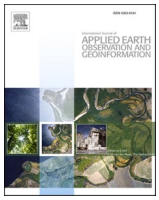
International Journal of Applied Earth Observations and Geoinformation 104 (2021) 102570

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Fusion of Hyperspectral-Multispectral images joining Spatial-Spectral Dual-Dictionary and structured sparse Low-rank representation

Nan Chen [a], Lichun Sui [a,*], Biao Zhang [b], Hongjie He [d], Kyle Gao [d], Yandong Li [a], José Marcato Junior [c], Jonathan Li [d,*]

[a] College of Geological Engineering and Geomatics, Chang'an University, Xi'an SX710054, China
[b] Geovis Spatial Technology Co., Ltd, Xi'an SX710199, China
[c] Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, MS 79070-900, Brazil
[d] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Hyperspectral super-resolution
Structured sparse low-rank
Spectral dictionary
Spatial dictionary
Superpixel segmentation

ABSTRACT

High spatial resolution hyperspectral images (HR-HSIs) have shown considerable potential in urban green infrastructure monitoring. A prevalent scheme to overcome spatial resolution limitations in HSIs is by fusing low-resolution hyperspectral images (LR-HSIs) and high-resolution multispectral images (HR-MSIs). Existing methods considering the spectral dictionary or spatial dictionary can only reflect the unilateral characteristics of the HSI and cannot completely restore full information in the latent HSI. To overcome this issue, we propose a novel HSI-MSI fusion method, named DDSSLR, which joins spatial-spectral dual-dictionary and structured sparse low-rank representation. The spectral dictionary characterizing generalized spectra and the corresponding spectral sparse coefficients are extracted from LR-HSI and HR-MSI, while sparse low-rank priors of the local structure are imposed on the spectral pixels within the same superpixel in HR-MSI. Additionally, in the spatial domain, we exploit the remaining high-frequency components to learn the spatial dictionary and use the unitary transformation to factorize the spatial sparse coefficient into the sparse low-rank matrix in subspace, establishing the relationship between low-rank and sparse. We formulate the two fusion models as variational optimization problems, which are effectively solved by the alternating direction methods of multipliers (ADMM). Experiments on three HSI datasets show that DDSSLR achieves state-of-the-art performance.

1. Introduction

The increasing development of modern urbanization has brought new challenging issues about the monitoring and management of urban expansion and urban green spaces. High-resolution satellite images can quickly discover the speed of urban expansion and the species distribution of urban green spaces (Rougier et al., 2016), and hyperspectral satellite images can quickly identify species of urban green spaces (Modzelewska et al., 2020; Shi et al., 2018) and land cover types in urban areas (Chen et al., 2018). Therefore, HR-HSIs have shown considerable potential in urban green infrastructure monitoring. Since the HSIs satisfy spectral continuity and spectral separability, they have the strong spectral diagnostic capability to distinguish materials, which facilitate image classification (Abbas et al., 2021), object detection (Noomen et al., 2015), and band selection (Long et al., 2019). The traditional HSIs contain rich and successive narrow bands in the same scene, but an acceptable signal-to-noise ratio for each spectral window requires long exposures to the imaging system to collect enough photons, which leads to the inevitable trade-off between the resolutions of the spatial-spectral modes. Therefore, fusing HSI having low spatial but high spectral resolution with MSI having high spatial but low spectral resolution over the same scene is a super-resolution (SR) method to obtain HR-HSI, which will play an important role in the construction and monitoring of urban green space.

Recently, there are roughly four categories of the HSI-SR methods. Spectral unmixing-based methods unmix the latent HR-HSI into underlying pure endmembers matrix and HR abundance matrix. For example, the coupled nonnegative matrix factorization (CNMF) method (Yokoya et al., 2012) alternately unmixes each image base on the linear mixture model to generate a new HR-HSI. The joint super-resolution and unmixing method (Lanaras et al., 2017) jointly unmixes the two input images into pure reflectance spectra and the associated mixing

* Corresponding authors.
E-mail addresses: sui1011@chd.edu.cn (L. Sui), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.jag.2021.102570
Received 17 July 2021; Received in revised form 14 September 2021; Accepted 1 October 2021
Available online 20 October 2021

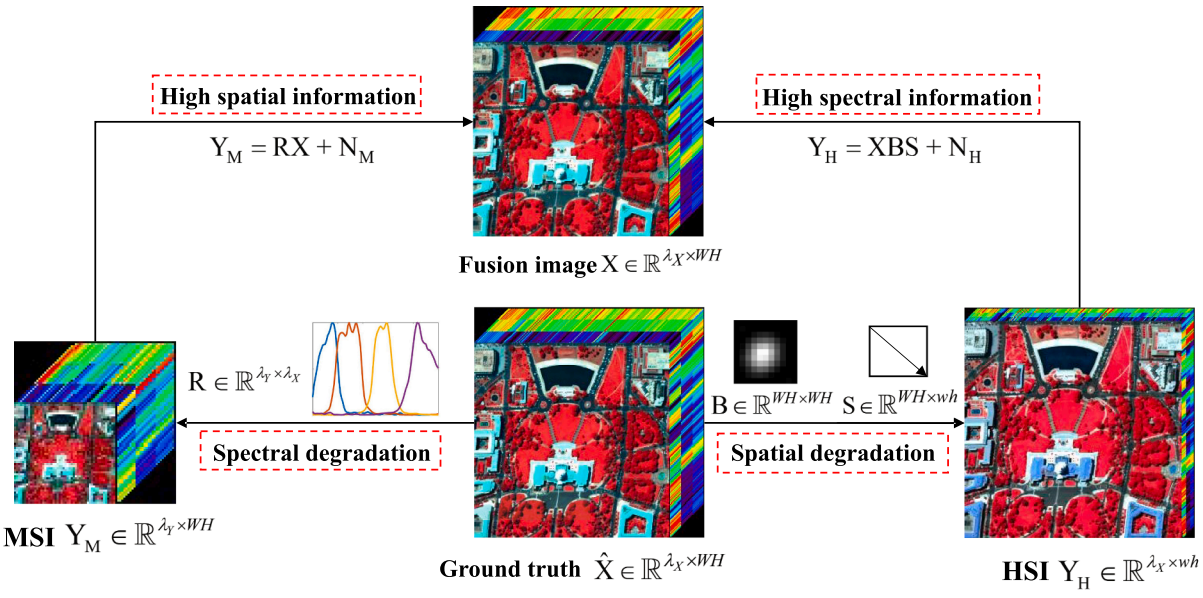


Fig. 1. Two degradation models in fusion task.

coefficients by a coupled matrix factorization with efficient proximal mappings. Sparsity representation-based methods reconstruct the latent HR-HSI by the learned redundant dictionary with the relative sparse coefficients. For example, a generalization of the Simultaneous Orthogonal Matching Pursuit (G-SOMP+) algorithm (Akhtar et al., 2014) estimates the sparse coefficients of each local patch by jointly exploiting non-negativity and similarities among the spectral pixels. Non-parametric Bayesian sparse representation scheme (BSR) (Akhtar et al., 2015) infers the probability distributions and proportions for the material spectra in the scene and uses Bayesian sparse coding strategy to infer the sparse coding coefficient. A novel superpixel-based sparse representation (SSR) scheme (Fang et al., 2018) employs a joint sparse regularization on the pixels within the same superpixel to improve the sparse decomposition. Tensor factorization-based methods represent an HSI image as a 3D tensor and reconstruct the latent HR-HSI via the multiplication of the dictionary and the core tensor. For example, the non-local sparse tensor factorization (NLSTF) method (Dian et al., 2017) introduces tensor factorization to HSI super-resolution and approximates the HSI cube by core tensor multiplication by dictionaries of three modes base on Tucker decomposition. A new low tensor-train rank representation (LTTR) (Dian et al., 2019) further reconstructs the HSI by grouping the similar cubes as a 4-D tensor and imposes the LTTR prior to grouped 4-D tensors. Deep learning-based methods use powerful feature extraction, often achieving high fusion quality but resulting in high time complexity. For example, a novel spatial and spectral fusion convolutional neural network (SSRCNN) (Han et al., 2018) improves the spatial resolution of the input HSI by concatenating the spatial feature of HR-RGB and the spectral feature of LR-HSI. A two-stream fusion network (TFNet) (Liu et al., 2020) can accomplish pan-sharpening in the CNN feature domain, producing promising pan-sharpened qualities.

Among them, the sparse representation-based methods have a wide range of applications, since they are concise and do not require additional training datasets. Although the existing methods have achieved great performance for the HSI-MSI fusion task, the efficient passing of information between spatial and spectral modes is still a challenge, which significantly impacts the fusion quality (Yi et al., 2018). For example, the HSI-MSI fusion methods (Ghasrodashti et al., 2017) only using a spatial dictionary can exploit detailed structural information, but they cause severe spectral distortion. On the contrary, the methods (Dian et al., 2018; Sui et al., 2019) that only consider the spectral dictionary often ignore the high-frequency spatial features. To preserve both spectral and spatial information simultaneously, the method (Nezhad et al., 2016) combines the spectral dictionary constructed by the spectral mixing model with sparse coding to inject HR information from a spatial dictionary of unrelated HR images. Typically, a fusion method with optimized twin dictionaries (OTD) is proposed (Han et al., 2020). It can fully integrate the spectral feature and the spatial features into the latent HSI, but it rarely incorporates the prior information of the raw image into the fusion model, so its fusion performance still has a large room for improvement. The HSIs possess abundant redundancy, that is, spectral global correlation and spatial nonlocal similarity, which have significantly improved HSI restoration methods (Chen et al., 2020; Xue et al., 2019). Low-rank representation (LRR) projects high-dimensional signals into a lower-dimensional subspace, then uses a sparse linear combination of dictionary atoms to recover the underlying structure hidden in the original data (Pan et al., 2018; Yang et al., 2020). Since there are the spectral correlations along with successive spectral bands and the spatial correlations among spatial nonlocal similarity patches in HSIs, the design of a rational LRR constraint to represent such correlations is key to regularizing the ill-posed problem of the HSI-MSI fusion task.

To this end, we recast the optimized twin dictionaries model used in the OTD method and propose a novel HSI-MSI fusion method, named DDSSLR, by joining spatial-spectral dual-dictionary fusion with relevant structured sparse low-rank representation. On one hand, a spectral dictionary characterizing generalized spectrum is learned from the two input images in the spectral domain. Given that spectral vectors selected neighborhoods represent the same materials, we employ a shape adaptive superpixel strategy to segment the HR-MSI and impose the low-rank prior of the local structure to the sparse coefficients of the spectral pixels within the same superpixel. Here, the variational regularization model of the regression coefficients for the spectral dictionary has been derived theoretically, for which the solutions are efficiently given by split augmented Lagrangian shrinkage algorithm (SALSA) (Afonso et al., 2010). On the other hand, we utilize the remaining high-frequency information that is unrepresented by the spectral dictionary to learn the spatial dictionary. Moreover, we impose the structured sparse constraint into the low-rank subspace to characterize the relationship between the low-rankness and sparsity of the regression coefficients in the spatial domain. Then, the variational regularization model of the regression coefficients for the spatial dictionary is derived theoretically, for which the solutions are efficiently given by ADMM (Boyd et al., 2011). The

three major advantages and innovations of this paper include:

(1) A novel structured sparse low-rank representation strategy for the HSI-MSI fusion task via joining spatial-spectral dual-dictionary is devised to construct the fusion models in both the spatial and spectral domain;
(2) In the spectral domain, an efficient shape adaptive superpixel strategy is exploited to segment the HR-MSI, and the low-rank prior of the local structure is imposed to the sparse coefficients of the spectral pixels within the same superpixel;
(3) In the spatial domain, an efficient low-rank structure constraint for the sparse subspace of the regression coefficients is exploited to characterize the inherent relationship between low-rankness and sparsity.

