Self-Similarity Constrained Sparse Representation for Hyperspectral Image Super-Resolution

Xian-Hua Han, *Member, IEEE*, Boxin Shi, *Member, IEEE*, and Yinqiang Zheng*, *Member, IEEE*

***Abstract*—Fusing a low-resolution hyperspectral image with the corresponding high-resolution multispectral image to obtain a high-resolution hyperspectral image is an important technique for capturing comprehensive scene information in both spatial and spectral domains. Existing approaches adopt sparsity promoting strategy, and encode the spectral information of each pixel independently, which results in noisy sparse representation. We propose a novel hyperspectral image super-resolution method via a self-similarity constrained sparse representation. We explore the similar patch structures across the whole image and the pixels with close appearance in local regions to create global-structure groups and local-spectral super-pixels. By forcing the similarity of the sparse representations for pixels belonging to the same group and super-pixel, we alleviate the effect of the outliers in the learned sparse coding. Experiment results on benchmark datasets validate that the proposed method outperforms the state-of-the-art methods in both quantitative metrics and visual effect.**

***Index Terms*—Hyper-spectral image super-resolution, global-structure self-similarity, local-spectral self-similarity, dictionary learning, non-negative sparse coding.**

I. Introduction

Hyperspectral (HS) imaging is an emerging technique for simultaneously obtaining a set of images of the same scene at many number of narrow band wavelengths. The rich spectra significantly enrich the captured scene information and greatly enhance performance in many tasks [1], such as object recognition and classification [2]–[7], tracking [8], segmentation [9], medical image analysis [10], and remote sensing [11]–[15]. HS imaging achieves abundant spectral information by simultaneously collecting a large number of spectral bands within a target wavelength interval. To guarantee sufficiently high signal-to-noise ratio, photon collection is usually performed in a much larger spatial region on the sensor thus results in a much lower spatial resolution than RGB or Multi-Spectral (MS) images. Such low resolution limits the applicability of HS imaging. To enhance the resolution of HS images, high resolution RGB or MS images of the same scene are usually captured and then fused into the HS image, a critical task referred to as HS image super-resolution (HSI SR) on the basis of hybrid fusion.

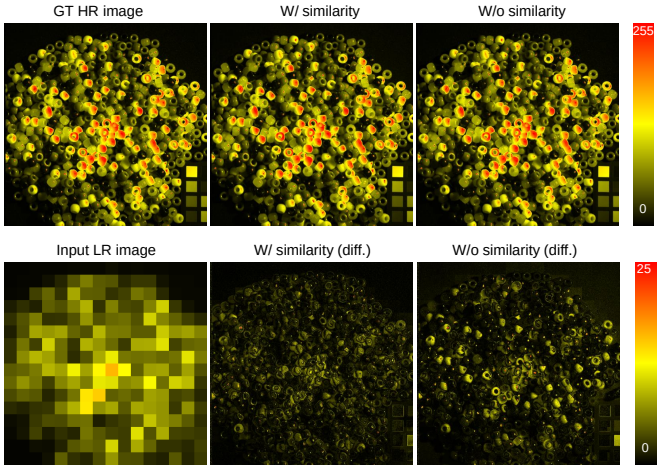


Figure 1. An example of HR images recovered from our method. The first column shows one channel of the ground truth HR image and the input LR image, respectively. The second and third columns show results from our method with and without self-similarity. In each column, the upper row shows the recovered image and the lower row shows the absolute difference map w.r.t. the ground truth.

Recent HSI SR methods are motivated by the sparse representation [16]–[22], which transfers the coded sparse representation of the high-resolution RGB (HR-RGB) images to guide the recovery of the high-resolution hyperspectral (HR-HS) image. Most existing methods encode the spectra of the HR-RGB image pixel-wise, which introduces a noisy sparse representation and degrades the final reconstruction. In addition, the sparse representation is usually optimized via orthogonal pursuit matching [2], [23], which has high computational cost.

For a natural scene, there mainly exist two types of similarities: One is the patches with close appearance across the whole image — we call it *global-structure similarity*, and the other one is the close color values for neighboring pixels — we call it *local-spectral similarity*. We observe that pixels sharing global and/or local similarities will have similar sparse representations, and such similarity constraints are investigated to build the inter-pixel relationship for more robust estimation.

Manuscript received December 27, 2017; revised May 31, 2018; accepted June 27, 2018. Date of publication ** **, ****

This work is supported in part by the open collaborative research program at National Institute of Informatics (NII), Japan (FY2017, FY2018), and the Recruitment Program for Young Professionals (a.k.a. 1000 Youth Talents) in China, National Natural Science Foundation of China (61702047), Beijing Natural Science Foundation (4174098), and the Fundamental Research Funds for the Central Universities (2017RC02).

This paper was recommended by Associate Editor **.

X.-H. Han is with Graduate School of Science and Technology for Innovation, Yamaguchi University, 1677-1 Yoshida, Yamaguchi City, Yamaguchi, 753-8511, Japan. (e-mail: hanxhua@yamaguchi-u.ac.jp).

B. Shi is with the National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Yiheyuan Road No. 5, Haidian district, Beijing, 100871, China (e-mail: shiboxin@pku.edu.cn).

Y. Zheng is with National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan (*Corresponding author, e-mail: yqzheng@nii.ac.jp).

We present a novel HSI SR method with self-similarity constrained sparse representation. The sparse representation is calculated via coupled minimizing the reconstruction error of the available low-resolution hyperspectral image (LR-HS) and HR-RGB image. Furthermore, the self-similarity is explored by clustering global-structure groups and creating local-spectral super-pixels in the HR-RGB image, and then applied to force the similarity of the estimated sparse representation within the same group and super-pixel. The self-similarity constraint significantly suppresses the noise in a pixel-wise representation, thus achieves more reliable HSI SR reconstruction, as shown in the example result of Fig. 1. Our complete pipeline includes a dictionary learning (Step 1) and a global and local self-similarity constrained sparse representation (Step 2) solved as convex optimization via the alternating direction multiplier method (ADMM) (Step 3), as shown in Fig. 2.

The three major advantages of the proposed method are summarized as:

1) We propose a coupled sparse representation strategy via simultaneously minimizing the reconstruction error of the available LR-HS and HR-RGB images;
2) We explore the global-structure and local-spectral self-similarity in the input HS-RGB image to constrain the sparse representation to achieve more robust performance;
3) We solve the constrained sparse representation problem with the ADMM optimization, which is much more efficient than the conventional sparse coding algorithms [23]–[26]. Experiment results on benchmark datasets validate that the proposed method outperforms the state-of-the-art methods [16], [17], [22], [27]–[30] in both quantitative metrics and visual effect.

The rest of the paper is organized as follows. We firstly review the related literature in Section II. We then formulate the relationship between the target HR-HS image and the input LR-HS, and HR-MS image pair, and model the HSI SR problem with sparse representation in Section III. The details of the proposed self-similarity constraint and optimization method are presented in Section IV. Experimental results compared with existing methods are evaluated in Section V. Finally, Section VI concludes the paper.

II. Related Work

Though the wide demand of high-resolution HS images in different application fields ranging from remote sensing to medical imaging, it is still difficult to simultaneously achieve high-resolution in both spatial and spectral domains due to technique and budget constraints [31]. Thus this has inspired considerable attention in the research literature to generate high resolution HS images via image processing and machine learning techniques based on the available LR-HS and HR-MS/HR-RGB images. In remote sensing, a high resolution pan-chromatic image is usually available accompanying with a low resolution MS or HS image, and the fusion of these two images is generally known as pan-sharpening [32]–[38]. In this scenario, most popular approaches concentrated on the reliable illumination restoration based on intensity substitution and projection with the explored hue saturation and principle component analysis [32], [33], which generally cause spectral distortion in the resulting image [39]. On the contrary, it is more common to use HR-RGB images in computer vision literature, since more spectral information is recorded in a RGB image than in a pan-chromatic image.

Recently, the HS image super resolution based on matrix factorization and spectral unmixing [40]–[43], which are mainly motivated by the fact the HS observations can be represented by the basis of reflectance functions (the spectral response of the pure material) and their corresponding sparse coefficients denoting the fractions of each material at each location, has been actively investigated [16], [17], [27], [28]. Yokoya et al. [28] proposed a coupled non-negative matrix factorization (CNMF) to estimate the HR-HS image from a pair of HR-MS and LR-HS images. Although the CNMF approach achieved impressive spectral recovery performance, its solution is generally not unique [44], and thus the spectral recovery results are not always satisfactory. On the other hand, Lanaras et al. [17] integrated coupled spectral unmixing strategy into HS super-resolution, and applied the proximal alternation linearized minimization scheme for optimization, which requires the initial points of the two decomposed reflectance functions and the endmember vectors. In addition, according to the physical meaning of the reflectance functions and the implementation consideration, the number of the pure materials in the observed scene is often assumed smaller than the spectral band number, which does not always meet the real application.

Motivated by the success of the sparse representation in natural image analysis [45], the sparsity promoting approaches without explicit physically meaningful constraints on the basis, which thus permit over-complete basis, have been applied for HS super-resolution [18], [19]. Grohnfeldt et al. [18] proposed a joint sparse representation by firstly learning the corresponding HS and MS (RGB) patch dictionaries using the prepared pairs, and estimated the same sparse coefficients of the combined MS and previously reconstructed HS patches for each individual band, which mainly focused on the approximation of the local structure (patch) and completely ignored the correlation between channels. Then several researches [19], [22] explored the sparse spectral representation instead of the local structure. Akhtar et al. [22] proposed a sparse spatio-spectral representation via assuming the same used atoms for reconstructing the spectra of the pixels in a local grid region, and developed generalized simultaneous orthogonal matching pursuit (G-SOMP) for estimating the sparse coefficients. Further, the same research group explored a Bayesian dictionary learning and sparse coding algorithm for HS image super-resolution and manifested improved performance. Most recently, Dong et al. [30] investigated a non-negative structured sparse representation (NSSR) approach to recover a HR-HS image from LR-HS and HR-RGB images and proposed to use the alternating direction multiplier method (ADMM) for solving, which gave impressive recovery performance compared the other existing approaches.

