Joint Nonlocal, Spectral, and Similarity Low-Rank Priors for Hyperspectral-Multispectral Image Fusion

Tatiana Gelvez-Barrera, Henry Arguello, Alessandro Foi

***Abstract*—The fusion of a low-spatial-and-high-spectral resolution hyperspectral image (HSI) with a high-spatial-and-low-spectral resolution multispectral image (MSI) allows synthesizing a high-resolution image (HRI), supporting remote sensing applications such as disaster management, material identification, and precision agriculture. Unlike existing variational methods using low-rank regularizations separately, we present an HSI-MSI fusion method promoting various low-rank regularizations jointly. Our method refines the HRI spatial and spectral correlations from the individual HSI and MSI data through the proper plug-and-play (PnP) of a nonlocal patch-based denoiser in the alternating direction method of multipliers (ADMM). Notably, we consider the nonlocal self-similarity, the spectral low-rank, and introduce a rank-one similarity prior. Furthermore, we demonstrate via an extensive empirical study that the rank-one similarity prior is an inherent characteristic of the HRI. Simulations over standard benchmark datasets show the effectiveness of the proposed HSI-MSI fusion outperforming state-of-the-art methods, particularly in recovering low-contrast areas.**

***Index Terms*—HSI-MSI Fusion, Low-rank regularizations, PnP-ADMM, Nonlocal patch based denoiser.**

I. Introduction

HYPERSPECTRAL imaging sensors face a trade-off between the spatial resolution and the signal-to-noise ratio because of the limited incident energy [1]. Hence, hyperspectral-multispectral image (HSI-MSI) fusion is a computational technique that combines the valuable information of two images obtained from different sensors, adding robustness and easing subsequent processing in remote sensing applications [2]. Precisely, a high spatial-spectral resolution image (HRI) is synthesized from a low-spatial-and-high-spectral resolution hyperspectral image (HSI) and a high-spatial-and-low-spectral resolution multispectral image (MSI) [3].

The HSI-MSI fusion has been mainly addressed through data-driven deep-learning (DL) and interpretable variational frameworks. DL uses training data to learn a non-linear operator that maps the HSI and MSI observations to the desired HRI. A DL approach first combines the HSI and MSI features and inputs them into a single network that maps the HRI [4]–[6]. An alternative DL approach first extracts the HSI and MSI features using two independent networks and then combines them to estimate the HRI [7]–[9]. However, DL methods are limited by the need for several pairs of HSI-MSI scenes during the training process [10]. Variational HSI-MSI fusion methods use instead prior information such as sparsity, smoothness, and low-rank properties to regularize the HRI [11]–[15]. Such priors can be coupled to take advantage of various properties simultaneously and more effectively [16], [17].

The low-rank prior contributes differently across the spatial, spectral, and nonlocal dimensions. Then, regularizers operating over individual dimensions can be coupled to improve the HRI estimation [18]; tensor decomposition models with nonlocal self-similarity, spectral unmixing, or sparse-based regularizations are among the most popular [19]–[22]. However, there is a gap in previous low-rank-based HSI-MSI fusion methods, neglecting strong correlations in the similarity dimension. Therefore, we present an HSI-MSI fusion that, unlike previous methods, synergistically combines the nonlocal self-similarity, the spectral low-rank, and an introduced low-rank similarity prior, dubbed *rank-one similarity*. We introduce this concept upon the idea that intrinsic nonlocal spatial structures occur across the entire spectral dimension. We demonstrate through a separate extensive empirical study that the rank-one similarity prior is an inherent characteristic of the HRI.

The low-rank priors are combined in the problem through an implicit regularization function, which is solved numerically by the versatile and flexible plug-and-play with alternating direction method of multipliers (PnP-ADMM) framework [23]–[27]. The PnP-ADMM enables employing multichannel block-matching and 3D filtering (M-BM3D) denoising, presented in [28] as a realization of BM3D to process color images. M-BM3D finds the matching block positions in the first channel and propagates them to filter the other channels. The proposed fusion method is tested over three benchmark datasets covering multispectral, hyperspectral, and satellite images through several experiments that compare the quantitative and qualitative results against various state-of-the-art variational methods. Our method provides superior fusion quality, particularly in the recovery of spectral signatures, and can obtain significant gains at low-contrast blocks. The main contributions are summarized as follows

- We introduce the concept of rank-one similarity prior (Section II-A), which we validate through an extensive study (Section II-B).
- We combine the nonlocal self-similarity, spectral low-rank, and rank-one similarity priors with an implicit regularization for HSI-MSI fusion (Section III-A).
- We present a practical algorithm based on the versatile

T. Gelvez-Barrera is with the Department of Electrical Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002, e-mail: tatiana.gelvez@correo.uis.edu.co. T. Gelvez-Barrera is supported by the Colciencias scholarship *Doctorados Nacionales - 785.*

This work was supported by the Academy of Finland (projects no. 310779, 326439, and 326473), and by the Sistema General de Regalías (SGR) under Project 8933.

H. Arguello is with the Department of Systems Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002, e-mail: henarfu@uis.edu.co.

A. Foi is with Tampere University, Tampere, Finland, 33720, e-mail: alessandro.foi@tuni.fi.

PnP-ADMM framework to solve the formulated problem via the M-BM3D denoising (Section III-B).

- We demonstrate a significant improvement in the visual quality and objective metrics, particularly in the mean spectrum-wise PSNR, with excellent recovery of spectral signatures even in low-contrast regions (Section IV).

II. Rank-One Similarity Prior

The geometry of the objects in an HRI is typically irrespective of the object reflectance and responsivity at different spectral bands. Therefore, we introduce the following rank-one similarity prior to formalize an intrinsic HRI characteristic under which all spectral bands share common structural similarities of small spatial blocks.

A. Concept

Let $\mathbf{z}\in\mathbb{R}^{N_m^2 L_h}$ represent the vector form of an HRI with N_m^2 spatial pixels and L_h spectral bands, and let $\mathbf{P}_{\lambda,i}\in\mathbb{R}^{p\times p}$ denote a $p\times p$ HRI block whose top-left corner is at the i^{th} spatial position, $i=1,\dots,N_m^2$, within the λ^{th} spectral band, $\lambda=1,\dots,L_h$. A dissimilarity map is a 2D array containing the dissimilarity measure of a *reference* block to its neighboring blocks. Let $\mathbf{D}_{\lambda,i}\in\mathbb{R}^{w\times w}$ denote the dissimilarity map for the $\mathbf{P}_{\lambda,i}$ reference block, where $w\in\mathbb{Z}_+$ is the neighborhood size and whose entries are calculated in terms of the ℓ_2-norm as

$$\mathbf{D}_{\lambda,i}(j) = \|\mathbf{P}_{\lambda,i} - \mathbf{P}_{\lambda,j}\|_2^2, \tag{1}$$

for $j\subseteq\Omega_i$, where Ω_i is a set of indexes of the blocks within the $w\times w$ neighbourhood of the reference block. We further denote by $\tilde{\mathbf{D}}_i\in\mathbb{R}^{w^2\times L_h}$ the matrix whose columns contain the normalized and vectorized dissimilarity maps for reference blocks at the same position index i over all spectral bands:

$$\tilde{\mathbf{D}}_i = [\text{vec}(\mathbf{D}_{1,i})/s(\mathbf{D}_{1,i}),\ \dots\ ,\text{vec}(\mathbf{D}_{L_h,i})/s(\mathbf{D}_{L_h,i})], \tag{2}$$

where $s(\mathbf{D}_{\lambda,i})$ denotes the sample standard deviation of $\mathbf{D}_{\lambda,i}$.