2. Method

2.1. Problem formulation

We consider the fusion task as a post-processing step which can be interpreted as the inverse problem of the forward models with instrument knowledge priors, where the input LR-HSI is supposed to be a blurred and down-sampled version of the latent HR-HSI in the spatial domain, while the HR-MSI image is viewed as a spectrally degraded version of the latent HR-HSI in the spectral domain, as illustrated in Fig. 1.

For representation convenience, $\mathrm{X} \in \mathbb{R}^{\lambda_X \times WH}$ denotes the targeted HR-HSI with $W \times H$ spatial resolution and λ_X spectral bands, $\mathrm{Y_M} \in \mathbb{R}^{\lambda_Y \times WH}$ denotes the available HR-MSI with λ_Y spectral bands ($\lambda_Y \ll \lambda_X$) and $\mathrm{Y_H} \in \mathbb{R}^{\lambda_X \times wh}$ denotes the available LR-HSI with $w \times h$ spatial resolution ($w \ll W, h \ll H$), let $n = w \times h, N = W \times H$. $\widehat{X} \in \mathbb{R}^{\lambda_X \times WH}$ denotes the ground truth as the reference image. The two forward observation degraded models can be formulated as follows

$$\begin{aligned} \mathrm{Y_H} &= \mathrm{XBS} + \mathrm{N_H} \\ \mathrm{Y_M} &= \mathrm{RX} + \mathrm{N_M} \end{aligned} \tag{1}$$

where $\mathrm{B} \in \mathbb{R}^{N \times N}$ denotes a block circulant convolution; $\mathrm{S} \in \mathbb{R}^{N \times n}$ denotes a spatial down-sampling operator; $\mathrm{R} \in \mathbb{R}^{\lambda_Y \times \lambda_X}$ represents the spectral degradation operator; $\mathrm{N_H} \in \mathbb{R}^{\lambda_X \times n}$ and $\mathrm{N_M} \in \mathbb{R}^{\lambda_Y \times N}$ represent noise matrices. It is assumed that the HR-HSI X can be represented as the linear combination of a small number of distinct spectral signatures,

$$\begin{aligned} \mathrm{X} &= \mathrm{D_S A} + E \\ \mathrm{E} &= \mathrm{D_P}^{\circ}\, \Psi \end{aligned} \tag{2}$$

where $\mathrm{D_S} \in \mathbb{R}^{\lambda_X \times K_S}$ denotes spectral dictionary, K_S is the size of $\mathrm{D_S}$, $\mathrm{A} \in \mathbb{R}^{K_S \times N}$ denotes the spectral sparse coefficients, $\mathrm{E} \in \mathbb{R}^{\lambda_X \times N}$ is the error component that cannot be represented by $\mathrm{D_S}$. The remaining spatial information in E can be further recovered by a spatial dictionary $\mathrm{D_P} \in \mathbb{R}^{B_P \times K_P}$ and the relevant spatial sparse coefficients $\Psi \in \mathbb{R}^{K_P \times N_P}$, where K_P is the size of $\mathrm{D_P}$, N_P is the number of the image patches. The symbol "$\circ$" represents the equivalent expression of multiplication in the image patch domain (Zhang et al., 2019).

By substituting Eq. (2) into Eq. (1), the fusion degradation observation model can be reformulation as

$$\begin{aligned} \mathrm{Y_H} &= \mathrm{D_S ABS} + \mathrm{E_H} \\ \mathrm{Y_M} &= \mathrm{RD_S A} + \mathrm{E_M} \\ \varepsilon_H &= P\mathrm{D_P}\Psi + n_H \\ \varepsilon_M &= \mathrm{D_P}\Psi \mathrm{R}_P + n_M \end{aligned} \tag{3}$$

where ε_H and ε_M are the stretched patches of the error matrixes $\mathrm{E_H}$ and $\mathrm{E_M}$, respectively. P, R_P, n_H, n_M are the equivalent expressions of $\mathrm{H} = \mathrm{BS}$, R, $\mathrm{N_H}$, $\mathrm{N_M}$ in image patch domain, respectively.

Based on the reformulation of the forward observation degraded models associated with the MSI-HSI fusion, the MSI-HSI fusion problem in this paper can be formulated as follows

$$\underset{\mathrm{D}_s, \mathrm{A}}{\operatorname{argmin}} \|\mathrm{Y_H} - \mathrm{D}_S ABS\|_F^2 + \eta \|\mathrm{Y_M} - R\mathrm{D}_S \mathrm{A}\|_F^2 + \lambda_1 \phi(\mathrm{A}) \tag{4}$$

$$\underset{\mathrm{D}_P, \Psi}{\operatorname{argmin}} \|\varepsilon_H - P\mathrm{D}_P \Psi\|_F^2 + \beta \|\varepsilon_M - \mathrm{D}_P \Psi R_P\|_F^2 + \lambda_2 \phi(\Psi) \tag{5}$$

where $\phi(\mathrm{A})$ denotes the regularization term derived from $\mathrm{Y_H}$ and $\mathrm{Y_M}$, while $\phi(\Psi)$ denotes the regularization term derived from ε_H and ε_M. $\eta > 0, \beta > 0$ are the trade-off parameters to balance the spatial error and the spectral error, respectively. λ_1, λ_2 are the regularization parameters. Eq. (4) and Eq. (5) are the fusion model in the spectral and spatial domain, respectively. In the following subsections, we will introduce the specific forms of $\phi(\mathrm{A})$ and $\phi(\Psi)$, which are introduced into the fusion models to ensure stable estimation for A and Ψ.

2.2. Spectral dictionary and sparse coefficient optimization

As is summarized in Algorithm 1, we derive theoretically the optimization process of the spectral dictionary $\mathrm{D_S}$ and spectral sparse coefficients A via using the spectral information provided by the observed image $\mathrm{Y_H}$ and $\mathrm{Y_M}$.

2.2.1. Spectral dictionary optimization

Since the urban green space scenes covered by the HSI often contain different materials, their reflectivity changes greatly, and the use of a common HSI dictionary cannot fully represent the spectral information of different materials, which tends to give considerable spectral distortion. The spectral dictionary $\mathrm{D_S}$ in this paper can be considered as a more generalized spectral signature, which may better represent the scene spectra. In this subsection, this paper firstly uses the K-SVD method to initialize $\mathrm{D_S}$, and then uses the biconvex iteration optimization strategy proposed in the OTD method (Han et al., 2020) to iteratively estimate $\mathrm{D_S}$ according to the following augmented Lagrangian function:

$$\mathscr{L}\ominus(\mathrm{D_S}, \mathrm{Z}, \mathrm{V_1}) = \|\mathrm{Y_H} - \mathrm{D_S A_H}\|_F^2 + \eta \|\mathrm{Y_M} - \mathrm{RZ}\|_F^2 + \mu \left\| \mathrm{D_S A} - \mathrm{Z} + \frac{\mathrm{V_1}}{2\mu_1} \right\|_F^2 \tag{6}$$

where $\mathrm{A_H} = \mathrm{ABS} \in \mathbb{R}^{K \times n}, \mathrm{V_1} \in \mathbb{R}^{\lambda_X \times N}$ is the Lagrangian multiplier, $\mathrm{Z} = \mathrm{D_S A}$ is the splitting variable, $\mu > 0$ is the Lagrangian parameter. Then, solving Eq. (6) will consist of the following alternative iterations:

$$\begin{aligned} &\operatorname{argmin}_{\mathrm{D_s}} \left\| \mathrm{Y_H} - \mathrm{D}_s^{(k)} \mathrm{A_H} \right\|_F^2 + \mu \left\| \mathrm{D}_s^{(k)} \mathrm{A} - \mathrm{Z} + \frac{\mathrm{V}_1^{(k)}}{2\mu} \right\|_F^2 \\ &\operatorname{argmin}_Z \eta \left\| \mathrm{Y_M} - \mathrm{RZ}^{(k+1)} \right\|_F^2 + \mu \left\| \mathrm{D_s A} - \mathrm{Z}^{(k+1)} + \frac{\mathrm{V}_1^{(k)}}{2\mu} \right\|_F^2 \\ &\mathrm{V}_1^{(k+1)} = \mathrm{V}_1^{(k)} + \mu \left(\mathrm{D}_s^{(k+1)} \mathrm{A} - \mathrm{Z}^{(k+1)} \right) \end{aligned} \tag{7}$$

The above optimization problems in Eq. (7) admit closed-form solutions, namely:

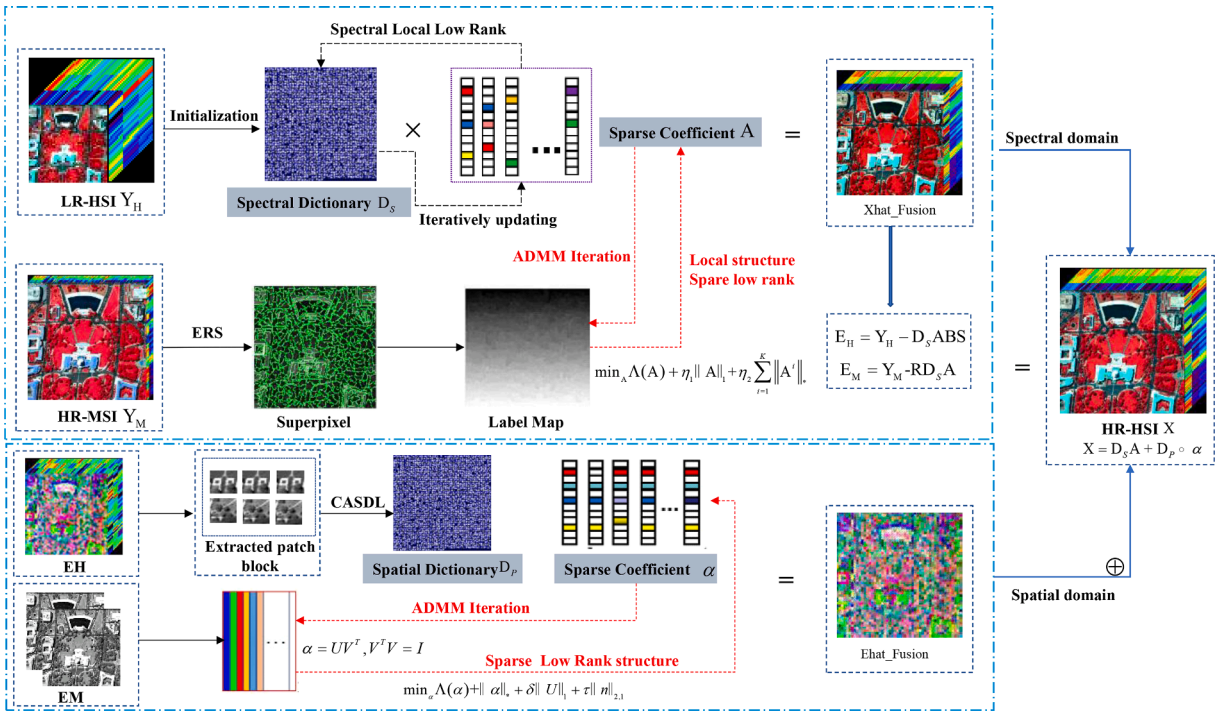


Fig. 2. The proposed DDSSLR framework.

