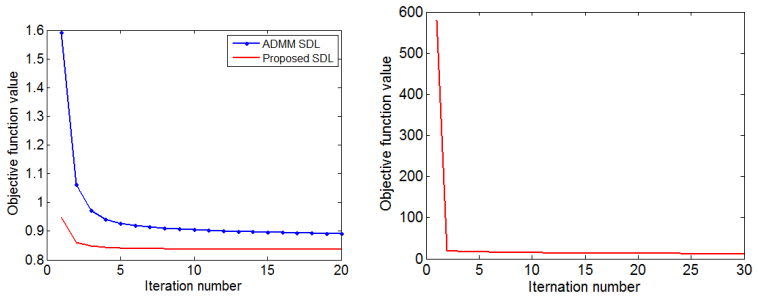
Fig. 8. The PSNR curve as a function of the number of atoms on the CAVE dataset.

It can be seen that **Algorithm 2** converges fast (usually after 25 $\sim$ 30 iterations).

Fig. 10 plots the curves of the PSNR values on the test image *Cloth* as the function of the regularization parameters η_1 and η_2. From Fig. 10, we can see that the performance of the proposed method is insensitive to the variation in the values of η_1 in the range of 0.005 $\sim$ 0.015 and in the values of η_2 in the range of $0.2 \times 10^{-4}$$\sim$$1.2 \times 10^{-4}$. Similar observations can be obtained for other test images.

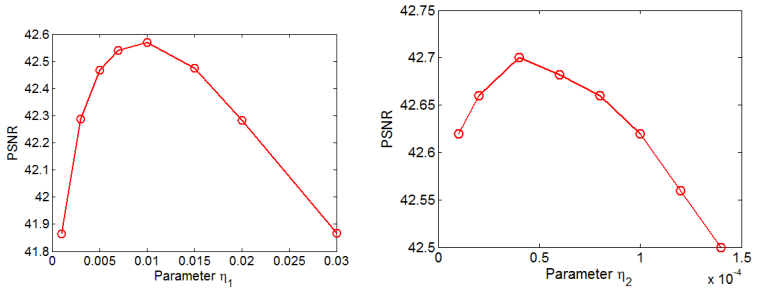
D. Computational Complexity

The computational complexity of the proposed method mainly consists of three parts: 1) the spectral dictionary learning; 2) computing the k-NN nearest neighbors searching per each patch within a window of size $h \times h$ for calculating μ_q;



(a) (b)

Fig. 9. The curves of the objective function values of Eqs. (7) and (24). (a) The curve of the objective function value of Eq. (7) versus iteration number; (b) The curve of the objective function value of Eq. (24) versus iteration number.



(a) (b)

Fig. 10. The PSNR curves as functions of the parameters η_1 and η_2. (a) The PSNR curve as a function of parameter η_1; (b) the PSNR curve as a function of parameter η_2.

TABLE V

RUNNING TIME (SECONDS) COMPARISON BETWEEN THE PROPOSED **NSSR** METHOD AND OTHER COMPETING METHODS ON A TEST IMAGE OF SIZE $512 \times 512 \times 31$ WITH SCALING FACTOR 16

G-SOMP+[26]	MF[20]	SNNMF[25]	CNMF[22]	BSR [37]	**Proposed NSSR**
221	697	3053	163	2984	**69**

3) the sparse coding of the sparse codes $\boldsymbol{A}$ by solving Eq. (24). The complexity of the spectral dictionary learning algorithm presented in **Algorithm 1** is $O(T_1 J(K^2 n + K^3 + KLn)$; while the complexity of computing the k-NN per-patch is $O(Nk(h^2 + logh^2))$. In **Algorithm 2**, the complexity of updating of $\boldsymbol{A}$, $\boldsymbol{Z}$ and $\boldsymbol{S}$ are $O(N)$, $O(\sqrt{\kappa}LNlogN)$, and $O(K^2L+K^3+KLN)$, respectively, where κ is the conditional number of the matrix to be inverted in Eq. (28) for $\boldsymbol{Z}$ update. Generally, $\kappa \in O(N)$. Therefore, the complexity for $\boldsymbol{Z}$ updating is $O(LN^{\frac{3}{2}}logN)$. The overall complexity of **Algorithm 2** is $O(T_2(LN^{\frac{3}{2}}logN + K^2L + K^3 + KLN))$. The overall complexity of the proposed hyperspectral image super-resolution under the assumption that $L \ll K \ll n \ll N$ and $T_1 Jn \geq N$ is