The *rank-one similarity prior* assumes that, for a high-contrast reference block at position index i, the dissimilarity maps differ across the spectral bands only by a proportionality factor, i.e. $\mathbf{D}_{\lambda,i}=\alpha_{\lambda,i}\mathbf{D}_{1,i}\ \forall\lambda$. Adopting this prior in practice means approximating $\tilde{\mathbf{D}}_i$ (2) by a rank-one matrix.

While alternative formulations are possible (e.g., adopting different norms for (1), or an histogram equalization of each $\mathbf{D}_{\lambda,i}(j)$ prior to forming (2)), for the sake of simplicity we adopt the above basic definitions throughout this work.

B. Validation

The rank-one similarity prior is validated through an extensive empirical study that analyzes the singular values of the matrix (2), whose decay is expected to be fast, especially from the first to the second singular value and on high-contrast blocks, as illustrated in Fig. 1 for the Colombia dataset [29].

The study follows the procedure below over ten datasets covering real-world objects, remote sensing, and satellite sources [30], [31].

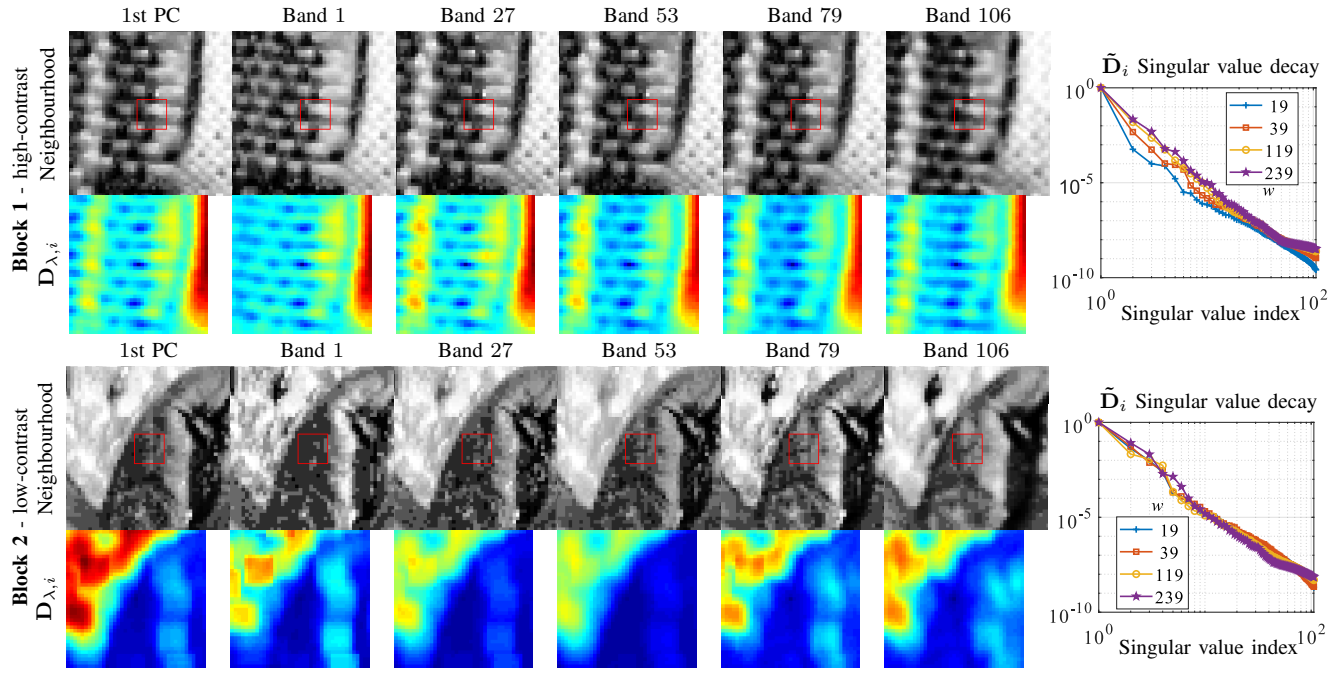

Figure 1. Illustration of the rank-one similarity prior for a high-contrast (top) and a low-contrast (bottom) reference block of size $p\times p=8\times 8$ from the Colombia dataset acquired in our optics laboratory [29]. A red square bounds each reference block into the RGB version of the search neighborhood employed to generate a dissimilarity map of size 39×39 across the first principal component (PC) and five spectral bands. There, the dissimilarity maps $\mathbf{D}_{\lambda,i}$, $\forall\lambda$, (1) appear to be all proportional to each other so that $\tilde{\mathbf{D}}_i$ (2) exhibits a fast singular value decay as observed at the bottom plots. The orange plot shows the decay when the dissimilarity map size is 39×39. The fast decay from the first to the second singular value indicates that $\tilde{\mathbf{D}}_i$ can be approximated as a rank-one matrix. We also show the singular-value decay when varying the dissimilarity map size from 19×19 to 239×239, while maintaining the reference block size fixed to 8×8. Observe that the decays for the high-contrast reference block are faster than for the low-contrast reference block and that the decays are faster for smaller dissimilarity map size.

1) Calculate the reference block contrast $\mathcal{C}(\mathbf{P}_{\lambda,i})$ as
$$\mathcal{C}(\mathbf{P}_{\lambda,i}) = \max(\mathbf{P}_{\lambda,i}) - \min(\mathbf{P}_{\lambda,i}). \tag{3}$$
2) Calculate the singular values of $\tilde{\mathbf{D}}_i$, $\forall i$.
3) Analyze the singular-value decay, where high-contrast blocks are expected to yield a fast decay, i.e. large $\mathcal{C}(\mathbf{P}_{\lambda,i})$ leads to fast decay of the singular values of $\tilde{\mathbf{D}}_i$.

We primarily focus on the case where the block size is $p\times p = 8\times 8$ and the neighborhood size is $w\times w = 39\times 39$, as in [28]. Figure 2 illustrates the singular value decay of $\tilde{\mathbf{D}}_i, \forall i$. We define a high-contrast reference block as a reference block whose contrast is higher than the half of the contrast's range for every analyzed dataset; other definitions are also possible. The expected behavior can be observed in the figure, where high-contrast blocks (on average 16.47% of all blocks) lead to a fast decay, i.e. red color must appear at the bottom and must be absent at the top of plots.

In particular, the second singular value is on average 3.2×10^{-3} times the first one (i.e. a decay from the first one of 99.7%) within the first quartile and 5.9×10^{-3} times (a 99.4% drop) for the first (i.e. lower) half of its distribution; this rate of decay continues through the further singular values, as can be seen in the figure. Such fast decay supports the incorporation of the rank-one similarity prior to the solution of spectral imaging problems. For instance, for the HSI-MSI fusion, the rank-one similarity prior is helpful to effectively propagate the detailed spatial information presented in the MSI across the narrow spectral bands presented in the HSI.