$$Z = \left(\eta R^T R + \mu I\right)^{-1}\left[\eta R^T Y_M + \mu\left(D_S A + \frac{V_1}{2\mu}\right)\right]$$
$$D_S = \left[Y_H A_H^T + \mu\left(Z - \frac{V_1}{2\mu}\right)A^T\right]\left(A_H A_H^T + \mu A A^T\right)^{-1} \quad (8)$$

2.2.2. Spectral sparse coefficient optimization

The spectral vectors of the HSI in selected neighborhoods often represent identical materials; they will provide a more promising reconstruction result by simultaneously encoding the pixels within a local fixed window than the pixel-wise sparse representation method which often ignores the similarity of neighbor pixels. Since HR-HSI preserves the same spatial information as HR-MSI counterparts, we try to sufficiently utilize the spatial information of MSI to search the local windows. The pixels in the adaptive windows are assumed to have similar spectral characteristics, so we can impose a constraint formulation based on locally low-rank to regularize the MSI-HSI fusion model in spectral-domain to achieve more robust performance. We firstly apply the entropy rate-based superpixel segmentation (ERS) method (Liu et al., 2011) to over-segment the HR-MSI into K superpixels $\{Y_M^i\}_{i=1}^K$. According to the corresponding superpixels segmentation maps, the latent HR-HSI is also segmented into K superpixels $\{X^i = D_S A^i\}_{i=1}^K$. We impose a constraint formulation based on local low-rank prior for the spectral pixels' sparse coefficients within the same superpixel. The nuclear norm is applied to promote local low-rankness in this paper. The obtained fusion model Eq. (4) can be specifically formulated as follows

Algorithm 1. (*The spectral dictionary and spectral sparse coefficient optimization algorithm*)

Input: Y_H, Y_M, B, S, R, regularization parameters η_1,η_2, superpixel number K, auxiliary parameter η, $\mu = \mu_1 = 10^{-3}$
Outer loop: For $k = 0, 1, \ldots, T_S$ do
Step 1: Spectral dictionary optimization
Inner loop: (Solving (6) via ADMM): **While not converged do**
(a) Compute Z via Eq. (8);
(b) Compute D_S via Eq. (8);
End while

(*continued on next column*)

(*continued*)

Step 2: Spectral sparse coefficient optimization
Superpixel segmentation
(a) over segment the HR-MSI into K superpixels $\{Y_M^i = \}_{i=1}^K$,
(b) over segment the HR-HSI into K superpixels, $\{X^i = D_S A^i\}_{i=1}^K$
Inner loop:(Solving (10) via SALSA): **While not converged do**
(a) Compute A via Eq. (11);
(b) Compute V_1, V_2, V_3, V_4 via Eq. (12)-Eq. (15);
(c) Compute G_1, G_2, G_3, G_4 via Eq. (16);
End while
End for
Output: Spectral dictionary D_S and spectral sparse coefficient A

$$\underset{A}{\operatorname{argmin}} \|Y_H - D_S ABS\|_F^2 + \eta\|Y_M - RD_S A\|_F^2 + \eta_1\|A\|_1 + \eta_2\sum_{i=1}^{K}\|A^i\|_* \quad (9)$$

where $\|A\|_1$ denotes the sparse constraint on the coefficient matrix; $\|A^i\|_*$ is the nuclear norm of the coefficients in i^{th} superpixel which promotes local low-rankness; $\eta_1 \geqslant 0$ denotes sparse regularization parameters, $\eta_2 \geqslant 0$ denotes low-rank regularization parameters. Eq. (9) is convex and can be efficiently solved via SALSA. By introducing four splitting variables $V_1 = AB$, $V_2 = A$, $V_3 = A$ and $V_4 = A$, we can obtain the following augmented Lagrangian function:

$$\begin{aligned}\mathscr{L}_\ominus = {} & \|Y_H - D_S V_1 S\|_F^2 + \mu_1\|V_1 - AB + G_1\|_F^2 + \eta\|Y_M - RD_S V_2\|_F^2 + \mu_1\|V_2 \\ & - A + G_2\|_F^2 + \eta_1\|V_3\|_1 + \mu_1\|V_3 - A + G_3\|_F^2 + \eta_2\sum_{i=1}^{K}\|V_4^i\|_* + \mu_1\|V_4 \\ & - A + G_4\|_F^2\end{aligned} \quad (10)$$

where $\mu_1 > 0$ is the Lagrangian parameter. Eq. (10) can be divided into the following subproblems, in which all variables are iteratively updated with the others fixed:

(1) Update A = $\operatorname{argmin}_A \mathscr{L}$:

$$A = [(V_1 + G_1)B^T + V_2 + G_2 + V_3 + G_3 + V_4 + G_4]\times(BB^T + 3I)^{-1} \tag{11}$$

(2) Update $V_1 = \text{argmin}_{V_1}\mathscr{L}$:

$$\begin{aligned} V_1 S &= \left(D_S^T D_S + \mu_1 I\right)^{-1}\left[D_S^T Y_H + \mu_1(AB - G_1)S\right] \\ V_1\overline{S} &= (AB - G_1)\overline{S} \end{aligned} \tag{12}$$

where $\overline{S}$ denotes the down-sampling matrix which selects the pixels not selected by S.

(3) Update $V_2 = argmin_{V_2}\mathscr{L}$:

$$V_2 = \left[\eta(RD_S)^T RD_S + \mu_1 I\right]^{-1}\left[\eta(RD_S)^T Y_M + \mu_1(A\text{-}G_2)\right] \tag{13}$$

(4) Update $V_3 = argmin_{V_3}\mathscr{L}$:

$$V_3 = \text{soft}\left(A\text{-}G_3, \frac{\eta_1}{2\mu_1}\right) \tag{14}$$

where $\text{soft}(\upsilon_1, \upsilon_2) = \text{sign}(\upsilon_1)\times\max(|\upsilon_1| - \upsilon_2, 0)$.

(5) Update $V_4 = \text{argmin}_{V_4}\mathscr{L}$:

The sparse coefficient of each superpixel is solved separately. For the i^{th} superpixel V_4^i, we calculate its sparse coefficients by:

$$V_4^i = U\left(\Sigma - \frac{\eta_2}{2\mu}\right)_+ V^T \tag{15}$$

where $U\Sigma V^T$ is the singular value decomposition (SVD) of $A^i - G_4^i$, and $(\cdot)^+$ denotes positive singular values.

(6) Update Lagrangian multipliers G_1, G_2, G_3 and G_4:

$$\begin{aligned} G_1 &= G_1 + V_1 - AB \\ G_2 &= G_2 + V_2 - A \\ G_3 &= G_3 + V_3 - A \\ G_4 &= G_4 + V_4 - A \end{aligned} \tag{16}$$

2.3. Spatial dictionary and sparse coefficient optimization

As is summarized in Algorithm 2, we derive theoretically the optimization process of the spatial dictionary D_P and spatial sparse coefficients Ψ via utilizing the remaining spatial information provided by the image ε_H and ε_M which cannot represent by the spectral dictionary D_S.

2.3.1. Spatial dictionary optimization

The spatial dictionary D_P represents the structure primitives of edges and textures in the remaining spatial information provided by the image ε_H and ε_M. However, it is challenging to obtain the exact values of the equivalent matrixes R_P and P in the image patch domain. In this section, we try to use variable substitution to eliminate the constraints of two equivalent matrixes (Han et al., 2020). From Eq. (17), we notice that ε_M and $\varepsilon_H R_P$ can sparsely represent by the spatial dictionary D_P and the degraded spatial dictionary PD_P with the same sparse coefficient $\Psi_P = \Psi R_P$, respectively.

$$\begin{aligned} \varepsilon_H R_P &= PD_P\Psi R_P + n_H \\ \varepsilon_M &= D_P\Psi R_P + n_M \end{aligned} \tag{17}$$

It can be found that Eq. (17) is a typical sparse representation problem, and the spatial dictionary D_P can be estimated by optimizing the following Eq. (18).

$$\underset{D_P}{argmin}\left\|\varepsilon_P - \widetilde{D}_P\Psi_P\right\|_F^2 + \lambda\|\Psi_P\|_0 \tag{18}$$

where $\widetilde{D}_P = \begin{bmatrix}\beta PD_P \\ D_P\end{bmatrix}$, $\varepsilon_P = \begin{bmatrix}\beta\varepsilon_H R_P \\ \varepsilon_M\end{bmatrix}$. β is a tradeoff parameter between matching LR patch and finding corresponding HR patch that is compatible with its neighbors. Here, we can utilize the consistent adaptive sequential dictionary learning algorithm (CASDL) (Seghouane and Iqbal, 2018) to efficiently estimate the spatial dictionary $D_P = [0\ I]\ \widetilde{D}_P$, the degraded spatial dictionary $PD_P = 1/\beta[I\ 0]\ \widetilde{D}_P$, and the degraded sparse coefficients $\Psi_P = \Psi R_P$.

Algorithm 2. (*The spatial dictionary and spatial sparse coefficient optimization algorithm*)

Input: Spectral dictionary D_S and spectral sparse coefficient A, auxiliary parameter, δ, β, $\tau = 1$, $\mu_2 = 10^{-3}$, $\rho = 1.1$;
Initialization: $E_H = Y_H - D_S ABS$, $E_M = Y_M - RD_S A$;
Step 1: Spatial Dictionary Optimization
(a) Compute D_P via Eq. (18) by CASDL;
Step 2: Spatial sparse coefficient optimization
Solving (36) via ADMM: **For** $k = 0, 1, \ldots, T_P$ **do**
(a) Compute U via Eq. (22);
(b) Compute V via Eq. (23);
(c) Compute J via Eq. (24);
(d) Compute Ψ via Eq. (25);
(e) Compute n_H via Eq. (26);
(f) Compute P_1, P_2, P_3 via Eq. (27);
(g) Compute $\mu : \mu = min(\rho\mu, \mu_{max})$;
End for
Step 3: $X = D_S A + D_P{}^{\circ}\ \Psi$
Output: The final fusion HSI X.