Our proposed HSI SR method is related to the sparsity

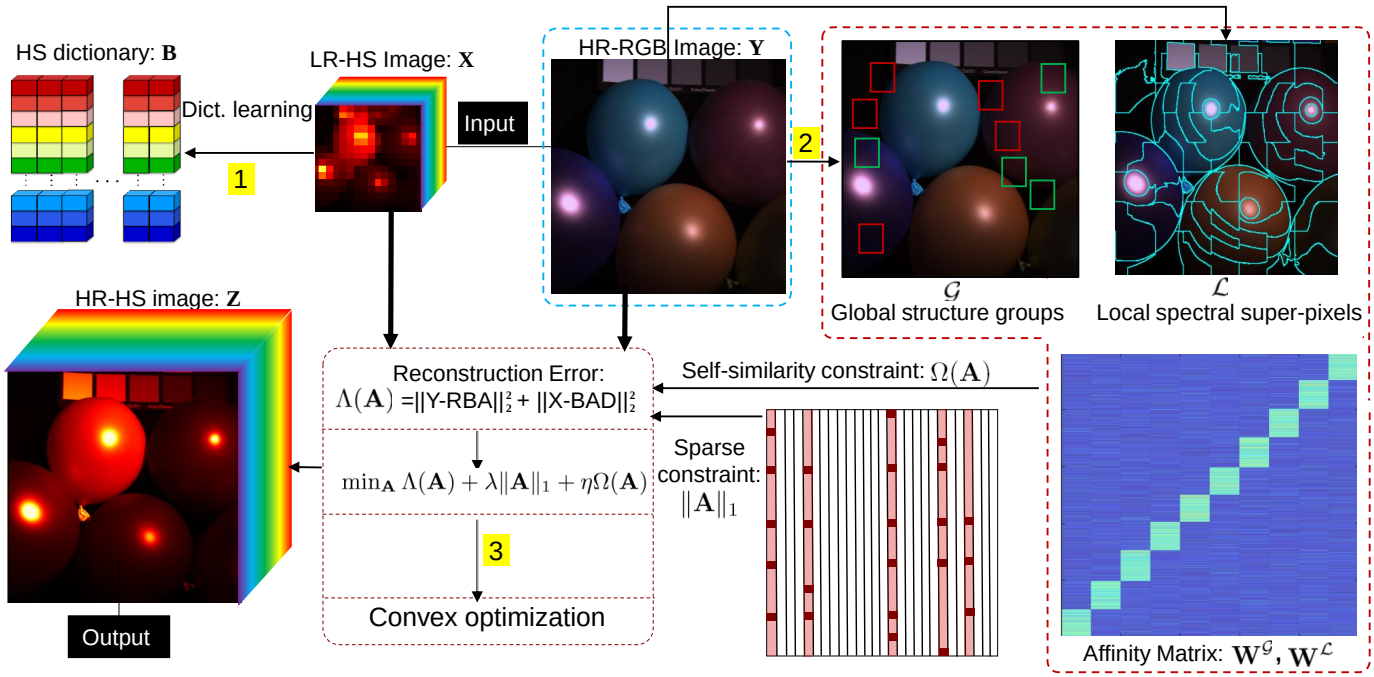


Figure 2. Schematics of the proposed approach: (1) Learn the HS dictionary $\mathbf{B}$ from the input LR-HS image $\mathbf{X}$; (2) Explore the global-structure and local-spectral self-similarity; (3) Convex optimization of the objective function with sparse and self-similarity constraints on the sparse matrix $\mathbf{A}$ to estimate the required HR-HS image $\mathbf{Z}$.

promoting strategy and the coupled unmixing strategy [17], but has major differences with all of the existing approaches. We optimize the HR-HS output via minimizing the coupled reconstruction error of the available LR-HR and HR-RGB images with the following characteristics: 1) the sparse representation with over-complete spectral dictionary (larger dictionary numbber than spectral dimension) instead of the sub-complete dictionary (endmember) in the coupled unmixing strategy [17]; 2) the self-similarity of the global structures and the local spectra present in the available HR-RGB image for sparse representation, which can achieve more robust performance.

III. Problem Formulation

Our goal is to estimate a HR-HS image $\mathbf{Z}' \in \mathbb{R}^{W\times H\times L}$, where W and H denote the spatial dimensions and L is the spectral band number, from a LR-HS image $\mathbf{X}' \in \mathbb{R}^{w\times h\times L}$ ($w \ll W$, $h \ll H$) and a HR-MS image $\mathbf{Y}' \in \mathbb{R}^{W\times H\times l}$ ($l \ll L$). In our experiments, the available HR-MS image is a RGB image with spectral band number $l = 3$. The matrix forms of $\mathbf{Z}'$, $\mathbf{X}'$, and $\mathbf{Y}'$ are denoted as $\mathbf{Z} \in \mathbb{R}^{L\times N}$ ($N = W \times H$), $\mathbf{X} \in \mathbb{R}^{L\times M}$ ($M = w \times h$), and $\mathbf{Y} \in \mathbb{R}^{3\times N}$, respectively. Both $\mathbf{X}$ (LR-HS) and $\mathbf{Y}$ (HR-RGB) can be expressed as a linear transformation from $\mathbf{Z}$ (the desired HS image) as

$$\mathbf{X} = \mathbf{ZD} \quad \text{and} \quad \mathbf{Y} = \mathbf{RZ}, \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{N\times M}$ is the decimation matrix blurring and down-sampling the HR-HS image to form the LR-HS image, and $\mathbf{R} \in \mathbb{R}^{3\times L}$ represents the camera spectral response matrix that maps the HR-HS image to the HR-RGB image. Given $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{Z}$ can be estimated by minimizing the following reconstruction error:

$$\mathbf{Z}^* = \underset{\mathbf{Z}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{ZD}\|_F^2 + \|\mathbf{Y} - \mathbf{RZ}\|_F^2. \tag{2}$$

Since the number of the unknowns (NL) is much larger than the number of available measurements ($ML + 3N$), the above optimization problem is highly ill-posed, and proper regularization terms are required to narrow the solution space and ensure stable estimation. A widely adopted constraint is that each pixel spectrum $\mathbf{z}_n \in \mathbb{R}^L$ of $\mathbf{Z}$ lies in a low-dimensional space, and it can be decomposed as [46]

$$\begin{aligned} \mathbf{z}_n &= \sum_{k=1}^{K} \mathbf{b}_k a_{k,n} = \mathbf{Ba}_n, \\ s.t.,\ \mathbf{b}_{i,k} &\geq 0, a_{k,n} \geq 0, \sum_{k=1}^{K} a_{k,n} = 1, \end{aligned} \tag{3}$$

where $\mathbf{B} \in \mathbb{R}^{L\times K}$ is the stack of the spectral signature ($\mathbf{b}_k$, also called endmember) of K distinct materials, and $\mathbf{a}_n$ denotes the fractional abundance of the K materials for the n-th pixel. Considering the physical property of the spectral reflectance, the elements in the endmember spectra and the fraction magnitude of the abundance are non-negative as shown in the first and second constraint terms, and the sum of abundance vector for each pixel is one, which means the fractional vector is sparse.

According to $\mathbf{Y} = \mathbf{RZ}$, each pixel $\mathbf{y}_n \in \mathbb{R}^3$ in the HR-RGB image can be decomposed as

$$\mathbf{y}_n = \mathbf{Rz}_n = \mathbf{RBa}_n = \hat{\mathbf{B}}\mathbf{a}_n, \tag{4}$$

where $\hat{\mathbf{B}}$ is the RGB spectral dictionary obtained via transforming the HS dictionary $\mathbf{B}$ with camera spectral response matrix $\mathbf{R}$. With a corresponding set of fixed spectral dictionaries $\hat{\mathbf{B}}$ and $\mathbf{B}$, the sparse fractional vector $\mathbf{a}_n$ can be predicted from the HR-RGB pixel $\mathbf{y}_n$.

The matrix representation of Eqs. 3 and 4 is denoted as

$$\mathbf{Z} = \mathbf{B}\mathbf{A} \quad \text{and} \quad \mathbf{Y} = \hat{\mathbf{B}}\mathbf{A}, \tag{5}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N] \in \mathbb{R}_+^{K \times N}$ is a non-negative sparse coefficient matrix. Substituting Eq. 5 into Eq. 2, we obtain[1]

$$\{\mathbf{B}^*, \mathbf{A}^*\} = \underset{\mathbf{B},\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{B}\mathbf{A}\mathbf{D}\|_F^2 + \|\mathbf{Y} - \hat{\mathbf{B}}\mathbf{A}\|_F^2. \tag{6}$$

Our goal is to solve both spectral dictionary $\mathbf{B}$ and coefficient matrix $\mathbf{A}$ with proper the regularization terms to achieve stable and accurate solution.

IV. Proposed Method

The complete pipeline of our approach is illustrated in Fig. 2. Our main contribution is to propose a non-negative sparse representation coupled with self-similarity constraint to regularize the solution of Eq. 6. Denoting $\Lambda(\mathbf{B}, \mathbf{A}) = \|\mathbf{X} - \mathbf{B}\mathbf{A}\mathbf{D}\|_F^2 + \|\mathbf{Y} - \hat{\mathbf{B}}\mathbf{A}\|_F^2$, we add two additional terms to Eq. 6 as

$$\{\mathbf{B}^*, \mathbf{A}^*\} = \underset{\mathbf{B},\mathbf{A}}{\operatorname{argmin}} \Lambda(\mathbf{B}, \mathbf{A}) + \lambda\|\mathbf{A}\|_1 + \eta\Omega(\mathbf{A}), \tag{7}$$

where $\|\mathbf{A}\|_1$ is the sparse constraint on the coefficient matrix, $\Omega(\mathbf{A})$ is the self-similarity regularization for the coefficient vector, and λ, η are weighting factors. Equation 7 will be optimized in three steps: 1) Online HS dictionary learning from the input LR-HS image; 2) Extracting the global-structure and local-spectral self-similarity from the input HR-RGB image; 3) The global convex optimization for estimating the HR-HS image given the learned HS dictionary and self-similarity constrained sparse representation. We will introduce the details of these operations in following three subsections.