III. Hyperspectral-Multispectral Image Fusion

Let $\mathbf{M} \in \mathbb{R}^{N_h^2 \times N_m^2}$ denote a uniform spatial sub-sampling matrix, $\mathbf{B} \in \mathbb{R}^{N_m^2 \times N_m^2}$ denote a spatial blurring convolution matrix, and $\mathbf{n_h} \in \mathbb{R}^{N_h^2 L_h}$ denote additive Gaussian noise. The HSI, denoted by $\mathbf{z_h} \in \mathbb{R}^{N_h^2 L_h}$, can be modeled as a spatial blurred and sub-sampled version of the HRI $\mathbf{z}$ as

$$\mathbf{z_h} = (\mathbf{I}_{L_h} \otimes \mathbf{M})(\mathbf{I}_{L_h} \otimes \mathbf{B})\mathbf{z} + \mathbf{n_h} = \bar{\mathbf{M}}\bar{\mathbf{B}}\mathbf{z} + \mathbf{n_h}, \tag{4}$$

where the Kronecker product $\otimes$ applies the sub-sampling and blurring across the spectral bands maintaining a vector notation, so that $\bar{\mathbf{M}} \in \mathbb{R}^{N_h^2 L_h \times N_m^2 L_h}$ and $\bar{\mathbf{B}} \in \mathbb{R}^{N_h^2 L_h \times N_m^2 L_h}$.

Similarly, let $\mathbf{R} \in \mathbb{R}^{L_m \times L_h}$ denote a multispectral sensor spectral response, and $\mathbf{n_m} \in \mathbb{R}^{N_m^2 L_m}$ denote additive Gaussian noise. The MSI, denoted by $\mathbf{z_m} \in \mathbb{R}^{N_m^2 L_m}$, can be modeled as a spectrally degraded version of the HRI $\mathbf{z}$ as

$$\mathbf{z_m} = (\mathbf{R} \otimes \mathbf{I}_{N_m^2})\mathbf{z} + \mathbf{n_m} = \bar{\mathbf{R}}\mathbf{z} + \mathbf{n_m}, \tag{5}$$

where $\otimes$ applies the spectral degradation along all spatial pixels maintaining a vector notation, so that $\bar{\mathbf{R}} \in \mathbb{R}^{N_m^2 L_m \times N_m^2 L_h}$.

The proposed HSI-MSI fusion follows a subspace-based formulation and a PnP-ADMM algorithm to estimate the HRI from the HSI and MSI degraded observations.

A. Subspace-based Inverse Problem Formulation

The proposed cost function comprises two ℓ_2-norms maintaining the data fidelity to the observations and one implicit regularization function $\phi : \mathbb{R}^{N_m^2 L_h} \to \mathbb{R}$ that aims to promote

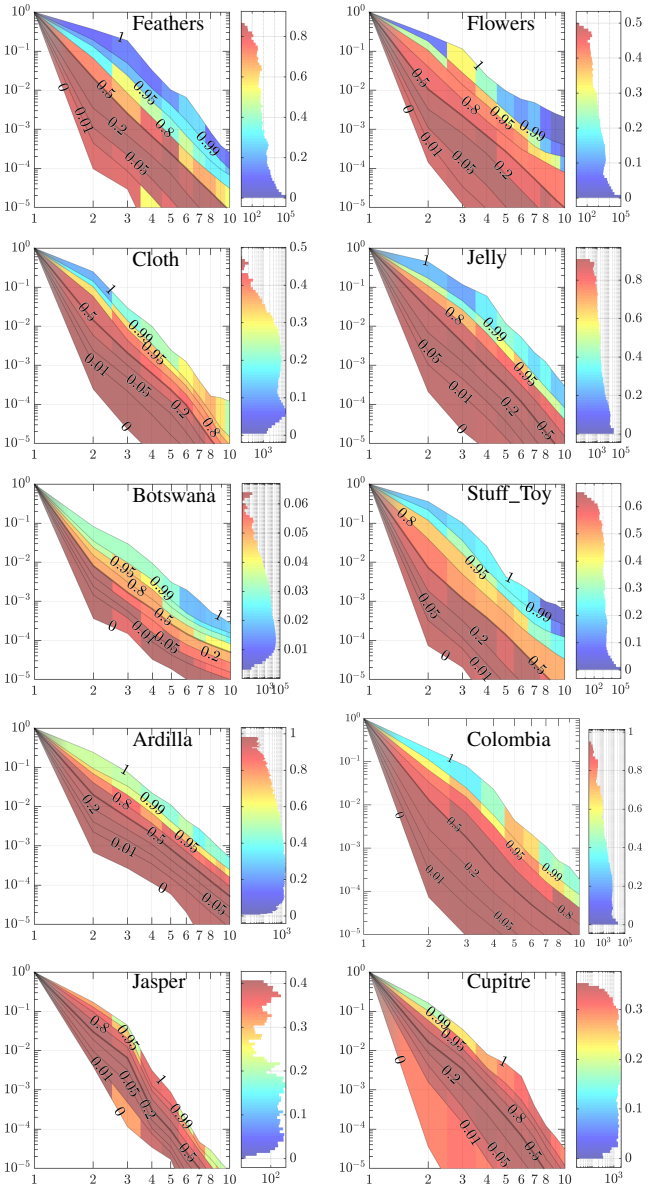

Figure 2. Singular value decay of the matrix $\tilde{\mathbf{D}}_i$ (2) for all reference blocks from ten publicly available datasets. The decays are grouped in quantiles ([0%-1%), [1%-5%), [5%-20%), [20%-50%), [50%-80%), [80%-95%), [95%-99%), [99%-100%]) and each singular value is colored according to the maximum contrast of the values that belong to the corresponding quantile; red color indicates high-contrast blocks, and blue color indicates low-contrast blocks. The color histogram on the right-hand side of each plot shows the distribution of the reference blocks according to their contrast; one can observe that in a typical natural image, the most frequent blocks have a relatively low-contrast (Cupitre being an exception, with a nearly uniform histogram). The plots show that only a very slim band of upper quantiles of the singular values (slowest decay) might feature only low-contrast blocks: higher-contrast blocks always have fast decay of the associated singular values and must appear at bottom.

jointly the spectral low-rank, the nonlocal self-similarities, and the rank-one similarity prior, taking full advantage of the intrinsic spatial-spectral structural correlations.

The HSI-MSI fusion is thus an inverse problem formulated as

$$\begin{aligned} &\underset{\mathbf{z}\in\mathbb{R}^{N_m^2 L_h}}{\text{minimize}} \;\; f(\mathbf{z}), \\ &f(\mathbf{z}) = \frac{1}{2}\|\mathbf{z_h} - \bar{\mathbf{M}}\bar{\mathbf{B}}\mathbf{z}\|_2^2 + \frac{\mu}{2}\|\mathbf{z_m} - \bar{\mathbf{R}}\mathbf{z}\|_2^2 + \lambda\phi(\mathbf{z}), \end{aligned} \tag{6}$$

where $\mu>0$ and $\lambda>0$ correspond to the regularization parameters that balance the three terms.

B. PnP-ADMM algorithm

The proposed algorithm to solve (6) follows the established PnP-ADMM framework [25], promoting the low-rank priors through a denoiser based on nonlocal regularization.