2.3.2. Spatial sparse coefficient optimization

Another important prior feature of the HSI in the urban green spaces scenes is that it has spatial nonlocal similarity, so the remaining spatial information unrepresented by the spectral dictionary appears more low-rank. To recover the high-frequency information hidden in the remaining part more accurately, in this subsection, we present a structured sparse low-rank representation strategy to recover the spatial sparse coefficient. Inspired by the idea of sparse transform learning (Xue et al., 2021), we take the low-rank structure U of Ψ into consideration when characterizing the sparse property of Ψ in the transform domain. As Ψ is simultaneously sparse and low-rank, so is U. The fusion model Eq. (5) can be specifically formulated is therefore

$$\underset{\Psi}{argmin}\|\varepsilon_H - PD_P\Psi\|_F^2 + \beta\|\varepsilon_M - D_P\Psi R_P\|_F^2 + \|\Psi\|_* + \delta\|U\|_1 + \tau\|n\|_{2,1} s.t.\ \Psi = UV^T, VV = I, \tag{19}$$

where $\|\Psi\|_*$ is the low-rank term, $\|U\|_1$ is the sparse term, $\|n\|_{2,1}$ is the error term, and V is the transform matrix. δ is a tradeoff parameter and τ is a balance parameter. The D_P, PD_P, and ΨR_P are obtained in the last

subsection, so we can simplify the optimization problem Eq. (19) as follows:

$$\underset{\Psi}{argmin}\|\varepsilon_H - P\mathrm{D}_P\Psi\|_F^2+\|\Psi\|_* + \delta\|U\|_1 + \tau\|n_H\|_{2,1} s.t.\ \Psi = UV^T, VV = I \tag{20}$$

By introducing an auxiliary variable $J = \Psi$, we can equivalently transform Eq. (20) into an augmented Lagrangian function:

$$\begin{aligned} G \ &= \|J\|_* + \delta\|U\|_1 + \tau\|n_H\|_{2,1} + \\ &\langle P_1, \varepsilon_H - P\mathrm{D}_P\Psi - n_H\rangle + \langle P_2, \Psi - J\rangle + \langle P_3, \Psi - UV^T\rangle + \\ &\frac{\mu_2}{2}\left(\|\varepsilon_H - P\mathrm{D}_P\Psi - n_H\|_F^2+\|\Psi - J\|_F^2+\left\|\Psi - UV^T\right\|_F^2\right) \end{aligned} \tag{21}$$

where P_1, P_2, P_3 are three Lagrange multipliers, μ_2 represents positive penalty parameters, $\mu_2 = \rho\mu_2$, $\rho \succ 0$. Eq. (21) can be solved by breaking into six subproblems in the ADMM framework:

(1) Update U:

$$U = \underset{U}{argmin}\frac{\delta}{\mu_2}\|U\|_1 + \frac{1}{2}\left\|\Psi - UV^T + \frac{P_3}{\mu}\right\|_F^2 U = S_{\delta/\mu}\left(\left(\Psi + \frac{P_3}{\mu_2}\right)V\right) \tag{22}$$

where $S_{\beta/\mu}()$ is the soft thresholding operator.

(2) Update V:

$$V = \underset{V}{argmin}\frac{\mu}{2}\left\|\Psi + \frac{P_3}{\mu_2} - UV^T\right\|_{\mathrm{F}}^2, s.t.\ V^TV = I,\ V = WC^T \tag{23}$$

where $[W, D, C] = \mathrm{SVD}\left((\Psi + P_3/\mu_2)^TU\right)$.

(3) Update J:

$$J = \underset{J}{argmin}\|J\|_* + \frac{\mu_2}{2}\left\|J - \left(\Psi + \frac{P_2}{\mu_2}\right)\right\|_F^2 J = ES_{1/\mu_2}(\Sigma)F^T \tag{24}$$

where $[E, \Sigma, F] = \mathrm{SVD}(\Psi + P_2/\mu_2)$.

(4) Update Ψ:

$$\begin{aligned} &\Psi = \underset{\Psi}{argmin}\left\|P\mathrm{D}_P\Psi - \left(\varepsilon_H - n_H + \frac{P_1}{\mu_2}\right)\right\|_F^2 + \left\|\Psi - J + \frac{P_2}{\mu_2}\right\|_F^2 + \left\|\Psi - UV^T + \frac{P_3}{\mu_2}\right\|_F^2 \\ &\Psi = \left(P\mathrm{D}_P^TP\mathrm{D}_P + 2I\right)^{-1}\left(P\mathrm{D}_P^TM_1 + M_2 + M_3\right) \end{aligned} \tag{25}$$

where $M_1 = \varepsilon_H - \Psi + P_1/\mu_2$, $M_2 = J - P_2/\mu_2$ and $M_3 = UV^T - P_3/\mu_2$.

(5) Update n_H:

$$n_H = \underset{n_H}{argmin}\frac{\tau}{\mu_2}\|n_H\|_{2,1} + \frac{1}{2}\left\|n_H - \left(n_H - P\mathrm{D}_P\Psi + \frac{P_1}{n_H}\right)\right\|_F^2$$

$$\left[n_{H:,i}\right] = \begin{cases} \dfrac{\left\|[\theta]_{:,i}\right\|_2 - \tau/\mu_2}{\left\|[\theta]_{:,i}\right\|_2}[\theta]_{:,i}, & if\ \left\|[\theta]_{:,i}\right\|_2 > \tau/\mu_2; \\ 0, & otherwise \end{cases} \tag{26}$$

where $\theta = \varepsilon_H$ - $P\mathrm{D}_P\Psi + P_1/\mu_2$. The i^{th} column of n_H and M are denoted by $[n_H]_{:,i}$ and $[M]_{:,i}$, respectively.

(6) Update Lagrange multipliers P_1, P_2, P_3:

$$\begin{aligned} P_1 &= P_1 + \mu_2(\varepsilon_H \text{ - } P\mathrm{D}_P\Psi - n_H) \\ P_2 &= P_2 + \mu_2(\Psi - J) \\ P_3 &= P_3 + \mu_2\left(\Psi - UV^T\right) \end{aligned} \tag{27}$$

The HSI-MSI fusion framework proposed in this paper reconstructs the latent HR-HSI from the input HR-MSI and LR-HSI by combining the spectral dictionary with the spatial dictionary to preserve full spatial and spectral information simultaneously. Moreover, structured sparse low-rank priors are imposed into two fusion models in the spectral domain and spatial domain, respectively, which can reduce less spectral distortions and ensure more accurate spatial features. The overall framework of our DDSSLR method is illustrated in Fig. 2.

3. Experiments

To evaluate the performance of our DDSSLR method, experiments are conducted on both two synthetic datasets and one real dataset from different urban green spaces scenes, featuring a variety of different sensor and geographical characteristics.

3.1. Experiment datasets

1) Synthetic datasets

We perform our experiments on two synthetic datasets, namely, Pavia University (Wei et al., 2015) and Washington DC Mall (Zhang et al., 2020). For the Pavia dataset, we crop a sub-region of size 128 × 128 × 93 with a ground sampling distance (GSD) of 1.3 m as the ground truth (GT) and generate LR-HSI of size 32 × 32 × 93 by applying a rotationally symmetric 5 × 5 Gaussian blur filter with a standard deviation of 2.5, followed by downsampling with a ratio of 4 along with both horizontal and vertical directions for all bands. The corresponding HR-MSI is obtained by the IKONOS-like reflectance spectral response filter that contains four multispectral bands. For the Washington dataset, we select a sub-region of size 250 × 250 × 191 with a GSD of 2.5 m as the GT and generate LR-HSI of size 50 × 50 × 191 by applying a rotationally symmetric 5 × 5 Gaussian blur filter with a standard deviation of 2.5, followed by downsampling with a ratio of 5 along with both horizontal and vertical directions for all bands. The corresponding HR-MSI is obtained by the Landsat TM bands 1–5 and 7, covering the spectral regions. Additionally, these simulated LR-HSIs are corrupted by Gaussian noise with SNR = 30 dB, and HR-HSIs are corrupted by Gaussian noise with SNR = 40 dB.

2) Real dataset

The real dataset was acquired by the Earth Observing-1 Mission (EO-1) satellite over Paris, France (Simões et al., 2015). we selected a sub-region of size 72 × 72 × 128 with a GSD of 30 m in Hyperion as the GT and a sub-region of size 72 × 72 × 9 with a GSD of 30 m in ALI as HR-MSI. We first generate the LR-HSI of size 24 × 24 × 128 by applying the Starck-Murtagh blur filter and downsampling with a ratio of 3. We normalize all bands of the EO-1 dataset before fusion making the 0.999

Table 1
Six PQIs of the eight methods over the testing image from Pavia University dataset.

PQIs	PSNR	RMSE	SAM	UIQI	ERGAS	DD
GSOMP+	38.831	1.179	2.248	0.988	1.122	0.883
BSR	42.785	0.801	1.199	0.994	0.737	0.481
NLSTF	40.923	0.905	1.562	0.992	0.849	0.634
SSR	39.306	1.092	1.897	0.989	1.025	0.759
SSFCNN	42.941	0.743	1.274	0.995	0.674	0.522
TFNet	45.047	0.575	0.972	0.996	0.525	0.395
OTD	44.985	0.578	0.998	0.996	0.529	0.404
Ours	**45.542**	**0.548**	**0.909**	**0.997**	**0.497**	**0.369**

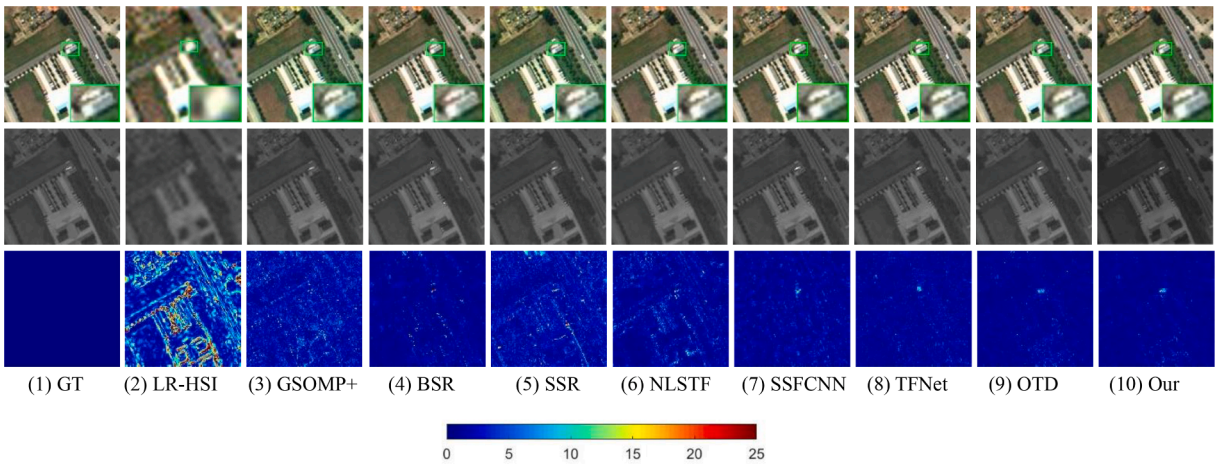


Fig. 3. 1st row: the GT and reconstructed images from Pavia University dataset, bands 45–25-8 as R-G-B at scaling factor s = 4. 2nd row: the reconstructed images of different methods at band 45. 3rd row: the error maps of the corresponding methods at band 45.