A. Online HS dictionary learning

Due to the large variety of the HS reflectance from different materials, learning a common HS dictionary for various scenes with different materials tends to give considerable spectral distortion. We instead learn the HS dictionary directly from the observed LR-HS image $\mathbf{X}$ in an online manner to build the HS dictionary that better represents the scene spectra as

$$\{\mathbf{B}^*, \hat{\mathbf{A}}^*\} = \underset{\mathbf{B},\hat{\mathbf{A}}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{B}\hat{\mathbf{A}}\|_F^2 + \lambda\|\hat{\mathbf{A}}\|_1, \tag{8}$$

where $\hat{\mathbf{A}}$ is the sparse matrix for the pixels in the LR-HS image. Since the non-negative constraints are applied to both sparse code $\hat{\mathbf{A}}$ and the dictionary $\mathbf{B}$, existing dictionary learning method such as K-SVD method cannot be applied here. We follow the optimization algorithm in [29] and apply ADMM technique to convert the constrained dictionary learning problem into an unconstrained version. The unconstrained dictionary learning is then solved with alternating optimization. After the HS dictionary $\mathbf{B}^*$ is learned from the observed LR-HS image via Eq. 8, we will fix it and only optimize $\mathbf{A}$ for solving Eq. 7.

[1]The non-negative constraints on both $\mathbf{B}$ and $\mathbf{A}$ are applied in the same manner as in Eq. 3. Unless otherwise noted, the non-negative constraint are imposed on both dictionary and sparse matrix in the following deductions.

B. Self-similarity constraint

We formulate the regularization term $\Omega(\mathbf{A})$ in Eq. 7 with two types of the self-similarities (see Fig. 2 for illustration):

- Global-structure self-similarity: Pixels whose concatenated RGB values within a local square windows are similar share similar hyperspectral information. This applies to both nearby patches and non-local patches in the whole image plane.
- Local-spectral self-similarity: The sparse vectors for different HR pixels are similar in a local region (superpixel) [47] by assuming that in HR-RGB images, pixels in a local region have the same material and RGB values. Note the superpixel is usually not a square region.

The global-structure self-similarity is represented by global-structure groups $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_P\}$ (in total P groups), which are formed by clustering all similar patches (both local and non-local) in the HR-RGB image with K-means [48]; $\mathbf{g}_p$ is a vector (each $\mathbf{g}_p$ may have different length) composed by the pixel indices of the p-th group. The local-spectral self-similarity is represented by super-pixels $\mathcal{L} = \{\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_Q\}$ (in total Q super-pixels), which are obtained via SLIC superpixel segmentation method [49]; $\mathbf{l}_q$ is also a vector including the pixel indices of the q-th superpixel. Since the pixels in the same global-structure group have similar spectral structure, we approximate the sparse code of a pixel by a weighted average of the sparse matrix for all pixels in this group. Similarly, the sparse vector of a pixel can also be approximated by a weighted average of the sparse matrix for all pixels in the same local-spectral superpixel. With both self-similarity constraints, the sparse vector for the n-th pixel is represented as

$$\begin{aligned} \mathbf{a}_n =& \gamma \sum_{i \in \mathbf{g}_p} w_{n,i}^{\mathcal{G}} \mathbf{a}_i + (1-\gamma) \sum_{j \in \mathbf{l}_q} w_{n,j}^{\mathcal{L}} \mathbf{a}_j, \\ & \text{with} \quad n \in \mathbf{g}_p \wedge n \in \mathbf{l}_q. \end{aligned} \tag{9}$$

Here, $w_{n,i}^{\mathcal{G}}$ is the global-structure weight for the n-th sparse vector $\mathbf{a}_n$; it adjusts and merges the contribution of the i-th sparse vector $\mathbf{a}_i$ belonging to the same global-structure group. Analoguely, $w_{j,n}^{\mathcal{L}}$ weights the j-th sparse vector $\mathbf{a}_j$ belonging to the same local-spectral superpixel. And γ is a parameter for balancing the contribution between the global-structure and local-spectral self-similarity.

To be more specific, $w_{n,i}^{\mathcal{G}}$ ($0 < w_{n,i}^{\mathcal{G}} < 1$ and $\sum_i w_{i,n}^{\mathcal{G}} = 1$) measures the similarity between the RGB intensities of patches $\mathbf{p}_n$ and $\mathbf{p}_i$ centered around the n-th and i-th pixels. Each patch is a set of pixels in a $R \times R$ window, so each $\mathbf{p}$ is a $3R^2$-dimensional ($R \times R \times$ RGB) vector. It is a decreasing function of the Euclidean distance between the spatial RGB values as

$$w_{i,n}^{\mathcal{G}} = \frac{1}{z_n^{\mathcal{G}}} \exp^{-\frac{||\mathbf{p}_i - \mathbf{p}_n||^2}{h^{\mathcal{G}}}}, \tag{10}$$

where $z_n^{\mathcal{G}}$ is a normalization term defined as $z_n^{\mathcal{G}} = \sum_{i \in \mathbf{g}_p} \exp -\frac{||\mathbf{p}_i - \mathbf{p}_n||^2}{h^{\mathcal{G}}}$ to ensure that $\sum_i w_{i,n}^{\mathcal{G}} = 1$ and $h^{\mathcal{G}}$ is a smoothing kernel for $3R^2$-dimensional vectors. The local-spectral weight $w_{n,j}^{\mathcal{L}}$ is defined in the exactly same format, but with $\mathbf{p}_n$ and $\mathbf{p}_i$ being the RGB values of the n-th and i-th pixels (so each $\mathbf{p}$ is a 3-dimensional vector here) and a smoothing kernel $h^{\mathcal{L}}$ for 3-dimensional vectors.

We then build affinity matrices $\mathbf{W}^{\mathcal{G}} \in \mathbb{R}^{N \times N}$ and $\mathbf{W}^{\mathcal{L}} \in \mathbb{R}^{N \times N}$, whose element encodes the pairwise similarity calculated using Eq. 10. Finally, the regularization term constrained by two types of self-similarities is represented as

$$\Omega(\mathbf{A}) = \|\mathbf{A} - \gamma \mathbf{W}^{\mathcal{G}} \mathbf{A} - (1-\gamma)\mathbf{W}^{\mathcal{L}} \mathbf{A}\|_F^2. \tag{11}$$

To determine the smoothing kernel width $h^{\mathcal{G}}$ and $h^{\mathcal{L}}$, we firstly calculate the average squared distance of all color channels (RGB) between the pixel pairs in the same superpixel of the available HR-RGB image, and then obtain the mean value on all images of a given dataset.

$$h = \frac{1}{|Img|} \sum_{Img} \frac{1}{Q} \sum_{q} \frac{2}{|l_q| \times |l_q|} \sum_{i,n \in \mathbf{l}_q} \frac{||\mathbf{p}_i - \mathbf{p}_n||^2}{3}, \tag{12}$$

where $|l_q|$ and Q denote the pixel number in the q-th superpixel and the superpixel number in a HR-RGB image, respectively. $|Img|$ represents the number of images in a given dataset. Finally, we set $h^{\mathcal{L}} = 2h \times 3$ (each $\mathbf{p}$ is a 3-dimensional vector here) and $h^{\mathcal{G}} = 2h \times 3R^2$ (each $\mathbf{p}$ is a $3R^2$-dimensional vector here). For all our experiments in the following section, R is set to 5 and the calculated h is about 45.

C. Convex optimization

Given the HS dictionary $\mathbf{B}^*$ pre-learned using Eq. 8 and the regularization term with self-similarity in Eq. 11, Eq. 7 is convex and can be efficiently solved by optimization algorithm. We apply the ADMM technique to solve Eq. 7 via transformation with the following augmented Lagrangian function

$$\begin{aligned} L(\mathbf{A}, \mathbf{Z}, \mathbf{V}, \mathbf{T}_1, \mathbf{T}_2) = \\ &\|\mathbf{Y} - \tilde{\mathbf{B}}\mathbf{V}\|_F^2 + \|\mathbf{X} - \mathbf{Z}\mathbf{D}\|_F^2 \\ &+ \lambda\|\mathbf{A}\|_1 + \eta\|\mathbf{A} - \gamma\mathbf{W}^G\mathbf{V} - (1-\gamma)\mathbf{W}^L\mathbf{V}\|_F^2 \\ &+ \langle \mathbf{T}_1, \mathbf{B}\mathbf{V} - \mathbf{Z}\rangle + \rho\|\mathbf{B}\mathbf{V} - \mathbf{Z}\|_F^2 \\ &+ \langle \mathbf{T}_2, \mathbf{V} - \mathbf{A}\rangle + \rho\|\mathbf{V} - \mathbf{A}\|_F^2, \end{aligned} \tag{13}$$

where $\mathbf{T}_1$ and $\mathbf{T}_2$ are the matrices of the Lagrangian multipliers, and $\langle \cdot, \cdot \rangle$ denotes the inner product and $\rho > 0$ is the penalty parameter. For convenience, we set $\mathbf{E} = \gamma\mathbf{W}^G\mathbf{V} - (1-\gamma)\mathbf{W}^L\mathbf{V}$ and iteratively solve Eq. 13 with each variable initialized to a matrix with all elements as 0.