The PnP-ADMM begins with the introduction of three auxiliary variables $\mathbf{v}_i$, for $i=1,2,3$, upon which (6) becomes

$$\begin{aligned}&\underset{\mathbf{z},\mathbf{v}_i}{\text{minimize}}\ \ \frac{1}{2}\|\mathbf{z_h}-\bar{\mathbf{M}}\mathbf{v}_1\|_2^2+\frac{\mu}{2}\|\mathbf{z_m}-\bar{\mathbf{R}}\mathbf{v}_2\|_2^2+\lambda\phi(\mathbf{v}_3),\\&\text{subject to}\ \ \mathbf{v}_1=\bar{\mathbf{B}}\mathbf{z};\quad \mathbf{v}_2=\mathbf{z};\quad \mathbf{v}_3=\mathbf{z},\end{aligned}\tag{7}$$

whose augmented Lagrangian optimization problem with dual variables $\mathbf{g}_i$, for $i=1,2,3$ is given by

$$\begin{aligned}&\underset{\mathbf{z},\mathbf{v}_i,\mathbf{g}_i}{\text{minimize}}\ \ \mathcal{L}(\mathbf{z},\mathbf{v}_i,\mathbf{g}_i)\\&\mathcal{L}(\mathbf{z},\mathbf{v}_i,\mathbf{g}_i)=\frac{1}{2}\|\mathbf{z_h}-\bar{\mathbf{M}}\mathbf{v}_1\|_2^2+\frac{\mu}{2}\|\mathbf{z_m}-\bar{\mathbf{R}}\mathbf{v}_2\|_2^2+\\&\qquad\lambda\phi(\mathbf{v}_3)+\frac{\rho}{2}\|\mathbf{v}_1-\bar{\mathbf{B}}\mathbf{z}+\mathbf{g}_1\|_2^2+\\&\qquad\frac{\rho}{2}\|\mathbf{v}_2-\mathbf{z}+\mathbf{g}_2\|_2^2+\frac{\rho}{2}\|\mathbf{v}_3-\mathbf{z}+\mathbf{g}_3\|_2^2,\end{aligned}\tag{8}$$

where $\rho>0$ is the dual regularization parameter.

Each primal variable is optimized by solving the iterative alternating process

$$\begin{aligned}\mathbf{z}^{k+1}&\in\underset{\mathbf{z}}{\text{argmin}}\ \ \mathcal{L}(\mathbf{z},\mathbf{v}_i^k,\mathbf{g}_i^k),\\\mathbf{v}_i^{k+1}&\in\underset{\mathbf{v}_i}{\text{argmin}}\ \ \mathcal{L}(\mathbf{z}^{k+1},\mathbf{v}_i,\mathbf{g}_i^k),\end{aligned}\tag{9}$$

implemented by Algorithm 1 with variables calculated as detailed in Table I. Specifically, the proposed algorithm is termed R1BM3D, referring to the inclusion of the rank-one similarity prior and the use of the M-BM3D [28].

1) Initialization: In Algorithm 1, line 3, the estimated HRI and the auxiliary and dual variables are initialized as

$$\begin{aligned}\mathbf{z}^0&\leftarrow\frac{1}{2}(\bar{\mathbf{B}}^T\bar{\mathbf{M}}^T\mathbf{z_h}+\bar{\mathbf{R}}^T\mathbf{z_m}).\\\mathbf{v}_1^0&\leftarrow\bar{\mathbf{B}}\mathbf{z}^0;\ \ \mathbf{v}_2^0\leftarrow\mathbf{z}^0;\ \ \mathbf{v}_3^0\leftarrow\mathbf{z}^0.\end{aligned}\tag{10}$$

Algorithm 1 Rank-one via BM3D for HSI-MSI fusion

1: **procedure** R1BM3D($\mathbf{z_h}$, $\mathbf{z_m}$, $\bar{\mathbf{M}}$, $\bar{\mathbf{B}}$, μ, λ, ρ, *iters*)
2: $k\leftarrow 0$
3: $(\mathbf{z}^k,\mathbf{v}_i^k,\mathbf{g}_i^k)\leftarrow$ (as in (10)) ▷ Initialization
4: $\mathbf{W}_i\leftarrow$ Table I ▷ Precomputation
5: **while** $k<$ *iters* **do** ▷ Update
6: $\mathbf{z}^{k+1}\leftarrow\mathbf{W}_0^{-1}\mathbf{d}_0^k$
7: $\mathbf{v}_1^{k+1}\leftarrow\mathbf{W}_1^{-1}\mathbf{d}_1^k$
8: $\mathbf{v}_2^{k+1}\leftarrow\mathbf{W}_2^{-1}\mathbf{d}_2^k$
9: $\mathbf{q}_3^k\leftarrow\mathbf{z}^{k+1}-\mathbf{g}_3^k$
10: $\mathbf{v}_3^{k+1}\leftarrow$ PnP-BM3D$(\mathbf{q}_3^k,\lambda/\rho)$ ▷ (see Algorithm 2)
11: $\mathbf{g}_1^{k+1}\leftarrow\mathbf{v}_1^{k+1}-\bar{\mathbf{B}}\mathbf{z}^{k+1}+\mathbf{g}_1^k$.
12: $\mathbf{g}_2^{k+1}\leftarrow\mathbf{v}_2^{k+1}-\mathbf{z}^{k+1}+\mathbf{g}_2^k$.
13: $\mathbf{g}_3^{k+1}\leftarrow\mathbf{v}_3^{k+1}-\mathbf{z}^{k+1}+\mathbf{g}_3^k$.
14: $k\leftarrow k+1$
15: **return** $\mathbf{z}^{iters}$

2) Precomputation: In Algorithm 1, line 4 the matrices that will be used to invert the optimization sub-problems are precomputed as detailed in Table I.

3) Plug-and-Play of the Multichannel BM3D: In Algorithm 1, line 10, the M-BM3D addresses the implicit regularization function $\phi(\mathbf{v}_3)$ that promotes the HRI low-rank property, entailing important considerations detailed below.

In practice, the rank-one similarity prior is verified because the structural features that determine the nonlocal block similarity are *repeated* across the different spectral bands. However, over real-world measurements, the structural features may be buried under noise and be distorted by the coarse sampling, depending on the particular band. Therefore, it is challenging to obtain a reliable dissimilarity map to build the block-matching by working over individual bands.

When applying the principal component analysis (PCA) to the spectral bands, the structural features typically end up being represented with a substantially higher signal-to-noise ratio (SNR) in the first principal component (PC). Therefore, it is convenient to operate the block-matching over the first PC; otherwise, noise can disrupt the block-matching as illustrated in Fig. 3. As the structural features drive the block-matching on the first PC, it can be used by the rank-one prior for all spectral bands.

In principle, the BM3D can be applied separately on each band using the common block-matching inherited from the first PC. However, the spectral decorrelation provided by the PCA is also beneficial to further sparsify the 3D spectrum of a group of blocks within each spectral band in the M-BM3D, which is therefore applied upon PCA.

In Algorithm 1 line 10, the update of $\mathbf{v}_3$ assumes that $\mathbf{q}_3^k=\mathbf{z}^{k+1}-\mathbf{g}_3^k$ is a noisy version of $\mathbf{v}_3$ whose effective noise is considered to be white Gaussian with variance $\sigma^2=\lambda/\rho$. Therefore, a denoised version of $\mathbf{v}_3$ is obtained by the PnP of the M-BM3D following the procedure in Algorithm 2, where the block-matching step is driven over the first PC as follows.

1) Calculate the PCs of $\mathbf{q}_3^k$ organized as the matrix $\mathbf{Q}_3^k\in\mathbb{R}^{L_h\times N_m^2}$.
2) Find the mutual similar blocks by applying the block-matching in the first PC.
3) Propagate the positions of the found similar blocks to filter and preserve a small number $r\ll L_h$ of PCs. Note in Algorithm 2, line 7, that $\text{PCA}_r^{-1}(\cdot)$ denotes an inverse truncated PCA on only the first r components, meaning that the remaining L_h-r components are treated as zero.