Table 2
Six PQIs of the eight methods over the testing image from Washington DC Mall dataset.

PQIs	PSNR	RMSE	SAM	UIQI	ERGAS	DD
GSOMP+	43.452	0.676	1.831	0.771	1.021	0.469
BSR	48.115	0.396	1.047	0.862	0.538	0.276
NLSTF	41.093	0.902	1.752	0.711	1.422	0.486
SSR	43.123	0.712	1.872	0.770	1.035	0.478
SSFCNN	47.627	0.442	1.204	0.867	0.552	0.309
TFNet	49.806	0.334	0.842	0.875	0.415	0.227
OTD	49.869	0.304	0.844	0.874	0.412	0.223
Ours	**50.919**	**0.261**	**0.695**	**0.898**	**0.339**	**0.186**

intensity quantile corresponded to 1. The GT and the LR-HSI were denoised by subspace projection through truncated SVD. This is a real dataset, so we estimate the real spatial blurring matrix B and the real spectral response matrix R by the Hysure (Simões et al., 2015).

3.2. Evaluation criterion

We use six quantitative image quality indices (PQIs) to gauge the reconstruction quality: peak signal-to-noise ratio (PSNR)(dB), universal image quality index (UIQI), root mean square error (RMSE)(in 10^{-2}), spectral angle mapper (SAM)(in degrees), degree of distortion (DD)(in 10^{-2}) and dimensionless global relative error of synthesis (ERGAS). The higher PSNR and UIQI values imply the better fusion results, whereas the lower RMSE, SAM, ERGAS, and DD values correspond to the better fusion results. For visual evaluation, the darker the error map, the better the fusion results. Brighter spots represent greater deviations.

3.3. Experimental settings

We benchmark our method with several state-of-the-art methods, including Generalization of Simultaneous Orthogonal Matching Pursuit (G-SOMP+) (Akhtar et al., 2014), Bayesian Sparse Representation (BSR) (Akhtar et al., 2015), Nonlocal Sparse Tensor Factorization (NLSTF) (Dian et al., 2017), Superpixel-based Sparse Representation (SSR) (Fang et al., 2018), Spatial and Spectral Fusion Convolutional Neural Network

Table 3
Six PQIs of the eight methods over the testing image from EO-1 dataset.

PQIs	PSNR	RMSE	SAM	UIQI	ERGAS	DD
GSOMP+	26.714	4.797	3.099	0.834	4.375	3.308
BSR	26.847	4.713	2.985	0.847	4.289	3.253
SSR	26.729	4.788	3.089	0.835	4.362	3.304
NLSTF	26.491	5.037	3.573	0.808	4.561	3.547
SSFCNN	26.776	4.761	3.289	0.847	4.328	3.338
TFNet	27.041	4.569	3.188	0.856	**4.143**	3.211
OTD	26.944	4.662	3.011	0.855	4.244	3.213
Ours	**27.246**	**4.563**	**2.858**	**0.873**	4.144	**3.165**

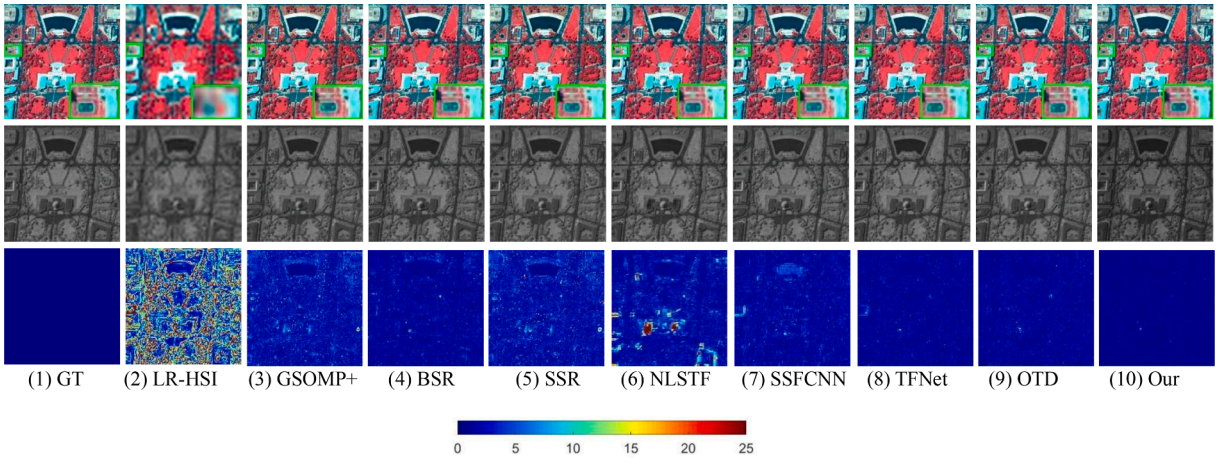


Fig. 4. 1st row: the GT and reconstructed images from Washington DC Mall dataset, bands 60–27-17 as R-G-B at scaling factor s = 5. 2nd row: the reconstructed images of different methods at band 60. 3rd row: the error maps of the corresponding methods at band 60.

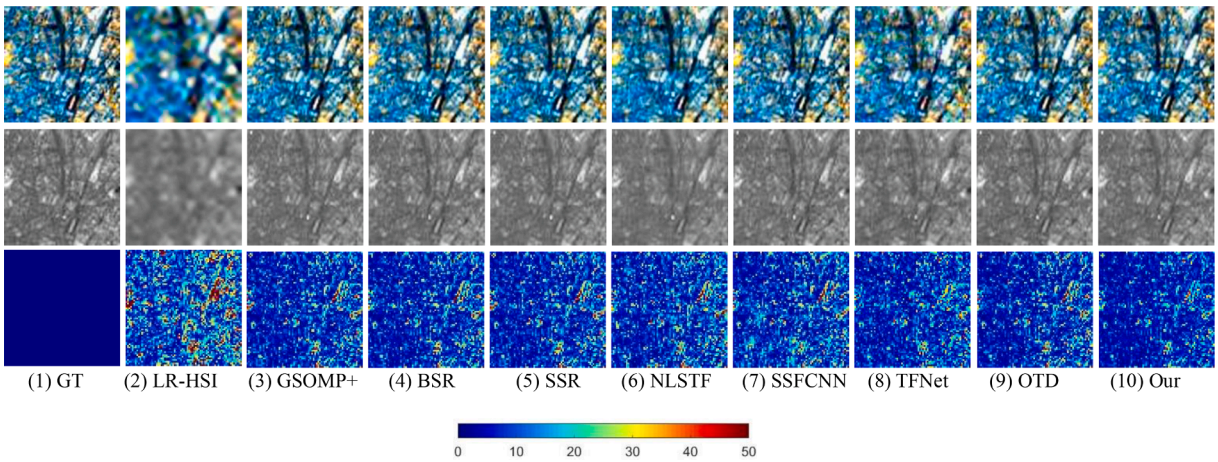


Fig. 5. 1st row: the GT and reconstructed images from the real EO-1 dataset, bands 51–30-20 as R-G-B at scaling factor s = 3, 2nd row: the reconstructed images of different methods at band 51. 3rd row: the error maps of the corresponding methods at band 51.

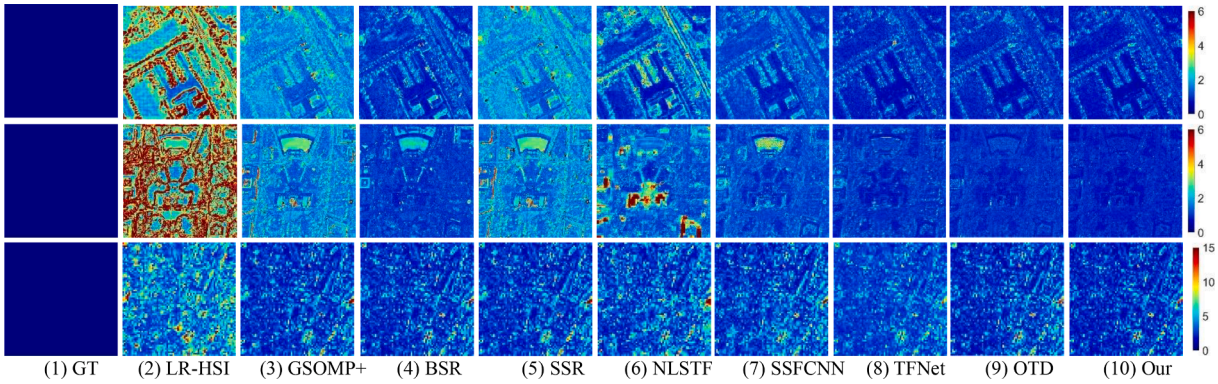


Fig. 6. SAM error images. 1st row: Pavia University (s = 4), 2nd row: Washington DC Mall (s = 5), 3rd row: real EO-1 (s = 3).

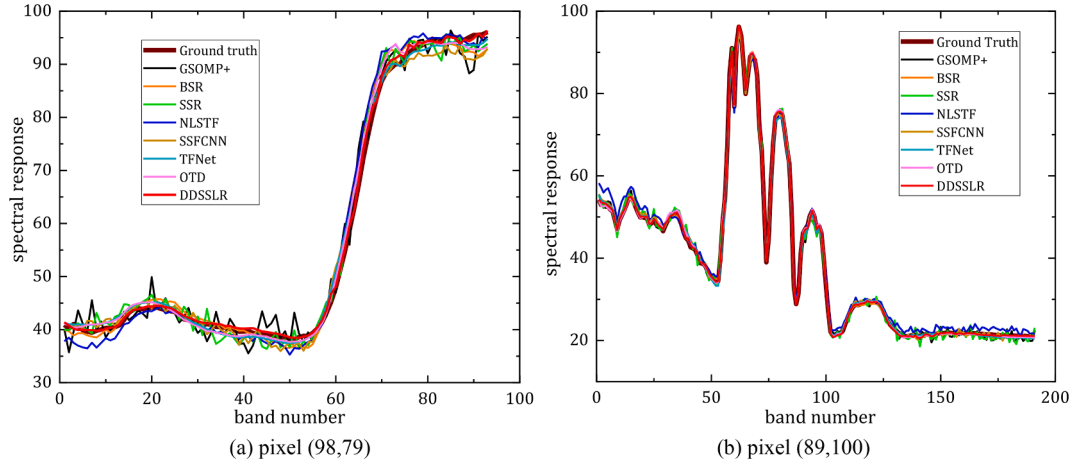


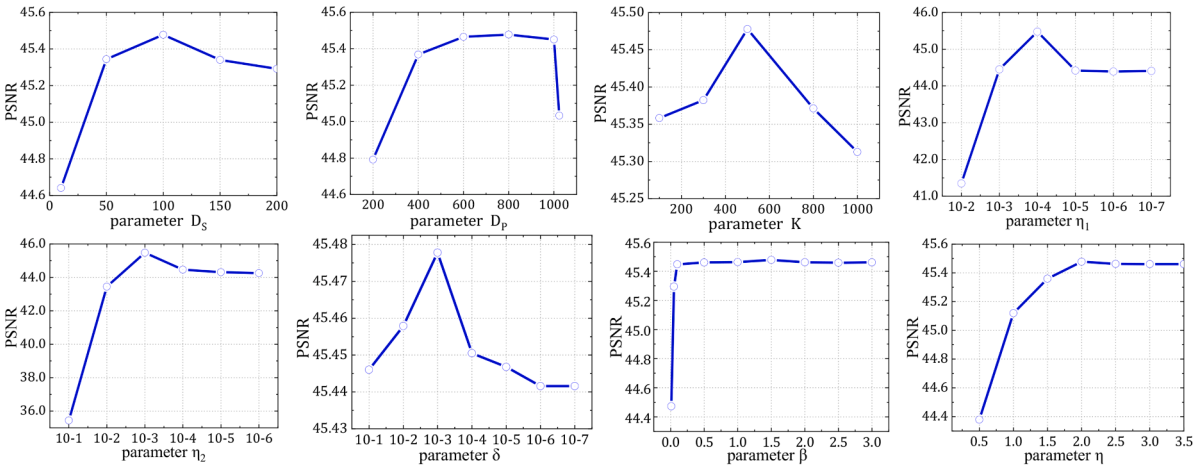
Fig. 7. The spectral curves of the representative pixels on (a) Pavia University, (b) Washington DC Mall.