$$O(T_1 JK^2 n + N(kh^2 + klogh^2) + T_2 LN^{\frac{3}{2}}logN) \quad (29)$$

The proposed algorithm was implemented with Matlab language on an Intel Core i7-3770 3.4GHz CPU. A running time comparison between the proposed method and other competing methods is shown in Table V. From Table V, it can be seen that the proposed algorithm is the fastest when comparing with the other test methods. It is interesting to see that the SNNMF [25] is the slowest. This is because it alternates the optimization of the sparse codes and the spectral dictionary using the ADMM technique, both of which are of high computational complexity. This demonstrates the advantage of the proposed optimization algorithms. For BSR method [37], we have optimized the parameters of the BSR method [37] for best performance, i.e. the iteration number and the number of dictionary atoms has been increased. This makes the BSR method [37] slower. It can be seen that the proposed algorithm runs much faster than the BSR method [37] and the G-SOMP+ method [26].

V. CONCLUSIONS

HR hyperspectral imaging is challenging due to various hardware limitations. In this paper, we propose an effective sparsity-based hyperspectral image super-resolution method to reconstruct a HR hyperspectral image from a LR hyperspectral image and a HR RGB image of the same scene. The hyperspectral dictionary representing the typical reflectance spectra signatures of the scene is first learned from the LR hyperspectral image. Specifically, an efficient non-negative dictionary learning algorithm is proposed using a block-coordinate descent algorithm. The sparse codes of the HR hyperspectral

image with respect to the learned dictionary are then estimated from the corresponding HR RGB image. To improve the accuracy of estimating sparse codes, a new clustering-based non-negative structured sparse representation framework is proposed to exploit both the spatial and spectral correlations. The estimated sparse codes are then used with the spectral dictionary to reconstruct the HR hyperspectral images. Experimental results on both public datasets and real-world LR hyperspectral images show that the proposed method can achieve smaller reconstruction errors and better visual quality on most test images than existing HR hyperspectral recovery methods in the literature.

REFERENCES

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

[2] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[4] H. V. Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2010, pp. 44–51.

[5] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition using 3D-DCT and partial least squares," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2013, pp. 57.1–57.10.

[6] 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 Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.

[7] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.

[8] W. J. Carper, T. M. Lillesand, and R. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogram. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.

[9] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogram. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[10] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.

[11] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva, "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 302–306, Apr. 2009.

[12] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[13] A. Minghelli-Roman, L. Polidori, S. Mathieu-Blanc, L. Loubersac, and F. Cauneau, "Spatial resolution improvement by merging MERIS-ETM images for coastal water monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 2, pp. 227–231, Apr. 2006.

[14] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, "Unmixing-based landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, Jul. 2008.

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

[16] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007, Art. no. 96.

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

[18] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.

[19] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2423–2435, Dec. 2011.

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

[21] C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin, "Acquisition of high spatial and spectral resolution video with a hybrid camera system," *Int. J. Comput. Vis.*, vol. 110, no. 2, pp. 141–155, Nov. 2014.

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

[23] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[24] D. D. Lee and S. H. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2001, pp. 556–562.

[25] E. Wycoff, T. H. Chan, K. Jia, W. K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 1409–1413.

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

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

[28] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 457–464.

[29] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

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

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

[32] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 193–200.

[33] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.

[34] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[35] *Tutorial on Transforming Hyperspectral Images*, accessed on Mar. 2013. [Online]. Available: http://personalpages.manchester.ac.uk/staff/david.foster/Tutorial_HSI2RGB/Tutorial_HSI2RGB.html

[36] J. Bieniarz, R. Müller, X. X. Zhu, and P. Reinartz, "Hyperspectral image resolution enhancement based on joint sparsity spectral unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2014, pp. 2645–2648.

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

[38] C. Grohnfeldt, X. X. Zhu, and R. Bamler, "Jointly sparse fusion of hyperspectral and multispectral imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 4090–4093.