Table I
SUMMARY OF VARIABLES COMPUTATION

$\mathbf{W}_0\in\mathbb{R}^{N_m^2L_h\times N_m^2L_h}$	$=\bar{\mathbf{B}}^T\bar{\mathbf{B}}+2\mathbf{I}_{N_m^2L_h}$.
$\mathbf{W}_1\in\mathbb{R}^{N_m^2L_h\times N_m^2L_h}$	$=\bar{\mathbf{M}}^T\bar{\mathbf{M}}+\rho\mathbf{I}_{N_m^2L_h}$.
$\mathbf{W}_2\in\mathbb{R}^{N_m^2L_h\times N_m^2L_h}$	$=\mu\bar{\mathbf{R}}^T\bar{\mathbf{R}}+\rho\mathbf{I}_{N_m^2L_h}$.
$\mathbf{d}_0^k\in\mathbb{R}^{N_m^2L_h}$	$=\bar{\mathbf{B}}^T(\mathbf{v}_1^k+\mathbf{g}_1^k)+(\mathbf{v}_2^k+\mathbf{g}_2^k)+(\mathbf{v}_3^k+\mathbf{g}_3^k)$.
$\mathbf{d}_1{}^k\in\mathbb{R}^{N_m^2L_h}$	$=\bar{\mathbf{M}}^T\mathbf{z_h}+\rho(\bar{\mathbf{B}}\mathbf{z}^{k+1}-\mathbf{g}_1^k)$.
$\mathbf{d}_2{}^k\in\mathbb{R}^{N_m^2L_h}$	$=\mu\bar{\mathbf{R}}^T\mathbf{z_m}+\rho(\mathbf{z}^{k+1}-\mathbf{g}_2^k)$.

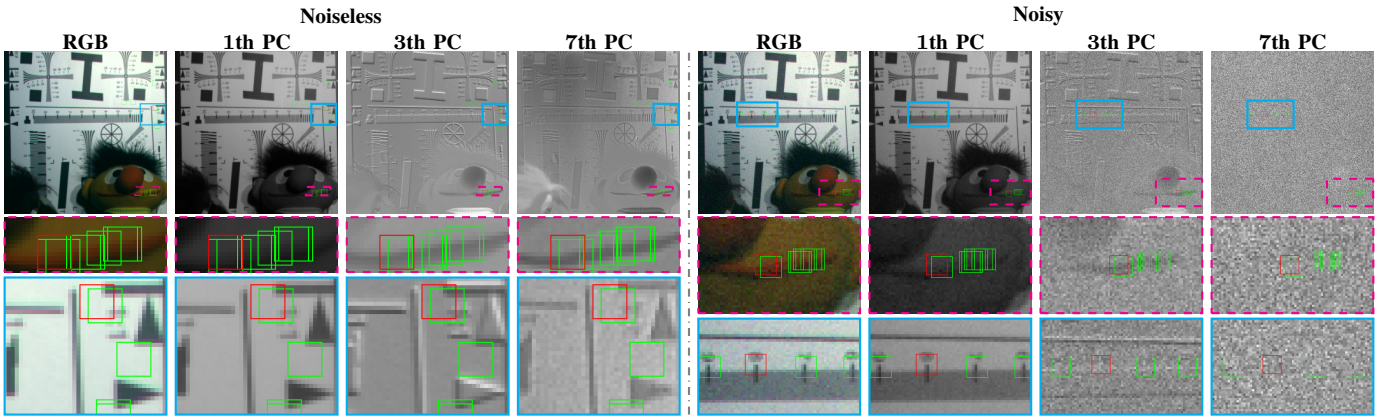

Figure 3. Suitability of the first PC to match similar blocks. The block-matching is executed in the first PC for a low-contrast and a high-contrast reference block that are bounded and zoomed in dashed magenta and cyan windows, respectively. The reference blocks are bounded with a red square and its corresponding matched similar blocks are bounded with a green square. The matched similar blocks found in the first PC that are superimposed over the 3^{th} and 7^{th} PCs result well suited for other few PCs, except for the noisy case where the last components are mainly driven by noise. Therefore, the block-matching found in the first PC can be propagated to filter only other PCs that are considered significant.

Algorithm 2 PnP rank-one prior via multichannel BM3D on top PCs

1: **procedure** PnP-BM3D($\mathbf{q}_3^k, \sigma^2$)

2: $\mathbf{Q}_3^k \leftarrow \mathcal{M}(\mathbf{q}_k^3)$ ▷ Unfold as a matrix
3: $\mathbf{P} \leftarrow \text{PCA}(\mathbf{Q}_k^3)$ ▷ Find principal components (PCs)
4: $r \leftarrow \text{HySime}(\mathbf{P})$ ▷ Find dimensionality
5: $\mathbf{P}_r \leftarrow [\mathbf{p}_1, \ldots, \mathbf{p}_i, \ldots, \mathbf{p}_r]$, $\mathbf{p}_i : i^{th}$ column of $\mathbf{P}$
6: $\hat{\mathbf{Q}}_3^k \leftarrow \text{M-BM3D}(\mathbf{P}_r, \sigma^2)$ ▷ Apply M-BM3D assuming noise variance σ^2
7: $\mathbf{V}_3^k \leftarrow \text{PCA}_r^{-1}\big(\hat{\mathbf{Q}}_3^k\big)$ ▷ Return filtered spectral bands
8: **return** $\mathbf{v}_3^k \leftarrow \text{vec}(\mathbf{V}_3^k)$

Solving the regularization $\phi(\mathbf{v}_3)$ through Algorithm 2 promotes the spectral low-rank, the nonlocal self-similarities, and the rank-one similarity *jointly*. The spectral low-rank is promoted when filtering and preserving just r PCs to represent the entire HRI, the nonlocal self-similarity is promoted when matching and filtering mutual nonlocal similar blocks, and the rank-one similarity is promoted when propagating the positions of similar blocks found in the first PC to filter the other PCs. The number r of preserved PCs is updated at each iteration by using the hyperspectral signal subspace identification by minimum error (HySime) method [32]. Notice that the algorithm's computational complexity is remarkably reduced, avoiding the calculation of multiple block-matching at each PC or each spectral band.

The first PC has the highest inter-pixel variance with well-defined blocks so that the filtering is conservative, and it is expected not to match together blocks that are dissimilar in other PCs, which are typically smoother. Nonetheless, in the rare but possible event where the block-matching on the first PC creates a group of dissimilar blocks in another PC, M-BM3D does not break down because the noise attenuation is done after shrinkage of the 3D transform of the group. Furthermore, matching significantly different blocks (e.g., above the noise level that exists in that PC) will elicit 3D-transform spectrum coefficients larger than the shrinkage threshold. Hence, these coefficients will be preserved, and the blocks' differences will be preserved after the inversion of the 3D transform.

C. Computational Complexity

The computational complexity of Algorithm 1 is mainly determined by the update of variable $\mathbf{v}_3$ in line 10 and unfolded in Algorithm 2. The computational complexity of Algorithm 2 is dominated by the steps from line 3 to line 6 corresponding to the calculation of the PCA, the dimension subspace estimation, and the BM3D filtering. The PCA implies the calculation of the covariance matrix and the singular value decomposition, entailing a complexity of $O(N_m^2 L_h^2 + L_h^3)$. The HySime method includes a noise estimation procedure with complexity $O(N_m^2 L_h^3 + L_h^4)$ [32] and subspace dimension estimation with complexity $O(N_m^4 L_h + L_h^3)$. The complexity of M-BM3D is proportional to $N_m^2(1+r)$ [28]. Finally, given that $r \ll L_h$, the asymptotic complexity of each iteration in Algorithm 1 can be estimated as $O(N_m^4 L_h + N_m^2 L_h^3 + L_h^4)$.