(SSFCNN) (Han et al., 2018), Two-stream Fusion Network (TFNet) (Liu et al., 2020), Optimized Twin Dictionary (OTD) (Han et al., 2020). All the competing experiments are implemented using an Intel(R) Core(TM) i9-9900X CPU and GeForce RTX 1080Ti GPU.

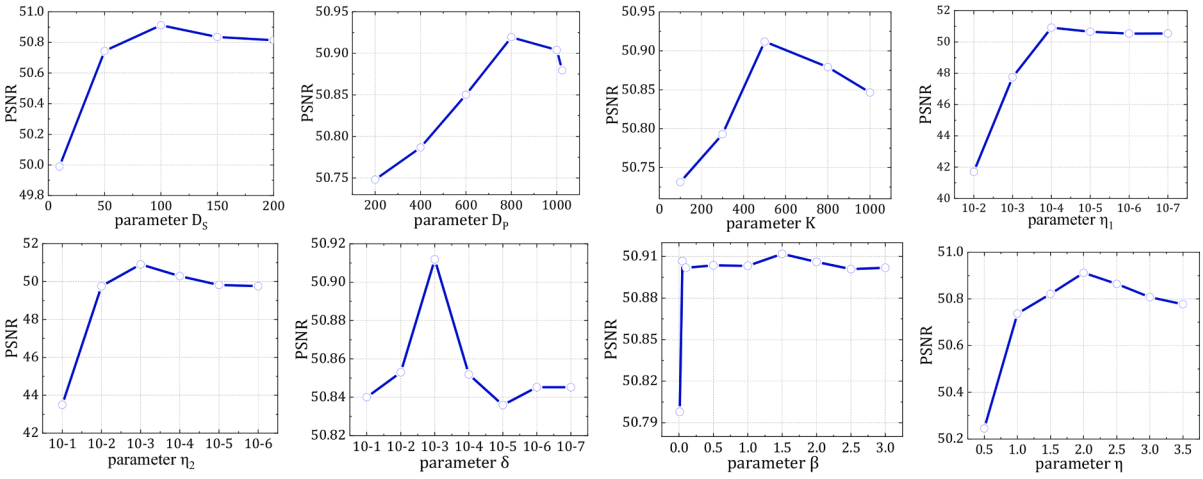
3.4. Result and analysis

3.4.1. Performance evaluation on the Pavia University dataset

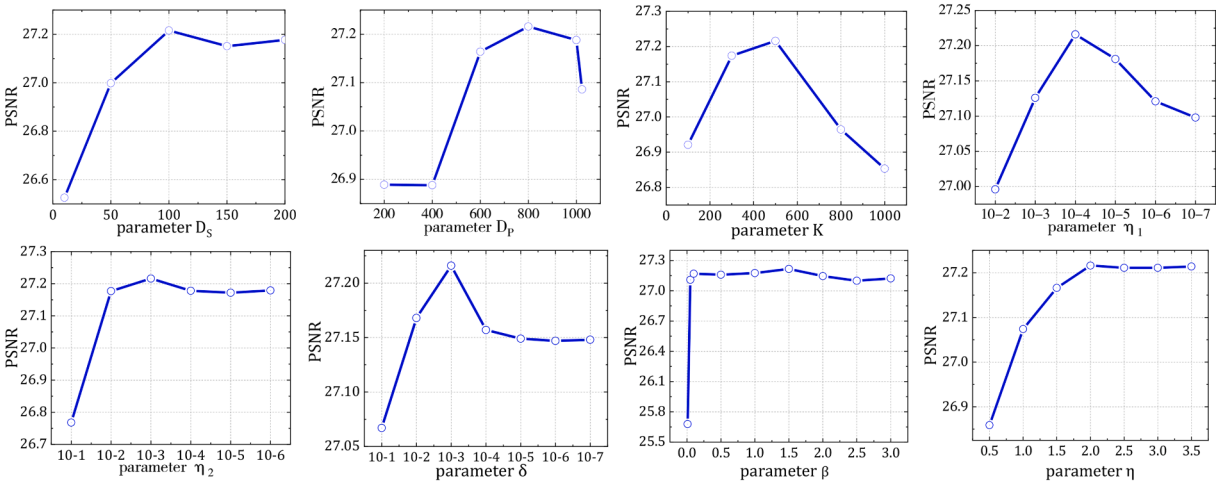
Table 1 tabulates the average performance in terms of six PQIs over the testing image from the Pavia University dataset. Our method outperforms other methods across all PQIs. The fusion results of GSOMP+, BSR, SSR are far inferior to that of OTD and our method, which is due to these methods only considering a single spectral dictionary, while OTD



(a) Pavia University (s=4)



(b) Washington DC Mall (s=5)



(c) real EO-1 (s=3)

Fig. 8. The PSNR change curves with different parameters settings.

and our method combine the advantages of using both spectral dictionary and spatial dictionary. We can easily observe that the PSNR of our method is higher than that of OTD by 0.557 dB. Our method imposes strong structure sparse low-rank prior in the spectral and spatial domains, therefore we can achieve higher fusion accuracy. The SSFCNN and TFNet utilize the convolution operation to explore the end-to-end mapping between LR-HSI and HR-HSI and obtain excellent fusion effects. Fig. 3 shows the false-color visual fusion results of the eight competing methods on the testing image with scaling factor s = 4. Moreover, we present fusion images and relative error maps on band 45

Table 4
Comparisons of fusion results on Pavia University dataset in four cases at s = 4.

PQIs	PSNR	RMSE	SAM	UIQI	ERGAS	DD
$Xhat_{\_OTD}$	44.450	0.626	1.065	0.996	0.572	0.433
$Xhat_{\_Our}$	44.923	0.584	0.972	0.997	0.530	0.394
$Xhat_{\_OTD}$ + $Ehat_{\_OTD}$	44.985	0.578	0.998	0.996	0.529	0.404
$Xhat_{\_Our}$ + $Ehat_{\_Our}$	45.542	0.548	0.909	0.997	0.497	0.369

for the different methods. From visual inspection, GSOMP+, NLSTF, and SSR cause blurred results, while BSR and SSFCNN seem to induce some texture deformation at the edge of the building. TFNet, OTD, and our method show noticeable improvement in the fusion performance. By fully considering the structure sparse low-rank prior and spatial-spectral features in the fusion model, our method recovers clearer texture structure and spectral information. Furthermore, the clear contour lines are not present in the error map of our method, indicating that our method contains fewer spectral distortions and maintains the main structural information at band 45.

3.4.2. Performance evaluation on Washington DC Mall dataset

Table 2 lists the average performance in terms of six PQIs over the testing image from the Washington DC Mall dataset. It can be observed that our method significantly outperforms other competing methods for almost all PQIs. This implies that DDSSLR makes better use of the latent spatial-spectral information of the HSI to maintain spectral reliability and restore more spatial details. The GSOMP + and SSR only utilize a spectral dictionary, resulting in information loss in the fusion results. The SSFCNN and TFNet present excellent fusion performance, but they may ignore the spectral correlation properties of HSI by using a black-box model. Fig. 4 demonstrates the false-color visual fusion results of the eight competing methods on the testing image from the Washington DC Mall dataset with scaling factor s = 5. We also present fusion images and relative error maps on band 60 for the different methods. As shown in Fig. 4, our method achieves better recovery of building structures and coarser-grained textures in the visual comparison. There are clear blurred details around the building in the error maps of GSOMP+, SSR, and NLSTF. This dataset mainly contains buildings and urban green spaces, in which the shadows can cause spectral distortion of the fusion result. However, the fusion result of our method is the closest to the GT and has the least bright spots on the error map. This shows our method can better preserve the spectral characteristics of the shadows caused by the buildings and trees.

3.4.3. Performance evaluation on the real EO-1 dataset

Table 3 presents the average performance in terms of six PQIs on the real EO-1 image for the competing methods. We can see the fusion results of GSOMP+, SSR are similar and NLSTF is slightly inferior, while our method achieves visually satisfying results. According to the quantitative metric results, the average PSNR value of our method is more

Table 5
Comparisons of running time for different HSI-SR methods (in seconds).

Datasets	Pavia University	Washington DC Mall	EO-1
GSOMP+	636	2301	524
BSR	125	612	44
SSR	888	4235	703
NLSTF	11	25	7
SSFCNN	78	552	29
TFNet	104	384	24
OTD	44	56	17
Ours	423	924	457

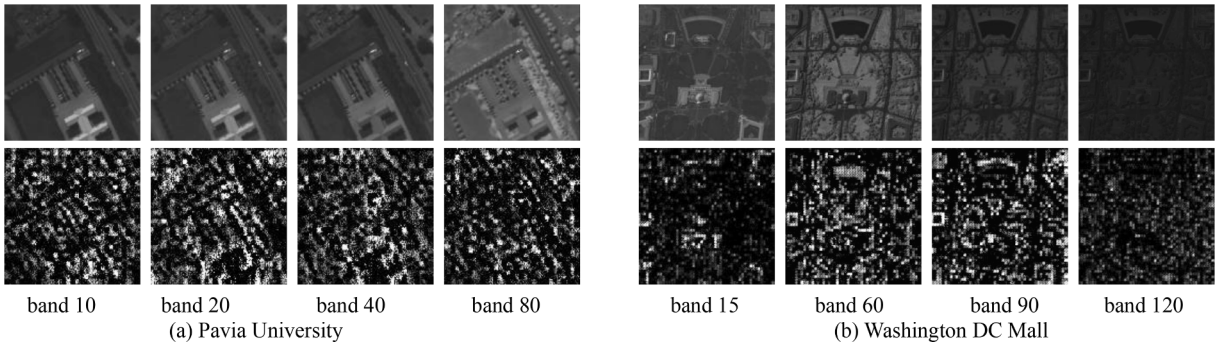


Fig.9. The upper row is the GT in typical bands and the lower row is the estimated images $Ehat_{\_Our}$ using D_P.