To solve the optimization subproblem of $\mathbf{V}$, we set the derivative of Eq. 13 w.r.t. $\mathbf{V}$ as 0 while fixing the other variables, and yield

$$\begin{aligned} \mathbf{V} =&[\tilde{\mathbf{B}}^T\tilde{\mathbf{B}} + (\eta+\rho)\mathbf{B}^T\mathbf{B} + \rho\mathbf{I}]^{-1} \\ &[\tilde{\mathbf{B}}^T\mathbf{Y} + \eta\mathbf{E}\rho\mathbf{B}^T(\mathbf{Z}^{(t)} + \frac{\mathbf{T}_1}{2\rho}) + \rho(\mathbf{A}^{(t)} + \frac{\mathbf{T}_2}{2\rho})], \end{aligned} \tag{14}$$

Table I
PARAMETER SETTINGS IN OUR EXPERIMENTS.

Parameters	CAVE dataset	Harvard dataset
P	4096	8192
Q	4096	16284
R	5	
$h^{\mathcal{G}}$	90×75	
$h^{\mathcal{L}}$	90×3	
γ	0.5	
η	0.025	
λ	0.0001	
ρ	0.001	

where $\mathbf{I}$ is the identity matrix and $\mathbf{E}$ can be calculated according to the previous $\mathbf{V}$, which would be updated at the end of each iteration. Similarly, by setting the derivative of Eq. 13 w.r.t. $\mathbf{Z}$ as 0, and $\mathbf{A}$ as $\mathbf{0}$, respectively, we obtain

$$\mathbf{Z} = [\mathbf{X}\mathbf{D}^T + \rho(\mathbf{B}\mathbf{V}^{(t)} + \mathbf{T}_1)](\mathbf{D}\mathbf{D}^T + \rho\mathbf{I})^{-1} \tag{15}$$

and

$$\mathbf{A} = \max(\mathbf{V}^{(t)} + \frac{\mathbf{T}_2}{2\rho} - \frac{\lambda}{2\rho}, 0). \tag{16}$$

Finally, $\mathbf{E}$ can be updated as the following

$$\mathbf{E} = \gamma\mathbf{W}^G\mathbf{V} - (1-\gamma)\mathbf{W}^L\mathbf{V}. \tag{17}$$

The last step includes the update of the Lagrangian multipliers $\mathbf{T}_1$ and $\mathbf{T}_2$ using the following two terms

$$\begin{aligned} \mathbf{T}_1^{(t+1)} &= \mathbf{T}_1^{(t)} + \rho(\mathbf{B}\mathbf{V}^{(t+1)} - \mathbf{Z}^{(t+1)}), \\ \mathbf{T}_2^{(t+1)} &= \mathbf{T}_2^{(t)} + \rho(\mathbf{V}^{(t+1)} - \mathbf{A}^{(t+1)}). \end{aligned} \tag{18}$$

The complete procedure for estimating the HR-HS image is summarized in Algorithm 1.

Algorithm 1 HR-HS estimation with self similarity.

Input: LR-HS image $\mathbf{X}$, HR-RGB image $\mathbf{Y}$, the camera spectral response $\mathbf{R}$, and decimation matrix $\mathbf{D}$.

1: Learn the HS dictionary $\mathbf{B}^*$ from $\mathbf{X}$ via convex optimizing Eq. 8 with ADMM method;
2: Build global-structure groups $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_P\}$ and local-spectral superpixels $\mathcal{L} = \{\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_Q\}$ and calculate the affinity matrix $\mathbf{W}^{\mathcal{G}}$ and $\mathbf{W}^{\mathcal{L}}$ using Eq. 10;
3: Solving Eq. 7 with the fixed dictionary $\mathbf{B}^*$ and the affinity matrix $\mathbf{W}^{\mathcal{G}}$, $\mathbf{W}^{\mathcal{L}}$ via ADMM:
Initialization: Set $\mathbf{A}$, $\mathbf{Z}$, $\mathbf{V}$, $\mathbf{T}_1$, and $\mathbf{T}_2$ to $\mathbf{0}$;
for $t = 0$ to max. iter. **do**
(a) Compute $\mathbf{V}^{(t+1)}$ via Eq. 14;
(b) Compute $\mathbf{Z}^{(t+1)}$ via Eq. 15;
(c) Compute $\mathbf{A}^{(t+1)}$ via Eq. 16;
(d) Update the Lagrangian multiplier $\mathbf{T}_1^{(t+1)}$ and $\mathbf{T}_2^{(t+1)}$via Eq. 18;
(e) Compute $\mathbf{E}^{(t+1)}$ as $\gamma\mathbf{W}^G\mathbf{V}^{(t+1)} - (1-\gamma)\mathbf{W}^L\mathbf{V}^{(t+1)}$;
Terminate: Converged or max. iter. reached;

Output: HR-HS image $\mathbf{Z}$.

Table II
QUANTITATIVE COMPARISON RESULTS USING THE CAVE DATASET. SMALLER RMSE, SAM AND ERGAS MEAN BETTER PERFORMANCE WHILE LARGER PSNR DENOTES BETTER RESULTS.

	RMSE	PSNR	SAM	ERGAS
Matrix Factorization [27]	3.03±1.44	39.37±3.76	6.12±2.17	0.40±0.22
Coupled Non-negative Matrix Factorization [28]	2.93±1.30	39.53±3.55	5.48±1.62	0.39±0.21
Sparse Non-negative Matrix Factorization [29]	3.26±1.57	38.73±3.79	6.50±2.32	0.44±0.23
Generalization of Simultaneous Orthogonal Matching Pursuit [22]	6.47±2.53	32.48±3.08	14.19±5.42	0.77±0.32
Bayesian Sparse Representation [16]	3.13±1.57	39.16±3.91	6.75±2.37	0.37±0.22
Couple Spectral Unmixing [17]	3.0±1.40	39.50±3.63	5.8±2.21	0.41±0.27
Non-Negative Structured Sparse Representation [30]	2.21±1.19	42.26±4.11	4.33±1.37	0.30±0.18
Proposed	**2.17±1.08**	**42.28±3.86**	**3.98±1.27**	**0.28±0.18**

Table III
QUANTITATIVE COMPARISON RESULTS USING THE HARVARD DATASET. SMALLER RMSE, SAM AND ERGAS MEAN BETTER PERFORMANCE WHILE LARGER PSNR DENOTES BETTER RESULTS.

	RMSE	PSNR	SAM	ERGAS
Matrix Factorization [27]	1.96±0.97	43.19±3.87	2.93±1.06	0.23±0.14
Coupled Non-negative Matrix Factorization [28]	2.08±1.34	43.00±4.44	2.91±1.18	0.23±0.11
Sparse Non-negative Matrix Factorization [29]	2.20±0.94	42.03±3.61	3.17±1.07	0.26±0.27
Generalization of Simultaneous Orthogonal Matching Pursuit [22]	4.08±3.55	38.02±5.71	4.79±2.99	0.41±0.24
Bayesian Sparse Representation [16]	2.10±1.60	43.11±4.59	2.93±1.33	0.24±0.15
Couple Spectral Unmixing [17]	1.7±1.24	43.40±4.10	2.9±1.05	0.24±0.20
Non-Negative Structured Sparse Representation [30]	1.76±0.79	44.00±3.63	2.64±0.86	0.21±0.12
Proposed	**1.64±1.20**	**45.20±4.56**	**2.63±0.97**	**0.16±0.15**

V. EXPERIMENT RESULTS

A. Experiment setup

We evaluate the proposed approach using two publicly available hyperspectral imaging database: the CAVE dataset [50] with 32 indoor images including paintings, toys, food, and so on, captured under controlled illumination, and the Harvard dataset [51] with 50 indoor and outdoor images recorded under daylight illumination. The dimensions of the images from the CAVE dataset are 512×512 pixels, with 31 spectral bands of 10 nm width, covering the visible spectrum from 400 to 700 nm; the images from the Harvard dataset have the dimensions of 1392×1040 pixels with 31 spectral bands of width 10 nm, ranging from 420 to 720 nm, from which we extract the top left 1024×1024 pixels in our experiments.

We treat the original images in the databases as ground truth $\mathbf{Z}$, and down-sample them by a factor of 32 to create 16×16 images for CAVE dataset and 32×32 images for Harvard dataset, which is implemented by averaging over 32×32 pixel blocks as done in [22], [27]. The observed HR-RGB images $\mathbf{Y}$ are simulated by integrating the ground truth over the spectral channels using the spectral response $\mathbf{R}$ of a Nikon D700 camera. To evaluate the quantitative accuracy of the estimated HS images, four objective error metrics including root-mean-square error (RMSE), peak-signal-to-noise ratio (PSNR), relative dimensionless global error in synthesis (ERGAS) [52], and spectral angle mapper (SAM) [16] are evaluated. The ERGAS metric [52] calculates the average amount of specific spectral distortion normalized by intensity mean in each band as defined below

$$\mathbf{ERGAS} = 100 \times \frac{N}{M} \sqrt{\frac{1}{L} \sum_{l=1}^{L} (\frac{\mathbf{RMSE}(i)}{\mu_i})}, \qquad (19)$$

where $\frac{N}{M}$ is the ratio between the pixel size of the available HR-RGB and LR-HS images, μ_i is the intensity mean of the i-th band of the LR-HS image, and L is the number of LR-HS bands. A smaller ERGAS denotes smaller spectral distortion. The SAM [16] measures the spectral distortion between the ground-truth and estimated HR-HS images, and the distortion of two spectral vectors $\mathbf{z}_n$ and $\hat{\mathbf{z}}_n$ is defined as follows

$$\mathbf{SAM}(\mathbf{z}_n, \hat{\mathbf{z}}_n) = \arccos(\frac{< \mathbf{z}_n, \hat{\mathbf{z}}_n >}{\| \mathbf{z}_n \|_2 \| \hat{\mathbf{z}}_n \|_2}), \qquad (20)$$

where $\mathbf{z}_n$ denotes the 31-band spectral vector of the n-th pixel in a ground-truth HR-HS image, and $\hat{\mathbf{z}}_n$ is the corresponding spectral vector in the estimated HR-HS image. The overall SAM is finally obtained by averaging the SAMs computed from all image pixels. Note that the value of SAM is expressed in degrees and thus belongs to (-90, 90]. The smaller the absolute value of SAM, the less significant the spectral distortion.