D. Convergence Analysis

Algorithm 1 is an instance of the block coordinate descent (BCD) optimization strategy. Therefore, the algorithm converges if the objective function decreases every iteration, entailing the different sub-problems' solutions [23]. Note that the update for variables $\mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{g}_i$, for $i = 1, \ldots, 3$ comes from a convex formulation with a closed-form solution, decreasing the objective function. In contrast, the update of variable $\mathbf{v}_3$ through the proposed PnP of the M-BM3D may not decrease the objective. However, according to the extensive simulations, the objective value is decreased at each iteration in most cases.

IV. Simulations and Results

To evaluate the proposed R1BM3D HSI-MSI fusion, we conducted extensive experiments across various publicly available multispectral and hyperspectral datasets described below.

1) Multispectral CAVE dataset: Stuff_Toy image from the standard CAVE dataset [30] of $512\times512\times31$ spatial-spectral resolution taken with a cooled CCD camera spanning the $(0.4-0.7)\mu$m spectral range in intervals of 0.01μm. The HSI was generated with a Gaussian blur kernel with standard deviation 2 pixel and spatial downsampling factor of $d_h=N_m/N_h=16$. The MSI was generated using a simulated spectral response $\mathbf{R}\in\mathbb{R}^{3\times31}$. The simulations were conducted with SNR level of 35dBs and 30dBs for the HSI and MSI.

2) Hyperspectral Pavia University Dataset: Semi-synthetic standard Pavia University dataset [31] of size $610\times340\times103$ taken with the ROSIS sensor spanning the $(0.43-0.86)\mu$m range, and spatial resolution of 1.3m. Following the experimental setup in [11], [14], [19], [33] only the top-left 256×256 spatial pixels and the last 93 spectral bands are considered after removing the water vapor absorption bands. The HSI was simulated with a Gaussian blur kernel with standard deviation 1.333 pixel and downsampling factor $d_h=N_m/N_h=4$ along vertical and horizontal spatial directions. The MSI was simulated using the spectral response of the IKONOS satellite sensor which captures 4 multispectral bands in the ranges $(0.45-0.52)$, $(0.52-0.60)$, $(0.63-0.69)$, and $(0.76-0.90)\mu$m [34]. As in [11], [19], [33], the simulations were conducted with SNR level of 35dB for the first 43 bands and SNR level of 30dB for the remaining 50 bands of the HSI, and with SNR level of 30dB for all bands of the MSI.

3) Hyperspectral Urban Dataset: The hyperspectral Urban dataset [35] contains information about six materials along 307×307 spatial locations and 210 spectral bands, spanning the $(400-2500)$nm spectral range with a resolution of 10nm [36]. The experiment considers a subset of spatial-spectral dimension $300\times300\times162$, after removing the $1-4$, $76, 87$, $101-111$, $136-153$, and $198-210$ noisy bands. The HSI was generated with a Gaussian blur kernel with standard deviation 1.2 pixel and downsampling factor of $d_h=N_m/N_h=6$. The MSI was generated using a simulated spectral response $\mathbf{R}\in\mathbb{R}^{6\times162}$. The simulations were conducted with SNR of 35dBs and 30dBs for the HSI and MSI.

We compare our approach against various state-of-the-art approaches that employ different strategies to obtain the fused image. Specifically, we compare against the subspace based Hyperspectral Superresolution (HySure) [12]; the HSI-MSI fusion based on Bayesian Sparse Representation (BSR) [11]; the Non-Negative Structured Sparse Representation (NSSR) [33]; the Couple Sparse Tensor Factorization (CSTF) [14]; the Clustering Manifold Structure (CMS) [37]; the super-resolution tensor-reconstruction (STEREO) [38], the super-resolution based on coupled Tucker Tensor approximation (SCOTT) [39], and the structured coupled LL1 decomposition (SC-LL1) [22]. Further, to analyze the effectiveness of the rank-one similarity prior, we compare the results against a baseline algorithm that replaces M-BM3D in Algorithm 2 line 6 by BM3D applied independently for each of the r preserved principal components, referred to as I-BM3D.

All comparison methods were implemented by their published code found in the hyperspectral super-resolution benchmark [40]–[43]. The main parameters of R1BM3D and of each comparison method have been tuned separately for each dataset to optimize the PSNR as detailed in Section IV-A.

The HSI-MSI fusion performance is evaluated in terms of the global, spatial, and spectral peak signal-to-noise ratio (PSNR) measured in decibels (dB), the Universal Image Quality Index (UIQI), the Spectral Angle Mapper (SAM) measured in degrees, and the dimensionless global relative error of synthesis (ERGAS), calculated as presented in [44], and the root-mean-square error (RMSE) calculated as in [11] for 8-bit representations of original and estimated images i.e. we scale the images to the range $[0, 255]$. Each quantitative metric is mathematically defined in the supplementary material.

A. Parameters Selection

Due to the very different spatial and spectral sampling conditions characterizing each dataset, the fusion algorithms all benefit from separate tuning of their key parameters. The comparison methods are tuned as follows in order to maximize the PSNR of the fused image separately for each dataset. For NSSR we set the number of dictionary atoms, the number of iterations, and the regularization terms to $K=256$, *iters*$=11$, $\eta_1=1e^{-4}$, and $\eta_2=8e^{-3}$ for Pavia University; $K=280$, *iters*$=12$, $\eta_1=1e^{-1}$, and $\eta_2=3.3e^{-4}$ for Stuff_Toy; and $K=300$, *iters*$=12$, $\eta_1=3.5e^{-5}$, and $\eta_2=1.7$ for Urban dataset. For CSTF we set the number of dictionary atoms of the spatial-spectral modes to $n_w=256$, $n_h=256$ and $n_s=11$ for Pavia University; $n_w=150$, $n_h=150$ and $n_s=12$ for Stuff_Toy; and $n_w=250$, $n_h=250$ and $n_s=5$ for Urban dataset. For CMS we set the size of the overlapped local full band patches, the number of clusters, and the interval between patches to $\sqrt{q}=2$, $K=250$, and $d=1$ for Pavia University; $\sqrt{q}=8$, $K=500$, and $d=2$ for Stuff_Toy; and $\sqrt{q}=2$, $K=300$, and $d=1$ for Urban dataset. For STEREO we set the tensor rank to $F=300$ for Pavia University, $F=85$ for Stuff_Toy and $F=300$ for Urban dataset. For SCOTT we set the main parameters according to the recoverability conditions to $R_1=R_2=250$ and $R_3=3$ for Pavia University, $R_1=R_2=150$ and $R_3=2$ for Stuff_Toy, and $R_1=R_2=300$ and $R_3=3$ for Urban dataset. For SC-LL1 we set the key regularization parameters as in [22] to $\theta=1e^{-3}$, $\lambda=0.8$, $\eta=1e^{-5}$ for Pavia University, $\theta=1e^{-3}$, $\lambda=0.8$, $\eta=5e^{-3}$ for Stuff_Toy, and $\theta=5e^{-3}$, $\lambda=0.8$, $\eta=9e^{-4}$ for Urban dataset.