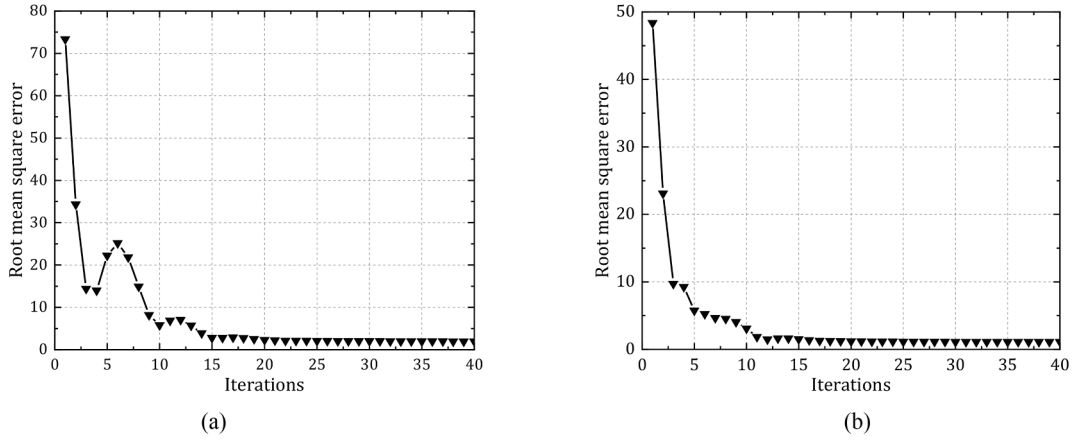


Fig.10. The convergence analysis of Algorithm 1 and Algorithm 2. (a) the RMSE curve of Eq. (9) versus iterations, (b) the RMSE curve of Eq. (20) versus iterations.

than 0.302 dB higher than that of OTD and 0.205 dB higher than that of the TFNet. Fig. 5 shows the false-color visual fusion result of the eight competing methods on the EO-1 dataset with scaling factor s = 3. Identically, we show fusion images and relative error maps on band 51 among eight competing methods. From these subfigures, we can notice that our method reconstructs a result most visually similar to the GT. The EO-1 image mainly shows a high-density building area in Paris. The texture feature is mostly clustered in dot-like patches. For all methods, the fusion results of building areas are blurred, and the bright spots on the error maps mainly appear around the buildings. Our method has relatively fewer bright spots on the error map compared to other methods, indicating that the fusion performance of our method on the real dataset is still more satisfactory.

4. Discussion

4.1. Model validation

To visually discern the differences in the spectral reconstruction quality across the different methods, Fig. 6 shows the SAM error images from the three datasets obtained by the eight competing methods. These SAM error images visualize the spectral distortion severity and distribution for the fusion results. Obviously, the GSOMP+, SSR, and NLSTF methods show more spectral distortion than other methods. Although the BSR, SSFCNN have better fusion performance, there are unsatisfactory SAM errors in edge regions. We can notice that the spectral distortion is mainly distributed in the shadow cast by buildings and trees in the SAM error maps (2)-(10). Our method best preserves the spectral properties of the shadow in the original LR-HSI. The spectral reflectance curves of the pixel (96,79) in the Pavia University dataset and the pixel (89,100) in the Washington DC Mall dataset are shown in Fig. 7 as representative examples. We can see that the trend of the spectral curves of the eight competing methods on representative pixels is roughly the same. Still, the ability of different methods to restore the spectral reflectance values on different wavelength bands is different. As shown in Fig. 7, the spectral response curve of the DDSSLR is the closest to that of the GT compared to other methods. This further implies that our method is better at preserving spectral information.

4.2. Parameter analysis

The key parameters in our experiment are set as follows: the size of D_S and D_P are set to 100 and 800, respectively; the number of superpixels K is set to 500; the sparsity regularization parameters $\eta_1 = 1 \times 10^{-4}$, $\delta = 1 \times 10^{-3}$; the low-rank regularization parameters $\eta_2 = 1 \times 10^{-3}$, the trade-off parameters $\beta = 1.5$, $\eta = 2$.

To assess the sensibility of these eight parameters on fusion performance, we evaluated model performance by varying these parameters individually. When evaluating the effect of one of the eight parameters for the fusion performance, we fixed the other parameters according to the above settings.

Fig. 8 (a) represents the PSNR change curve concerning the different parameters in the fusion experiment on the Pavia University dataset at s = 4. Fig. 8 (b) describes the PSNR change curve with the different parameters in the fusion experiment on the Washington DC Mall dataset at s = 5. Fig. 8 (c) demonstrates the PSNR change curve with the different parameters in the fusion experiment on the real EO-1 at s = 3.

We can see the PSNR values in these three datasets are gradually increasing when the size of D_S changes within the range [0,100], but gradually decreasing when D_S is greater than 100; Similarly, the PSNR change curves with the size of D_P in these three datasets show that the PSNR values are maximized at D_P of 800. The PSNR change curves with K in these three datasets imply that the PSNR values reach their peaks when $K = 500$. In addition, the effects of the sparsity regularization parameters η_1, δ and the low-rank regularization parameters η_2 on fusion performance for the three datasets are shown in Fig. 8. The PSNR values are maximized when $\eta_1 = 1 \times 10^{-4}$, $\eta_2 = 1 \times 10^{-3}$ and $\delta = 1 \times 10^{-3}$, respectively. Also, the PSNR change curves versus the trade-off parameters η and β show that our method achieves the optimal performances on these three datasets when $\beta = 1.5$ and $\eta = 2$, respectively.

Due to the similar trends of the PSNR change curves of the same parameter across the three datasets in Fig. 8, we believe that our method has stable performances across different datasets and that our parameter settings are universal and effective.

4.3. Ablation experiments

Our method achieves better fusion performance on the three datasets by combining spatial-spectral dual-dictionary fusion with structured sparse low-rank representation. To assess the effectiveness of our method joining spatial-spectral dual-dictionary with structured sparse low-rank representation, we perform ablation experiments to test the effect of the individual components. In this section, we consider four cases of fusion experiments on the testing image of the Pavia University dataset at s = 4.

The first case is to only consider the spectral dictionary in OTD (denoted as $Xhat_{\_OTD}$); the second case is to consider only the spectral dictionary in our method (denoted as $Xhat_{\_Our}$); the third case is to use the OTD (denoted as $Xhat_{\_OTD} + Ehat_{\_OTD}$,) and the fourth case is to use our method (denoted as $Xhat_{\_Our} + Ehat_{\_Our}$), which $Ehat_{\_OTD}$ and $Ehat_{\_Our}$ denote the fusion results using the spatial dictionary in OTD and our method, respectively. The fusion results of these four cases are shown in Table 4. By comparing $Xhat_{\_OTD}$ with $Xhat_{\_Our}$, we observe that constructing local low-rank structures by using superpixels to constrain the fusion model in the spectral domain can greatly improve the fusion effect. Similarly, comparing $Xhat_{\_Our}$ and $Xhat_{\_Our} + Ehat_{\_Our}$, we can find that the joint optimization of the spectral dictionary and the spatial dictionary can fully restore the spatial information on the latent HSI and reduce the spectral distortion.

Fig. 9 presents the GT and the estimated images $Ehat_{\_Our}$ using D_P in typical bands. The upper and lower row images have high similarity, which demonstrates the effectiveness of the spatial dictionary D_P in our method.

4.4. Convergence analysis

We can efficiently solve the fusion model in the spectral domain via SALSA, which involves optimizing iteratively multiple variables. Fig. 10 (a) plots the RMSE of the fusion result in Eq. (9) in each iteration to analyze the convergence of Algorithm 1. We can see Algorithm 1 coverages after 20–25 iterations. Similarly, we efficiently solve the fusion model in the spatial domain via ADMM. Fig. 10 (b) plots the RMSE of the fusion result in Eq. (20) in each iteration to analyze the convergence of Algorithm 2. It shows that Algorithm 2 coverages faster and coverage to a small RMSE value after 10 iterations. This convergence analysis of two algorithms is performed on the testing image of the Pavia University dataset with scaling factor s = 4.

4.5. Computational cost

A speed benchmark on two synthetic datasets and a real dataset is shown in Table 5. The GSOMP+, BSR, SSR, NLSTF, OTD, and our method are implemented with MATLAB, while SSFCNN and TFNet are implemented in PyTorch. Our method was slower than BSR, NLSTF, and OTD, but faster than GSOMP + and SSR. On one hand, our method is more complex due to alternating the optimization of the spectral sparse coefficients and the spectral dictionary using ADMM. On the other hand, two fusion models in both domains are constrained by the sparse low-rank prior and those optimization processes will be more costly. Since GSOMP + and SSR compute the cumulative correlation between

dictionary atom and the image patch during each iteration, they have the slowest computing time. It is worth noting that the deep learning-based methods of SSFCNN and TFNet are different from traditional methods. They require the preparation of a training set with the same structure as the testing image, as well as a long training time. Only the time spent in the testing phase is recorded in Table 5.

5. Conclusion

We construct a novel HSI-MSI fusion framework that combines spatial-spectral dual-dictionary and relevant structured sparse low-rank representation to fuse the input LR-HSI and HR-MSI. Specifically, the spectral dictionary characterizing the generalized spectrum is learned and sparse low-rank priors of the local structure are imposed to the spectral pixels within the same superpixel in HR MSI/HSI. Additionally, in the spatial domain, the spatial dictionary is learned by utilizing the remaining high-frequency information, and a unitary transformation is used to factorize the regression coefficient into a low-rank matrix in the subspace. Experiments on two synthetic HSI datasets and a real HSI dataset demonstrate the PSNR of our method exceeds that of OTD by 0.557 dB, 1.050 dB, and 0.302 dB, respectively. The structured sparse low-rank representation imposed into two fusion models as constraints can improve the fusion performance. This paper serves as basic research on the data source for urban green infrastructure monitoring. The obtained HR-HSI can provide an intuitive decision basis for the monitoring and management of urban green infrastructure. Future research will focus on urban vegetation mapping and classification by species using the obtained HR-HSI.

CRediT authorship contribution statement

Nan Chen: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Lichun Sui:** Writing – review & editing, Supervision. **Biao Zhang:** Formal analysis, Validation. **Hongjie He:** Methodology, Formal analysis, Validation. **Kyle Gao:** Formal analysis, Writing – review & editing. **Yandong Li:** Data curation, Validation. **José Marcato Junior:** Formal analysis, Writing – review & editing. **Jonathan Li:** Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the National Natural Scientific Foundation of China under Grants 41601345, 41571346, 41871380, and 4141379, and the China Scholarship Council (CSC). We acknowledge CAPES PrInt for supporting the internationalization project.