In Table I, we list the values of all adjustable parameters used in our experiments, including the global-structure and local-spectral group numbers P and Q, patch size R for representing spatio-RGB structure, the smoothing kernel width $h^{\mathcal{G}}$ and $h^{\mathcal{G}}$ for the self-similarity constraints as explained in Section V: B, constraint balance ratio γ, the weighting factors λ and η for sparse and self-similarity constraints, and the penalty parameter ρ for convex optimization.

Table IV
RESULTS WITHOUT LOCAL, GLOBAL, AND BOTH SIMILARITIES ON THE CAVE AND HARVARD DATASETS.

	CAVE dataset			Harvard dataset		
	Without both	Local simil. only	Global simil. only	Without both	Local simil. only	Global simil. only
RMSE	2.81±1.42	2.25±1.15	2.32±1.20	1.83±1.30	1.66±1.20	1.78±1.32
PSNR	40.05±3.87	42.00±3.91	41.78±4.05	44.16±4.39	45.01±4.51	44.60±4.56
SAM	5.46 ± 1.89	4.24±1.36	4.59±1.46	2.86±1.06	2.69±1.00	2.79±1.09
ERGAS	0.37±0.20	0.30±0.18	0.31±0.19	0.23±0.16	0.19±0.15	0.18±0.16

B. Comparison with state-of-the-art methods

Firstly, we show the performance of our complete method (including the on-line dictionary learning procedure and self-similarity constraints), compared with state-of-the-art HSI SR methods including: Matrix Factorization method (MF) method [27], Coupled Non-negative Matrix Factorization (CNMF) method [28], Sparse Non-negative Matrix Factorization (SNNMF) method [29], Generalization of Simultaneous Orthogonal Matching Pursuit (GSOMP) method [22], Bayesian Sparse Representation (BSR) method [16], Couple Spectral Unmixing (CSU) method [17], and Non-Negative Structured Sparse Representation (NSSR) method [30]. The average RMSE, PSNR, SAM and ERGAS results of the 32 recovered HR-HS images from the CAVE dataset [50] are shown in Table II, while the average results of the 50 images from the Harvard dataset [51] are given in Table III.

From Tables II and III, we observe that for all error metrics our approach achieves the best performance, and the improvement on the CAVE dataset is in general more significant than on the Harvard dataset. The NSSR method [30] has the closest performance to ours, and both methods show relatively larger advantage over the other methods. Our method shows the most noticeable improvement on SAM values over NSSR [30]. This is because of the facts that, for SAM, a slight spectral distortion of the pixels with small magnitudes affect its value greatly, and that our proposed approach not only robustly recovers the HS image, but also suppresses the artifacts and noise in the original HS image, especially for those pixels with small spectral magnitudes, due to the imposed constraints of the global-structure and local-spectral self-similarities.

C. Results without self-similarity constraints

One of the key difference of our method from existing ones (such as MF [27]) is the two types of self-similarities encoded by $\Omega(\mathbf{A})$ in Eq. 7. We can still recover the HR-HS image $\mathbf{Z}$ by optimizing Eq. 7 without the $\Omega(\mathbf{A})$ term with the ADMM method. Furthermore, we can also apply either global or local self-similarity separately, *i*.e., by considering only the $\mathbf{W}^{\mathcal{G}}$ or $\mathbf{W}^{\mathcal{L}}$ terms in Eq. 11. We perform such experiments using the same error metrics as in Tables II and III for both datasets, and show the results in Table IV. Considering local self-similarity only significantly improves the results on both datasets for all error metrics, which verifies the effectiveness of this self-similarity. However, to further integrate the global self-similarity as in our complete approach could consistently improve the results.

D. Evaluation on Hyperparameters

In addition, we evaluate the HR-HS image recovery performance by changing the parameter γ, which actually adjusts the contribution of global-structure and local-spectral self-similarity while fixing the other parameters as in I. The parameter γ is changed from 0 (local-spectral self-similarity only) to 1 (global-structure self-similarity only) with an interval of 0.1, and apply the same measure metrics for manifesting the contribution of the global and local self-similarity. Figures 3 (a)-(d) give the curves of the quantitative measures: RMSE, PSNR, SAM and ERGAS, respectively, on both CAVE and Harvard datasets, which manifests that $\gamma = 0.3$ gives the best performance and our parameter setting $\gamma = 0.5$ in Table I is very close to the optimal setting.

Furthermore, we also evaluate the reconstruction performance of the HR-HS images by changing one parameter but fixing the contribution balance ratio $\gamma = 0.3$ (the best performance according to Figure 3). All the other parameters are the same as in Table I. We set $\lambda = [0.00001, 0.00005, 0.0001, 0.0005, 0.001]$, $\rho = [0.00001, 0.00005, 0.0001, 0.0005, 0.001]$ and $\eta = [0.015, 0.02, 0.025, 0.03, 0.035]$, respectively, and the average quantitative measures RMSE and ERGAS are shown in Figure 4 for both CAVE and Harvard datasets, which manifests the performance of our algorithm is insensitive to the parameter setting of λ, ρ and η in a wide range.

E. Visual quality comparison

Figures 5, 6, 7 and 8 show the examples of the recovered HS images and the difference images with respect to the ground truth, with two examples from the CAVE and Harvard dataset, respectively. Since in addition to our method, the CSU [17] and NSSR [30] methods show the promising performance compared with all evaluated methods as shown in Tables II and III, we only compare our method with the CSU [17] and NSSR [30] methods for checking the differences in visual quality. It can be seen that the recovered HS images by our approach have smaller absolute difference magnitude for most pixels than the result by the NSSR method. It is also worth noting when self-similarity is not applied, our results show quite similar appearance to those from the NSSR method [30], which in turn reflects the effectiveness of the self-similarity constraint.

VI. CONCLUSIONS

We proposed a global-structure and local-spectral constrained sparse representation for hybrid fusion based HS

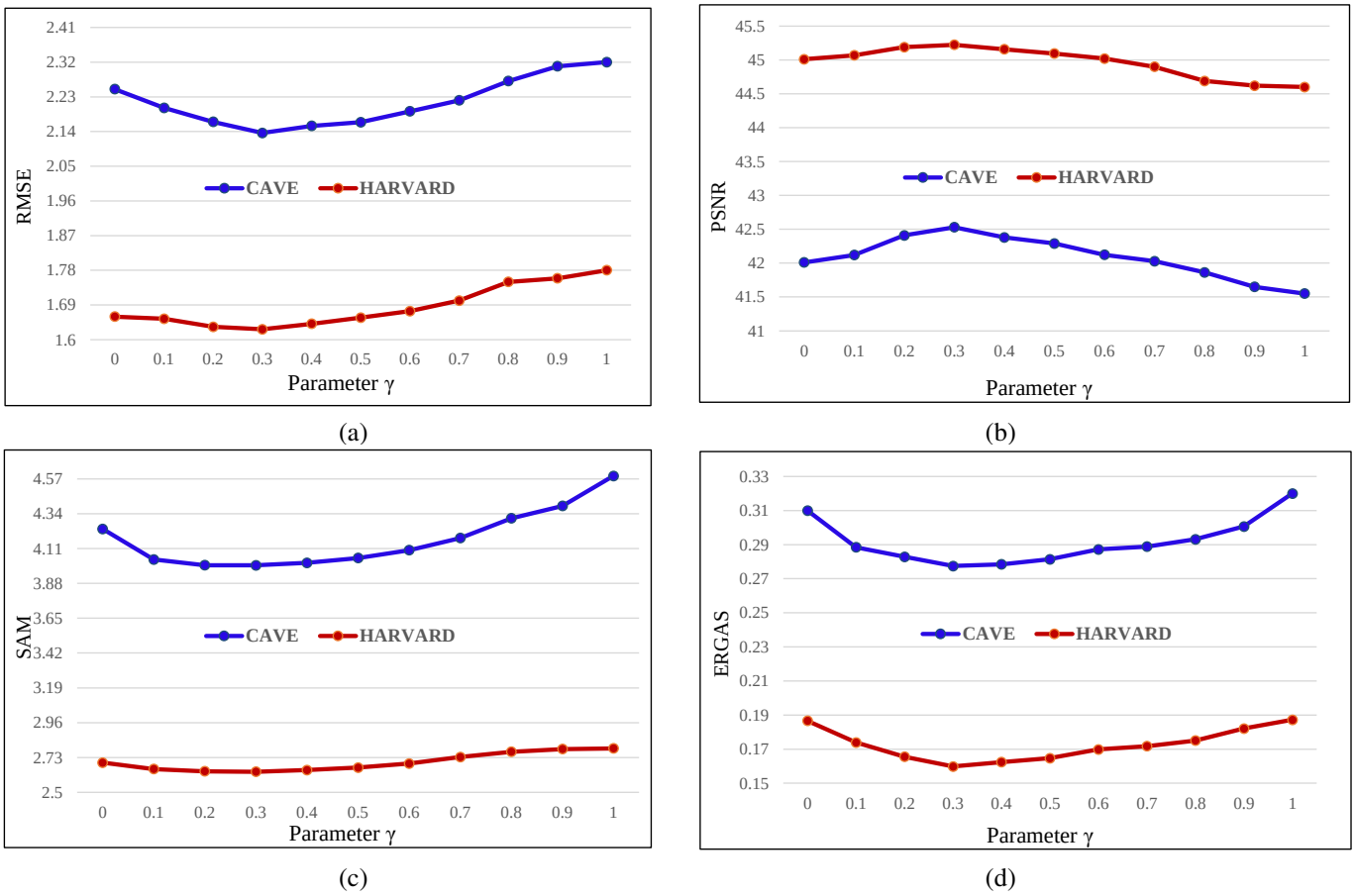


Figure 3. The evaluated performances with different values of the parameter γ on both CAVE and Harvard datasets. (a) RMSE, (b) PNSR, (c) SAM, and (d) ERGAS.

image super-resolution. The proposed approach first learned the HS dictionary online from the input LR-HS image, and then transformed it into RGB dictionary for calculating the sparse representation of all HR pixels in the input HR-RGB image. In order to obtain robust sparse vectors for the HR pixels, we explored the global-structure and local-spectral self-similarity in the HR-RGB image and then constrained similarity of the sparse representation, which is solved by the ADMM method. Experiments on two public HS datasets compared with state-of-the-art methods showed that our proposed approach achieved best performances, and validated that our self-similarity constrained sparse representation can alleviate the effect of outliers in the learned sparse coding.