R1BM3D has three key parameters, λ controlling the implicit regularization term, ρ controlling the dual variable weights, and μ controlling the weight of the fidelity term to the MSI. In the experiments, we fix $\lambda=\sigma^2\rho$, where σ stands for the MSI effective noise standard deviation, which is assumed by the M-BM3D for denoising. Then, ρ and μ are tuned through simulations for each dataset. Figure 4 shows the global PSNR as function of ρ and μ. Based on these simulations, we set $\mu=7$ and $\rho=0.16$ for Pavia University; $\mu=0.6$ and $\rho=0.02$ for Stuff_Toy; and $\mu=0.65$ and $\rho=0.09$ for Urban dataset. In general, the selection of ρ affects the convergence speed. Very small values will produce a fast convergence but will reduce the importance of the dual terms, affecting the quality. Very large values of ρ will instead produce slow convergence. Note that μ is commonly larger than ρ, hence the fidelity to the MSI in (8) is more relevant than the fidelity

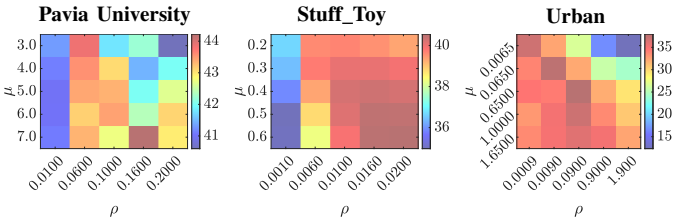

Figure 4. PSNR for Pavia University, Stuff_Toy, and Urban datasets. Each dataset benefits from a different combination of parameters ρ and μ.

to the HSI towards obtaining good quality. Based on empirical tuning, for the I-BM3D algorithm we set $\mu=5$ and $\rho=0.16$ for Pavia University, $\mu=0.06$ and $\rho=0.2$ for Stuff_Toy, and $\mu=0.65$ and $\rho=0.09$ for Urban dataset.

When the HRI ground-truth is unknown, the regularization parameters μ and ρ can be tuned, e.g. as suggested in [22], according to the RMSE between the observed HSI or MSI and the corresponding estimated HSI or MSI, which can be defined from the estimated HRI by the degradation matrices.

B. Comparative Benchmarking

Table II summarizes the objective results averaged over ten independent noise realizations for each case. It can be observed that R1BM3D outperforms previous methods by up to 3 dBs in the global PSNR, where the improvement mainly comes from the gain in the spectral domain as shown with the spectral PSNR.

To visualize the spectral reconstruction quality improvement reported in Table II, Fig. 5 shows the spectral signatures and difference in absolute value with respect to the ground-truth for a random spatial location P1 in the reconstructions for each fused image. These results show that the R1BM3D spectral reconstructions are more accurate than those of the comparison methods. Figure 6 shows an RGB mapping of the obtained reconstructions, where the spatial improvement can be visualized in the zoomed sub-region, especially in the reconstruction of smooth regions.

Figure 7 illustrates a comparison of quality with finer granularity, as determined by the contrast of individual ground-truth blocks. It can be observed that although the rank-one similarity prior is primarily verified on high-contrast blocks, R1BM3D shows significant improvements also for low-contrast blocks, which are more impacted by noise, and their reconstruction can be challenging for the comparison methods. The superior performance is verified for both global PSNR and spectral PSNR.

C. Semi-blind Comparison

This section evaluates the performance of R1BM3D for the semi-blind case, where the spatial degradation operator is unknown. In this scenario, the HSI-MSI fusion can be addressed with our approach R1BM3D by estimating the spatial degradation with an effective literature algorithm such as the presented in [12]. We compare the results against the publicly available semi-blind versions for STEREO and SCOTT methods denoted by BSTEREO and BSCOTT, respectively. We employed the Pavia University dataset, where the MSI and HSI were simulated as degraded downsampled versions of the HRI as explained in IV-2. The regularization parameters were tuned to $F=200$ for BSTEREO, $R_1=R_2=60$ and $R_3=2$ for BSCOTT, and $\mu=1$ and $\rho=0.016$ for R1BM3D. Table III summarizes the objective results averaged over ten realizations for each case, varying the noise. It can be observed that the proposed R1BM3D fusion provides an improvement of up to 4 dBs in the semi-blind scenario.

D. Rank analysis

We also analyze the number of PCs that are preserved at each iteration. This number is directly related to the underlying HRI rank, which is expected to be low. Figure 8 shows the nuclear norm, the number of preserved PCs, and the global PSNR as the iterations progress. It can be observed that the number of preserved PCs converges to a relatively small number, inducing a low-rank structure. Notice how the nuclear norm fluctuates opposite to the PSNR; this is most noticeable

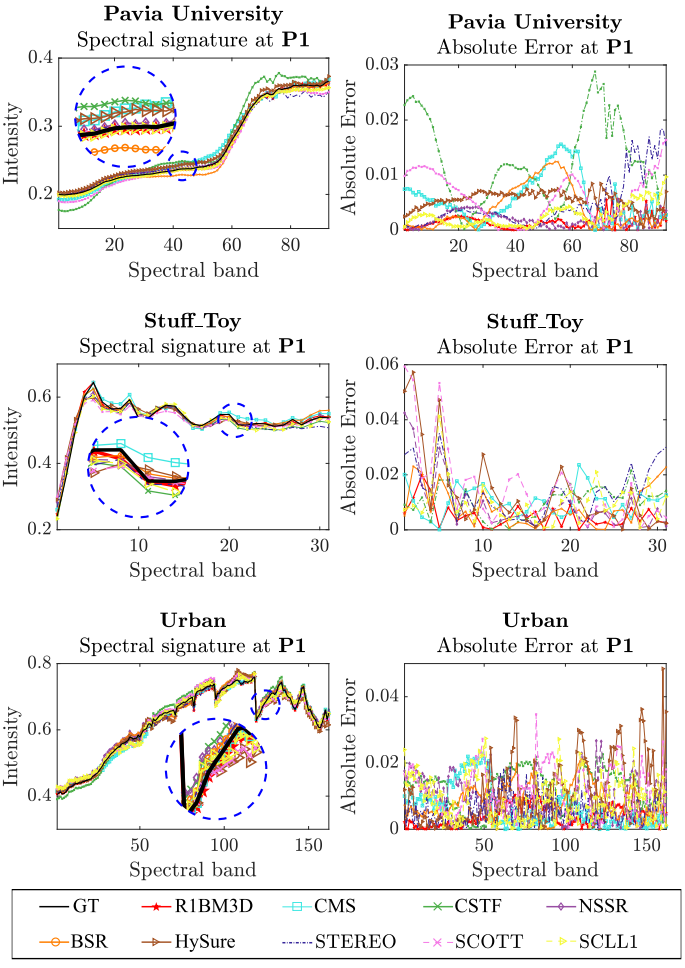

Figure 5. Comparison of the recovered spectral signatures at a random spatial location P1 for each dataset. The absolute error plots confirm that the spectral signatures obtained by the R1BM3D method are more accurate than those obtained by the comparison methods.