Appendix A

Notations

Notation	Description	Notation	Description
$X \in \mathbb{R}^{\lambda_X \times WH}$	the targeted HR-HSI	U	the low-rank structure ofΨ
$Y_M \in \mathbb{R}^{\lambda_Y \times WH}$,$Y_H \in \mathbb{R}^{\lambda_X \times wh}$	the available HR-MSI/LR-HSI	V	the transform matrix ofΨ
$D_S \in \mathbb{R}^{\lambda_X \times K_S}$,$D_P \in \mathbb{R}^{B_P \times K_P}$	the spectral/spatial dictionary	K	the superpixel number
$A \in \mathbb{R}^{K_S \times N}$,$\Psi \in \mathbb{R}^{K_P \times N_P}$	the spectral/spatial coefficients matrix	λ_1,λ_2	the regularization parameters
$E_H \in \mathbb{R}^{\lambda_X \times wh}$,$E_M \in \mathbb{R}^{\lambda_Y \times WH}$	the error components that cannot be represented by D_S in Y_M andY_H	$\eta_1 \geqslant 0$	the sparse regularization parameters
ϵ_H,ϵ_M	the stretched patches of the error matrixes E_H and E_M image patch domain	$\eta_2 \geqslant 0$	the low-rank regularization parameters
$B \in \mathbb{R}^{N \times N}$	the block circulant convolution	η, β,δ	the trade-off parameters
$S \in \mathbb{R}^{N \times n}$	the spatial down-sampling operator	P, R_P, n_H,n_M	the equivalent expressions of H = BS,R, N_H, N_M in image patch domain
$R \in \mathbb{R}^{\lambda_Y \times \lambda_X}$	the spectral response transform factor	$\|\cdot\|_1$,$\|\cdot\|_F$	the 1-norm and Frobenius norm
$N_H \in \mathbb{R}^{\lambda_X \times n}$,$N_M \in \mathbb{R}^{\lambda_Y \times N}$	noise matrices	V_1	the Lagrangian multiplier for solvingD_S
μ_2	the positive penalty parameters.	μ,μ_1	the Lagrangian parameter for solving D_S,A
$S_{\beta/\mu}()$	the soft thresholding operator	$G_1G_2G_3G_4$	the Lagrangian multipliers for solvingA
$[n_{H:,i}]$, $[\theta_{:,i}]$	The $i-th$ column of n_H andθ	P_1,P_2 P_3	the Lagrange multipliers for solvingΨ
τ	a balance parameter	$\| \|_*$,$\| \|_{2,1}$	the nuclear norm and$l_{2,1}$ norm

References

Abbas, S., Peng, Q., Wong, M.S., Li, Z., Wang, J., Ng, K.T.K., Kwok, C.Y.T., Hui, K.K.W., 2021. Characterizing and classifying urban tree species using bi-monthly terrestrial hyperspectral images in Hong Kong. ISPRS J. Photogramm. Remote Sens. 177, 204–216. https://doi.org/10.1016/j.isprsjprs.2021.05.003.

Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A.T., 2010. Fast image recovery using variable splitting and constrained optimization. IEEE Trans. Image Process. 19 (9), 2345–2356. https://doi.org/10.1109/TIP.2010.2047910.

Akhtar, N., Shafait, F., Mian, A., 2014. Sparse spatio-spectral representation for hyperspectral image super-resolution. Proc. ECCV 63–78. https://doi.org/10.1007/978-3-319-10584-0_5.

Akhtar, N., Shafait, F., Mian, A., 2015. Bayesian sparse representation for hyperspectral image super resolution. Proc. CVPR 3631–3640. https://doi.org/10.1109/CVPR.2015.7298986.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. FNT Mach. Learn. 3, 1–122. https://doi.org/10.1561/2200000016.

Chen, F., Jiang, H., Van de Voorde, T., Lu, S., Xu, W., Zhou, Y., 2018. Land cover mapping in urban environments using hyperspectral APEX data: A study case in Baden, Switzerland. Int. J. Appl. Earth Obs. Geoinf. 71, 70–82. https://doi.org/10.1016/j.jag.2018.04.011.

Chen, Y., He, W., Yokoya, N., Huang, T.-Z., 2020. Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition. IEEE Trans. Cybern. 50 (8), 3556–3570. https://doi.org/10.1109/TCYB.622103610.1109/TCYB.2019.2936042.

Dian, R., Fang, L., Li, S., 2017. Hyperspectral image super-resolution via non-local sparse tensor factorization. Proc. CVPR 3862–3871. https://doi.org/10.1109/CVPR.2017.411.

Dian, R., Li, S., Fang, L., 2019. Learning a low tensor-train rank representation for hyperspectral image super-resolution. IEEE Trans. Neural Networks Learn. Syst. 30 (9), 2672–2683. https://doi.org/10.1109/TNNLS.2018.2885616.

Dian, R., Li, S., Fang, L., Bioucas-Dias, J., 2018. Hyperspectral image super-resolution via local low-rank and sparse representations. Proc. IGARSS 4003–4006. https://doi.org/10.1109/IGARSS.2018.8519213.

Fang, L., Zhuo, H., Li, S., 2018. Super-resolution of hyperspectral image via superpixel-based sparse representation. Neurocomputing 273, 171–177. https://doi.org/10.1016/j.neucom.2017.08.019.

Ghasrodashti, E.K., Karami, A., Heylen, R., Scheunders, P., 2017. Spatial resolution enhancement of hyperspectral images using spectral unmixing and Bayesian sparse representation. Remote Sens. 9, 541. https://doi.org/10.3390/rs9060541.

Han, X., Yu, J., Xue, J.-H., Sun, W., 2020. Hyperspectral and multispectral image fusion using optimized twin dictionaries. IEEE Trans. Image Process. 29, 4709–4720. https://doi.org/10.1109/TIP.8310.1109/TIP.2020.2968773.

Han, X.H., Shi, B., Zheng, Y., 2018. SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution. In: in: 25th IEEE International Conference on Image Processing (ICIP), pp. 2506–2510. https://doi.org/10.1109/ICIP.2018.8451142.

Lanaras, C., Baltsavias, E., Schindler, K., 2017. Hyperspectral super-resolution with spectral unmixing constraints. Remote Sens. 9, 1196. https://doi.org/10.3390/rs9111196.

Liu, M., Tuzel, O., Ramalingam, S., Chellappa, R., 2011. Entropy rate superpixel segmentation. Proc. CVPR 2097–2104. https://doi.org/10.1109/CVPR.2011.5995323.

Liu, X., Liu, Q., Wang, Y., 2020. Remote sensing image fusion based on two-stream fusion network. Inf. Fusion 55, 1–15. https://doi.org/10.1016/j.inffus.2019.07.010.

Long, Y., Rivard, B., Rogge, D., Tian, M., 2019. Hyperspectral band selection using the N-dimensional spectral solid angle method for the improved discrimination of spectrally similar targets. Int. J. Appl. Earth Obs. Geoinf. 79, 35–47. https://doi.org/10.1016/j.jag.2019.03.002.

Modzelewska, A., Fassnacht, F.E., Stereńczak, K., 2020. Tree species identification within an extensive forest area with diverse management regimes using airborne hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 84, 101960. https://doi.org/10.1016/j.jag.2019.101960.

Nezhad, Z.H., Karami, A., Heylen, R., Scheunders, P., 2016. Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (6), 2377–2389. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2016.2528339.

Noomen, M., Hakkarainen, A., van der Meijde, M., van der Werff, H., 2015. Evaluating the feasibility of multitemporal hyperspectral remote sensing for monitoring bioremediation. Int. J. Appl. Earth Obs. Geoinf. 34, 217–225. https://doi.org/10.1016/j.jag.2014.08.016.

Pan, L., Li, H.-C., Sun, Y.-J., Du, Q., 2018. Hyperspectral image reconstruction by latent low-rank representation for classification. IEEE Geosci. Remote Sens. Lett. 15 (9), 1422–1426. https://doi.org/10.1109/LGRS.885910.1109/LGRS.2018.2844555.

Rougier, S., Puissant, A., Stumpf, A., Lachiche, N., 2016. Comparison of sampling strategies for object-based classification of urban vegetation from very high resolution satellite images. Int. J. Appl. Earth Obs. Geoinf. 51, 60–73. https://doi.org/10.1016/j.jag.2016.04.005.

Seghouane, A.-K., Iqbal, A., 2018. Consistent adaptive sequential dictionary learning. Signal Process. 153, 300–310. https://doi.org/10.1016/j.sigpro.2018.07.018.

Shi, Y., Skidmore, A.K., Wang, T., Holzwarth, S., Heiden, U., Pinnel, N., Zhu, X., Heurich, M., 2018. Tree species classification using plant functional traits from LiDAR and hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 73, 207–219. https://doi.org/10.1016/j.jag.2018.06.018.

Simoes, M., Bioucas-Dias, J., Almeida, L.B., Chanussot, J., 2015. A convex formulation for hyperspectral image superresolution via subspace-based regularization. IEEE Trans. Geosci. Remote Sens. 53 (6), 3373–3388. https://doi.org/10.1109/TGRS.2014.2375320.

Sui, L., Li, L.i., Li, J., Chen, N., Jiao, Y., 2019. Fusion of hyperspectral and multispectral images based on a Bayesian nonparametric approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (4), 1205–1218. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2019.2902847.

Wei, Q.i., Bioucas-Dias, J., Dobigeon, N., Tourneret, J.-Y., 2015. Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Trans. Geosci. Remote Sens. 53 (7), 3658–3668. https://doi.org/10.1109/TGRS.2014.2381272.

Xue, J., Zhao, Y., Liao, W., Chan, J.-W., 2019. Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising. IEEE Trans. Geosci. Remote Sens. 57 (7), 5174–5189. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2019.2897316.

Xue, J., Zhao, Y.-Q., Bu, Y., Liao, W., Chan, J.-W., Philips, W., 2021. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. IEEE Trans. Image Process. 30, 3084–3097. https://doi.org/10.1109/TIP.8310.1109/TIP.2021.3058590.

Yang, Y., Zheng, J., Chen, S., 2020. Local low-rank matrix recovery for hyperspectral image denoising with $\ell 0$ gradient constraint. Pattern Recognit. Lett. 135, 167–172. https://doi.org/10.1016/j.patrec.2020.04.012.

Yi, C., Zhao, Y., Chan, J.C., 2018. Hyperspectral image super-resolution based on spatial and spectral correlation fusion. IEEE Trans. Geosci. Remote Sens. 56, 4165–4177. https://doi.org/10.1109/TGRS.2018.2828042.

Yokoya, N., Yairi, T., Iwasaki, A., 2012. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. Geosci. Remote Sens. 50, 528–537. https://doi.org/10.1109/TGRS.2011.2161320.

Zhang, F., Wu, Y., Xiao, Z., Geng, L., Wu, J., Wen, J., Wang, W., Liu, P., 2019. Super resolution reconstruction for medical image based on adaptive multi-dictionary learning and structural self-similarity. Comput. Assist. Surg. 24, 81–88. https://doi.org/10.1080/24699322.2018.1560092.

Zhang, X., Huang, W., Wang, Q., Li, X., 2020. SSR-NET: spatial-spectral reconstruction network for hyperspectral and multispectral image fusion. IEEE Trans. Geosci. Remote Sens. 59, 5953–5965. https://doi.org/10.1109/TGRS.2020.3018732.