Although the proposed self-similarity constrained sparse representation outperforms state-of-the-art methods for HSI SR, there are still several prospective research lines for further improvement. The HS dictionary was learned online from the available LR-HS image and fixed in the following convex optimization, so it is guaranteed to be consistent with the representation for the HR-HS image. As future work, we are planning to optimize sparse representation and HS dictionary in an integrated procedure. Currently, we explored the self-similarity without considering spectral constraint, and it is promising to explore the relationship among spectral bands for more robust estimation.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, and P. ScheunScheunders, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 2, pp. 6–36, 2013. 1

[2] M. Fauvel, Y. Tarabalka, J. Benediktsson, J. Chanusssot, and J. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, 2013. 1

[3] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition using 3d-dct and partial least squares," *BMVC*, pp. 57.1–57.10, 2013. 1

[4] D. Zhang, W. Zuo, and F. Yue, "A comparative study of palmprint recognition algorithm," *ACM Comput. Surv.*, vol. 44, no. 1, pp. 2:1–2:37, 2012. 1

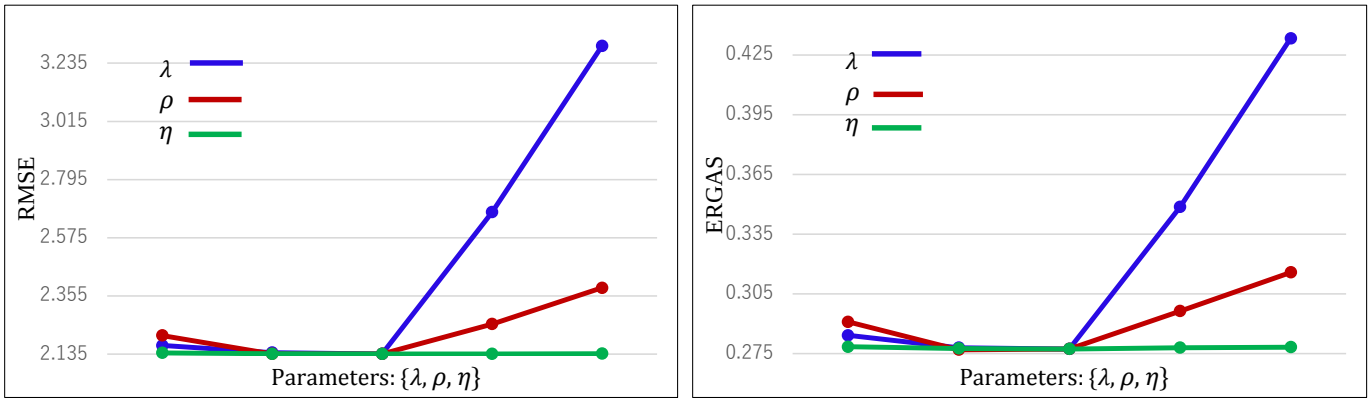
[5] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 50, pp. 809–823, 2012. 1

[6] K. Tan, X. Jin, A. Plaza, X. Wang, L. Xiao, and P. Du, "Automatic change detection in high-resolution remote sensing images by using a multiple classifier dystem and spectral spatial features," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, pp. 3439–3451, 2016. 1

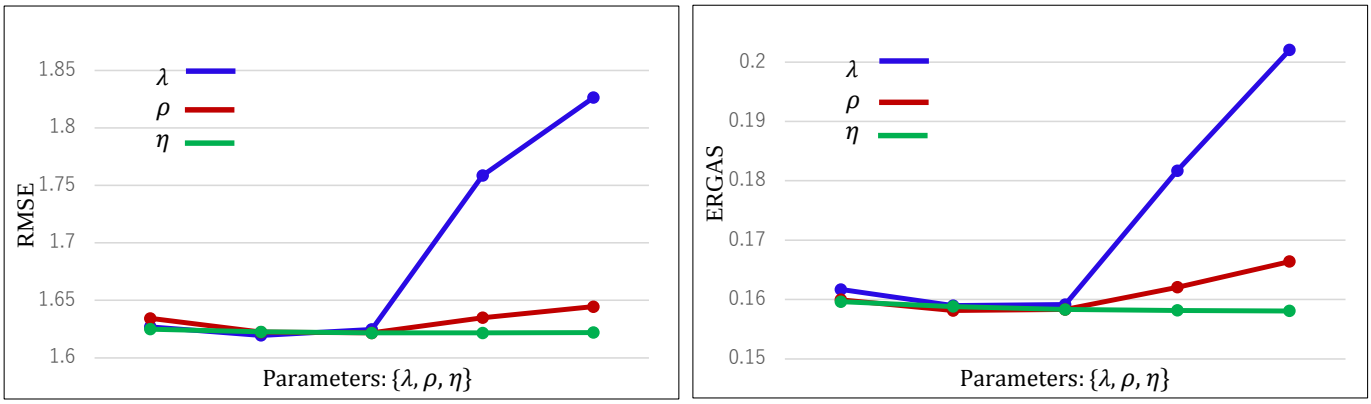
[7] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recogn.*, vol. 63, pp. 371–383, 2017. 1

[8] H. Nguyen, A. Benerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," *CVPRW*, pp. 44–51, 2010. 1

[9] Y. Tarabalka, J. Chanusssot, and J. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 40, no. 5, pp. 1267–1279, 2010. 1



(a) CAVE dataset



(b) Harvard dataset

Figure 4. The evaluated performances with different values of the parameters $\lambda = [0.00001, 0.00005, 0.0001, 0.0005, 0.001]$, $\rho = [0.00001, 0.00005, 0.0001, 0.0005, 0.001]$, $\eta = [0.015, 0.02, 0.025, 0.03, 0.035]$ on both CAVE and Harvard datasets. (a) CAVE dataset and (b) Harvard dataset.

[10] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin, "Classification of histology sections via multispectral convolutional sparse coding," *CVPR*, pp. 3081–3088, 2014. 1

[11] J. 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, 2013. 1

[12] N. Akhtar, F. Shafait, and A. Mian, "Sungp: A greedy sparse approximation algorithm for hyperspectral unmixing," *ICPR*, pp. 3726–3731, 2014. 1

[13] M. Nelson, L. Shi, L. Zbur, R. Priore, and P. Treado, "Real-time shortwave infrared hyperspectral conformal imaging sensor for the detection of threat materials," *Proceedings of the SPIE Defense + Commercial Sensing (DCS) Symposium*, vol. 9824, pp. 1–9, 2016. 1

[14] S. Asadzadeh, C. Roberto, and S. F. De, "A review on spectral processing methodes for geological remote sensing," *Int. J. Appl. Earth Obs. Geoinform.*, vol. 47, pp. 69–90, 2016. 1

[15] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, 2014. 1

[16] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Toureret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015. 1, 2, 6, 7

[17] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," *ICCV*, pp. 3586–3595, 2015. 1, 2, 3, 6, 7, 10, 11, 12, 13

[18] C. Grohnfeldt, X. X. Zhu, and R. Bamler, "Jointly sparse fusion of hyperspectral and multispectral imagery," *IGARSS*, 2013. 1, 2

[19] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," *CVPR*, pp. 3631–3640, 2015. 1, 2

[20] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *NIPS*, pp. 2223–2231, 2009. 1

[21] J. Wang, J. Yang, K. Yu, F. L. T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," *CVPR*, pp. 1–8, 2010. 1

[22] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super resolution," *ECCV*, pp. 63–78, 2014. 1, 2, 6, 7

[23] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006. 1, 2

[24] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007. 2

[25] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012. 2

[26] J. A. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009. 2

[27] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," *CVPR*, pp. 2329–2336, 2011. 2, 6, 7

[28] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization for hyperspectral and multispectral data fusion," *IEEE Trans Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2012. 2, 6, 7

[29] E. Wcoff, T. Chan, K. Jia, W. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," *ICASSP*, pp. 1409–1413, 2013. 2, 4, 6, 7

[30] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative strutured sparse representation,"