[39] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[40] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

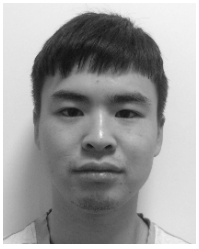
[41] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3631–3640.

[42] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.
[43] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009.
[44] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 3176–3180.
[45] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3586–3594.
[46] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Nice, France, Jan. 2000, pp. 99–103.
[47] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, Dec. 2007.

Weisheng Dong (M'11) received the B.S. degree in electronics engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree in circuits and system from Xidian University, Xi'an, China, in 2010.

He was a Visiting Student with Microsoft Research Asia, Beijing, China, in 2006. From 2009 to 2010, he was a Research Assistant with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. In 2010, he joined the School of Electronic Engineering, Xidian University, as a Lecturer, where he has been an Associate Professor since 2012. His current research interests include inverse problems in image processing, sparse signal representation, and image compression.

Dr. Dong was a recipient of the best paper award at the SPIE Visual Communication and Image Processing in 2010. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Fazuo Fu received the B.S. degree in information and computing science from Xidian University, Xi'an, China, in 2014, where he is currently pursuing the M.S. degree in circuits and system.

His current research interests include spectral image super-resolution and denoising.

Guangming Shi (SM'10) received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, in 1985, 1988, and 2002, respectively.

He joined the School of Electronic Engineering, Xidian University, in 1988. From 1994 to 1996, as a Research Assistant, he cooperated with the Department of Electronic Engineering, The University of Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University, and he was the Head of the National Instruction Base of Electrician and Electronic in 2004. In 2004, he had studied in the Department of Electronic Engineering at the University of Illinois at Urbana– Champaign. He is currently the Deputy Director of the School of Electronic Engineering with Xidian University, and the Academic Leader in the subject of Circuits and Systems. He has authored or co-authored over 60 research papers. His research interests include compressed sensing, theory and design of multirate filter banks, image denoising, low bit-rate image/video coding, and implementation of algorithms for intelligent signal processing (using DSP and FPGA).

Xun Cao (S'10–M'12) received the B.S. degree from Nanjing University, Nanjing, China, in 2006, and the Ph.D. degree from the Department of Automation, Tsinghua University, Beijing, China, in 2012. He was visiting Philips Research, Aachen, Germany, in 2008, and Microsoft Research Asia, Beijing, China, from 2009 to 2010. He was a Visiting Scholar with the University of Texas at Austin, Austin, TX, USA, from 2010 to 2011. He is currently an Associate Professor with the School of Electronic Science and Engineering, Nanjing University. His research interests include computational photography, image-based modeling and rendering, and 3DTV systems.

Jinjian Wu received the B.Sc. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore. From 2013 to 2014, he was a Post-Doctoral Research Fellow with Nanyang Technological University. From 2013 to 2015, he was a Lecturer with Xidian University. Since 2015, he has been an Associate Professor with the School of Electronic Engineering, Xidian University.

His research interests include visual perceptual modeling, saliency estimation, quality evaluation, and just noticeable difference estimation. He has served as the TPC Member of ICME2014, ICME2015, PCM2015, and ICIP2015. He received the best student paper of ISCAS 2013.

Guangyu Li received the B.S. degree in detection guidance and control technology and the M.S. degree in electronics and communications engineering from Xidian University, Xi'an, China, in 2013 and 2016, respectively. He is currently with Huawei Technologies Company, Ltd., Xi'an.

Xin Li received the B.S. (Hons.) degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000. He was a Technical Staff Member with the Sharp Laboratories of America, Camas, WA, USA, from 2000 to 2002. Since 2003, he has been a Faculty Member with the Lane Department of Computer Science and Electrical Engineering. His research interests include image/video coding and processing. He is currently serving as a member of the Image, Video and Multidimensional Signal Processing Technical Committee and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He was a recipient of the Best Student Paper Award at the Conference of Visual Communications and Image Processing in 2001, the Best Student Paper Award at the IEEE Asilomar Conference on Signals, Systems and Computers in 2006, and the best paper award at the Conference of Visual Communications and Image Processing in 2010.