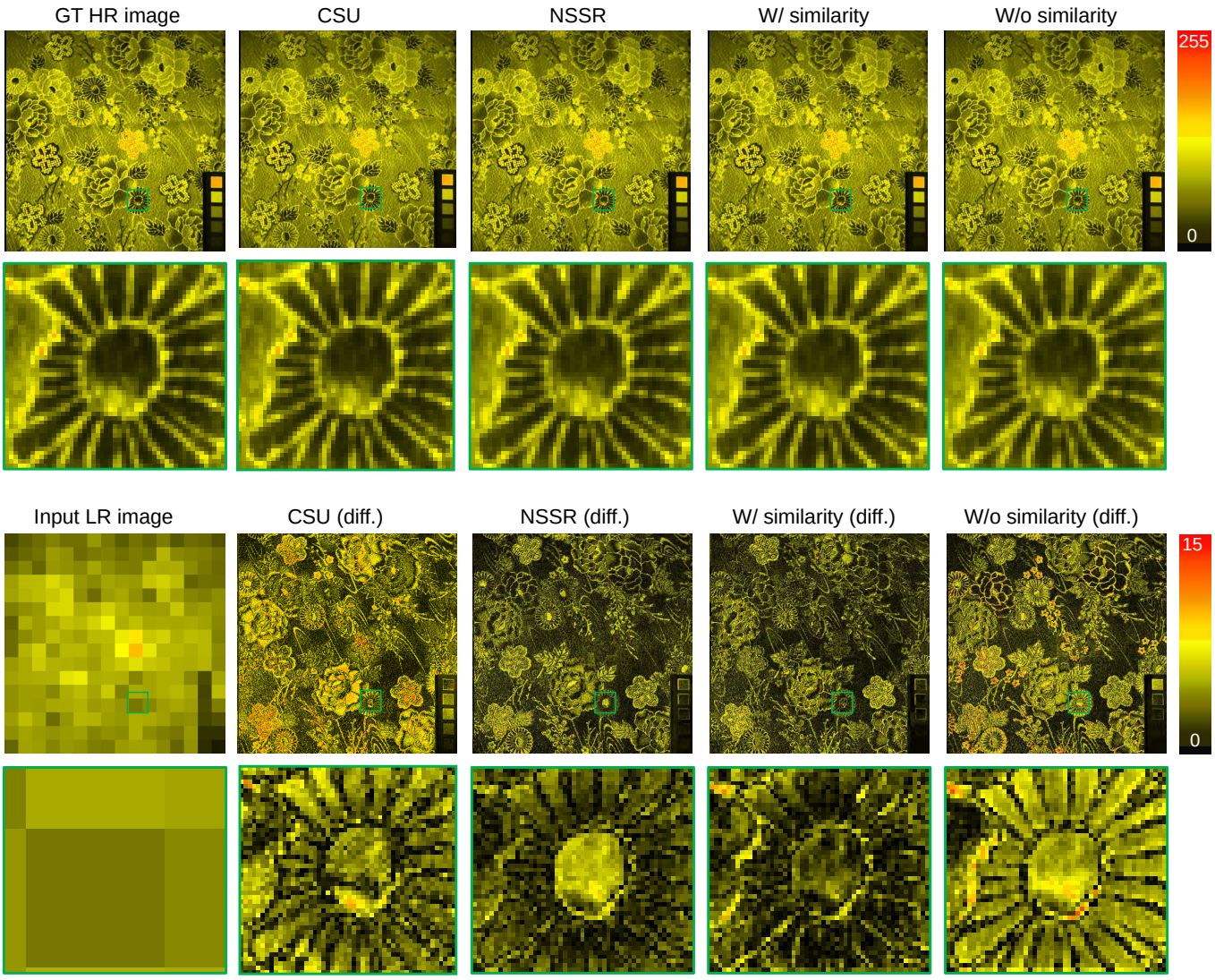


Figure 5. The recovered HR images of 'cloth' image in the CAVE dataset. The first column shows the ground truth HR image and the input LR image, respectively. The second to fifth columns show results from CSU [17], NSSR [30], our method with and without self-similarity, with the upper part showing the recovered images and the lower part showing the absolute difference maps w.r.t. the ground truth. Close-up views are provided below each full resolution image.

IEEE Transaction on Image Processing, vol. 25, no. 3, pp. 2337–2352, 2016. 2, 6, 7, 10, 11, 12, 13

[31] 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, 2014. 2

[32] P. Chavez, S. Sides, and J. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat tm and spot panchromatic," *Photogramm. Eng. Rem. S.*, vol. 30, no. 7, pp. 1779–1804, 1991. 2

[33] R. Haydn, G. Dalke, J. Henkel, and J. Bare, "Application of the ihs color transform to the processing of multisensor data and image enhancement," *Int. Symp on Remote Sens. of Env.*, 1982. 2

[34] 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, 2009. 2

[35] 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, 2006. 2

[36] R. Zurita-Milla, J. 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, 2008. 2

[37] J. Duran, A. Buades, C. Sbert, and G. Blanchet, "A survey of pansharpening methods with A new band-decoupled variational model," *CoRR*, vol. abs/1606.05703, 2016. [Online]. Available: http://arxiv.org/abs/1606.05703 2

[38] K. Kidiyo, C. E.-M. Miloud, and T. Nasreddine, "Recent trends in satellite image pan-sharpening techniques," *1st International Conference on Electrical, Electronic and Computing Engineering*, 2014. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01075703 2

[39] M. Cetin and N. Musaoglu, "Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis," *Int. J. Remote Sens.*, vol. 30, no. 7, pp. 1779–1804, 2009. 2

[40] A. Halimi, J. Bioucas-Dias, N. Dobigeon, G. Buller, and S. McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodelling effects," *Transactions on Computational Imaging*, vol. 3, no. 2, pp. 146–159, 2017. 2

[41] J. Sigurdsson, M. Ulfarsson, J. Sveinsson, and J. Bioucas-Dias, "Sparse distributed multitemporal hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2017. 2

[42] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M.Chen, and S. S. Godsill, "Multi-band image fusion based on spectral unmixing," *IEEE Transactions on Image Processing*, vol. 54, no. 12, pp. 7236–7249,

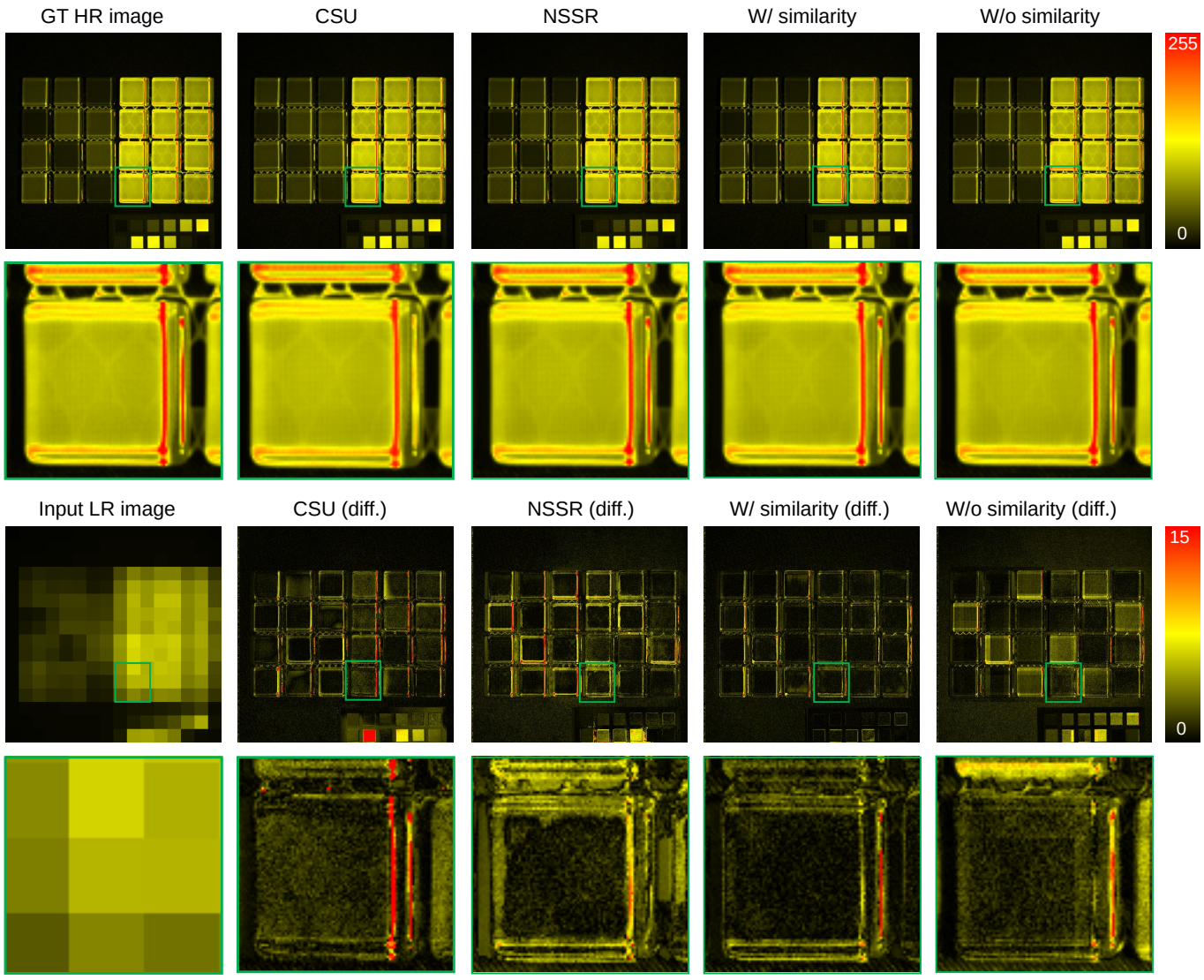


Figure 6. The recovered HR images of 'glass tiles' image in the CAVE dataset. The first column shows the ground truth HR image and the input LR image, respectively. The second to fifth columns show results from CSU [17], NSSR [30], our method with and without self-similarity, with the upper part showing the recovered images and the lower part showing the absolute difference maps w.r.t. the ground truth. Close-up views are provided below each full resolution image.

2016. 2

[43] X. Fu, W.-K. Ma, J. Bioucas-Dias, and T.-H. Chan, "Semiblind hyperspectral unmixing in the presence of spectral library mismatches," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5171–5184, 2016. 2

[44] D. D. Lee and S. H. Seung, "Algorithms for non-negative matrix factorization," *NIPS*, pp. 556–562, 2001. 2

[45] J. Li, "Sparse representation based single image super-resolution with low-rank constraint and nonlocal self-similarity," *Multimedia Tools and Applications*, vol. 77, pp. 1693–1714, 2018. 2

[46] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012. 3

[47] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transaction on Pattern Analysis and Machine Intrlligence*, vol. 34, no. 11, pp. 2274–2282, 2012. 4

[48] A. David and S. Vassilvitskii, "K-means++: The advantages of careful seeding," *The Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA2007)*, pp. 1027–1035, 2007. 4

[49] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels," *EPFL Technical Report*, no. 149300, 2010. 4

[50] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum," *IEEE Transaction on Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010. 6, 7

[51] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," *CVPR*, pp. 193–200, 2011. 6, 7

[52] L. Wald, "Quality of hige resolution synthesisted images: Is there a simple criterion?" *Proc. of Fusion Earth Data*, pp. 99–103, 2000. 6

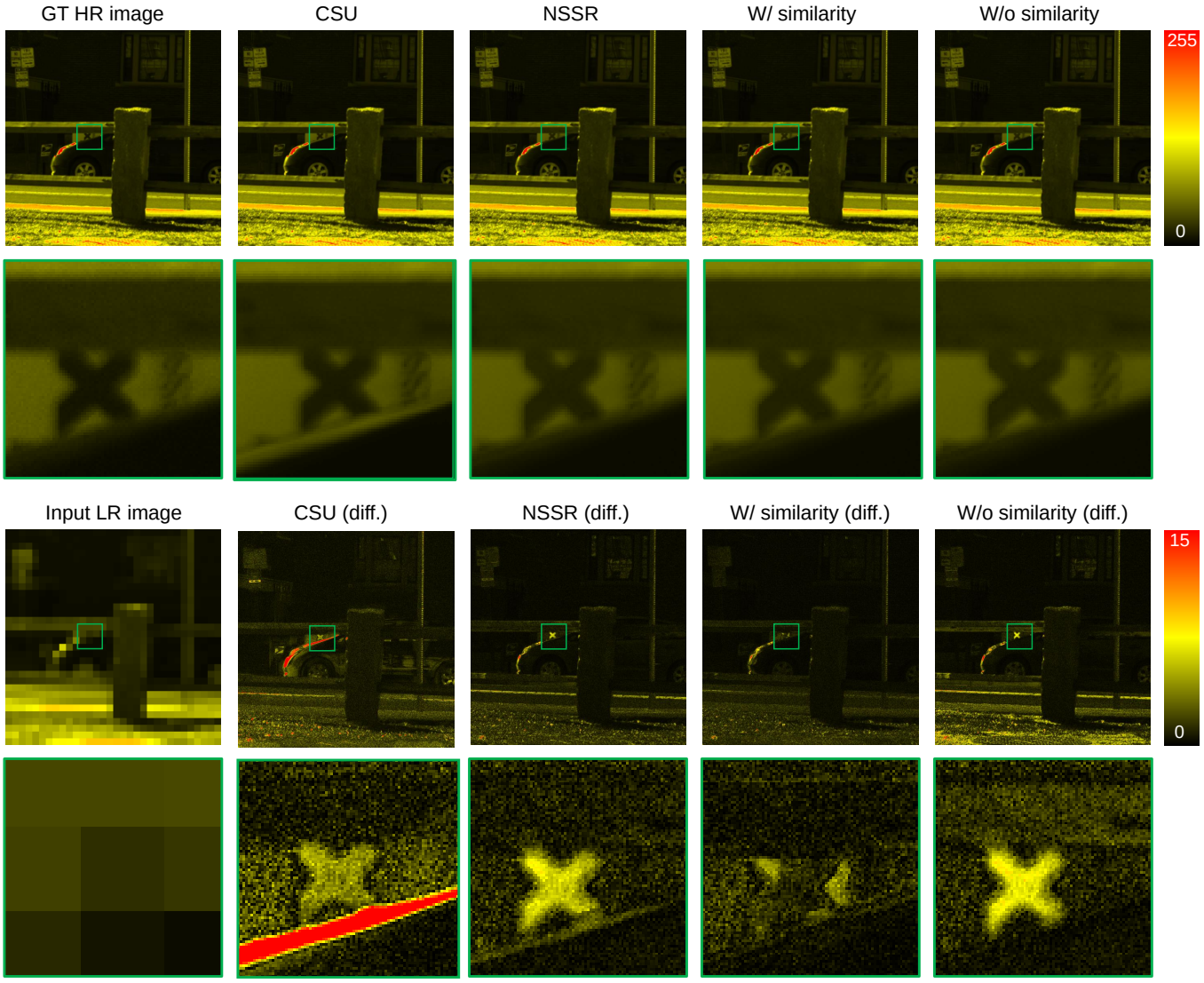


Figure 7. The recovered HR image of 'imgf1' image in the Harvard dataset. The first column shows the ground truth HR image and the input LR image, respectively. The second to fifth columns show results from CSU [17], NSSR [30], our method with and without self-similarity, with the upper part showing the recovered images and the lower part showing the absolute difference maps w.r.t. the ground truth. Close-up views are provided below each full resolution image.

Xian-Hua Han (M'11) received the B.E. degree from Chongqing University, Chongqing, China, the M.E. degree from Shandong University, Jinan, China, and the D.E. degree from the University of Ryukyus, Japan, in 2005. From Apr. 2007 to Mar. 2016, she was a post-doctoral fellow and an associate professor with the College of Information Science and Engineering, Ritsumeikan University, Japan, and from Apr. 2016 to Feb. 2017, was a senior researcher at the Artificial Intelligence Researcher Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan. From Mar. 2017, she is now an associate professor at Yamaguchi University, Japan. Her current research interests include image processing and analysis, feature extraction, machine learning, computer vision, and pattern recognition. She is a member of the IEEE, IEICE.

Boxin Shi received the BE degree from the Beijing University of Posts and Telecommunications, the ME degree from Peking University, and the PhD degree from the University of Tokyo, in 2007, 2010, and 2013. He is currently an assistant professor (1000 Youth Talents Professorship) with Peking University, where he leads the Camera Intelligence Group. Before joining PKU, he did postdoctoral research with MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University from 2013 to 2016, and worked as a researcher in the National Institute of Advanced Industrial Science and Technology from 2016 to 2017. He won the Best Paper Runner-up award at International Conference on Computational Photography 2015. He has served as Area chairs for ACCV 2018, MVA 2017. He is a member of the IEEE.

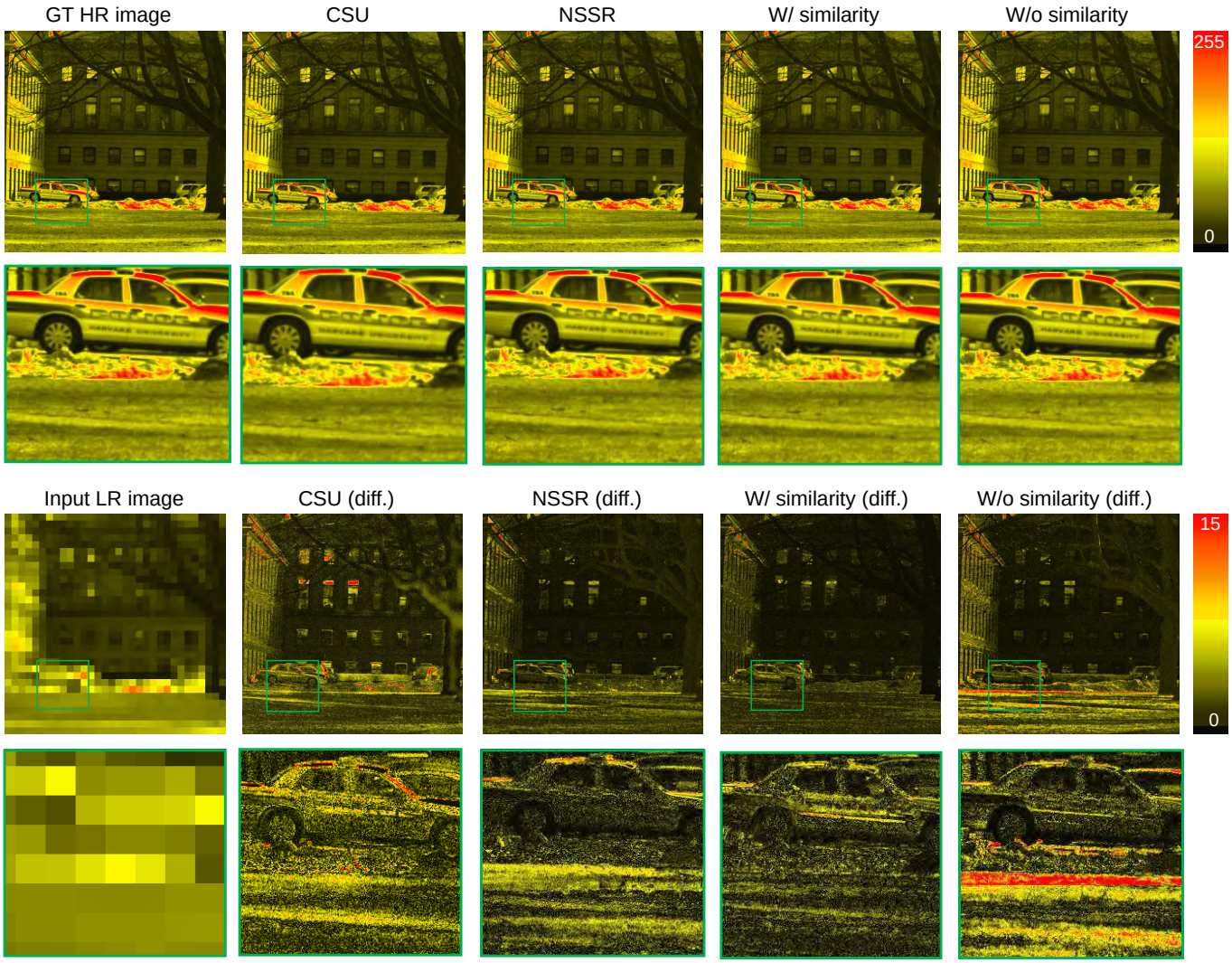


Figure 8. The recovered HR image of 'imgb3' image in the Harvard dataset. The first column shows the ground truth HR image and the input LR image, respectively. The second to fifth columns show results from CSU [17], NSSR [30], our method with and without self-similarity, with the upper part showing the recovered images and the lower part showing the absolute difference maps w.r.t. the ground truth. Close-up views are provided below each full resolution image.

YinQiang Zheng received his Bachelor degree from the Department of Automation, Tianjin University, Tianjin, China, in 2006, Master degree of engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, and Doctoral degree of engineering from the Department of Mechanical and Control Engineering, Tokyo Institute of Technology, Tokyo, Japan, in 2013. He is currently an assistant professor in the National Institute of Informatics, Japan. His research interests include image processing, computer vision and mathematical optimization.