Table II
AVERAGE QUALITY AND STANDARD DEVIATION OVER 10 REPLICATIONS CHANGING THE RANDOM NOISE.
HSI-MSI FUSION METRICS: RMSE, ERGAS, SAM (IN DEGREES), UIQI, PSNR (IN DB) AND RUNNING TIME (IN SECONDS)

Pavia University								
Methods	RMSE ↓	ERGAS ↓	SAM ↓	UIQI ↑	Global PSNR ↑	Spatial PSNR ↑	Spectral PSNR ↑	Time [s] ↓
HySure [12]	2.323 ± 0.015	0.849 ± 0.006	1.506 ± 0.016	$0.992 \pm 8.094 \times 10^{-5}$	39.524 ± 0.055	39.754 ± 0.057	32.706 ± 0.075	74.21 ± 14.91
BSR [11]	2.363 ± 0.004	0.846 ± 0.001	1.493 ± 0.003	$0.992 \pm 2.995 \times 10^{-5}$	39.399 ± 0.014	39.802 ± 0.012	32.746 ± 0.016	43.54 ± 5.710
NSSR [33]	2.341 ± 0.010	0.832 ± 0.003	1.495 ± 0.006	$0.992 \pm 8.519 \times 10^{-5}$	41.835 ± 0.482	40.285 ± 0.039	32.753 ± 0.034	145.9 ± 19.02
CSTF [14]	2.432 ± 0.035	0.870 ± 0.010	1.576 ± 0.039	$0.991 \pm 2.793 \times 10^{-4}$	41.706 ± 0.329	39.915 ± 0.104	32.389 ± 0.183	29.29 ± 3.250
CMS [37]	3.171 ± 0.010	1.090 ± 0.002	2.109 ± 0.009	$0.984 \pm 8.442 \times 10^{-5}$	39.440 ± 0.502	38.327 ± 0.015	30.387 ± 0.026	181.3 ± 24.22
NLTD [19]	$\underline{2.114}$	$\underline{0.756}$	$\underline{1.298}$	0.985	41.627	–	–	–
STEREO [38]	2.736 ± 0.008	0.978 ± 0.003	1.873 ± 0.005	$0.989 \pm 6.705 \times 10^{-5}$	40.807 ± 0.582	38.904 ± 0.027	31.133 ± 0.023	26.11 ± 6.084
SCOTT [39]	2.663 ± 0.005	0.961 ± 0.002	1.774 ± 0.003	$0.989 \pm 3.698 \times 10^{-5}$	40.964 ± 0.601	39.070 ± 0.026	31.558 ± 0.013	$\mathbf{2.05 \pm 0.18}$
SC-LL1 [22]	2.164 ± 0.072	0.769 ± 0.024	1.325 ± 0.051	$\underline{0.993} \pm 5.450 \times 10^{-4}$	$\underline{42.639} \pm 0.288$	$\underline{41.022} \pm 0.228$	34.089 ± 0.207	83.99 ± 6.340
I-BM3D	3.648 ± 0.013	1.337 ± 0.008	2.290 ± 0.015	$0.979 \pm 1.960 \times 10^{-4}$	38.056 ± 0.5836	36.883 ± 0.044	30.556 ± 0.071	2219 ± 114.7
R1BM3D	$\mathbf{1.960} \pm 0.014$	$\mathbf{0.697} \pm 0.007$	$\mathbf{1.188} \pm 0.005$	$\mathbf{0.994} \pm 1.042 \times 10^{-4}$	$\mathbf{43.673} \pm 0.214$	$\mathbf{41.858} \pm 0.099$	$\mathbf{35.198} \pm 0.042$	588.4 ± 58.28
Stuff_Toy								
Methods	RMSE ↓	ERGAS ↓	SAM ↓	UIQI ↑	Global PSNR ↑	Spatial PSNR ↑	Spectral PSNR ↑	Time [s] ↓
HySure [12]	3.858 ± 0.102	0.437 ± 0.013	3.775 ± 0.202	$0.994 \pm 2.547 \times 10^{-4}$	35.294 ± 0.228	37.309 ± 0.149	24.899 ± 0.360	204.17 ± 55.51
BSR [11]	3.080 ± 0.007	0.318 ± 0.001	2.643 ± 0.008	$0.997 \pm 1.223 \times 10^{-5}$	37.248 ± 0.019	37.209 ± 0.020	27.581 ± 0.016	143.01 ± 28.59
NSSR [33]	3.108 ± 0.049	0.328 ± 0.005	2.433 ± 0.035	$0.997 \pm 9.114 \times 10^{-5}$	37.170 ± 0.137	37.570 ± 0.168	28.108 ± 0.115	169.76 ± 29.95
CSTF [14]	$\underline{2.941} \pm 0.070$	$\underline{0.309} \pm 0.008$	$\underline{2.094} \pm 0.066$	$0.997 \pm 2.760 \times 10^{-4}$	$\underline{37.649} \pm 0.205$	$\underline{38.080} \pm 0.161$	$\underline{29.846} \pm 0.206$	16.08 ± 2.36
CMS [37]	3.197 ± 0.018	0.328 ± 0.002	3.007 ± 0.023	$0.997 \pm 2.520 \times 10^{-5}$	36.922 ± 0.050	36.681 ± 0.053	27.045 ± 0.054	505.12 ± 100.4
STEREO [38]	4.349 ± 0.139	0.485 ± 0.020	3.182 ± 0.131	$0.991 \pm 7.312 \times 10^{-4}$	34.252 ± 0.277	35.571 ± 0.102	26.929 ± 0.385	$\underline{5.851} \pm 1.30$
SCOTT [39]	5.135 ± 0.004	0.588 ± 0.001	5.298 ± 0.006	$0.990 \pm 2.695 \times 10^{-5}$	32.807 ± 0.007	35.167 ± 0.004	22.521 ± 0.006	$\mathbf{1.350 \pm 0.13}$
SC-LL1 [22]	3.375 ± 0.325	0.346 ± 0.033	2.691 ± 0.196	$\underline{0.997} \pm 6.871 \times 10^{-4}$	36.491 ± 0.816	36.761 ± 0.677	27.833 ± 0.540	537.66 ± 57.70
I-BM3D	7.083 ± 0.131	0.717 ± 0.013	3.613 ± 0.490	$0.989 \pm 3.857 \times 10^{-4}$	30.014 ± 0.159	29.779 ± 0.1544	24.798 ± 1.069	18228 ± 304.4
R1BM3D	$\mathbf{2.183} \pm 0.005$	$\mathbf{0.226} \pm 0.001$	$\mathbf{1.252} \pm 0.006$	$\mathbf{0.999} \pm 6.748 \times 10^{-6}$	$\mathbf{40.237} \pm 0.018$	$\mathbf{40.638} \pm 0.021$	$\mathbf{35.401} \pm 0.037$	5893 ± 334
Urban								
Methods	RMSE ↓	ERGAS ↓	SAM ↓	UIQI ↑	Global PSNR ↑	Spatial PSNR ↑	Spectra PSNR ↑	Time [s] ↓
HySure [12]	3.119 ± 0.086	0.497 ± 0.014	1.373 ± 0.046	$0.989 \pm 2.257 \times 10^{-4}$	36.173 ± 0.235	35.029 ± 0.185	34.144 ± 0.208	101.57 ± 17.28
BSR [11]	$\underline{2.662} \pm 0.005$	$\underline{0.422} \pm 0.001$	$\underline{1.145} \pm 0.003$	$\underline{0.992} \pm 2.784 \times 10^{-5}$	$\underline{37.546} \pm 0.015$	$\underline{36.354} \pm 0.014$	$\underline{35.679} \pm 0.019$	57.96 ± 10.4
NSSR [33]	3.088 ± 0.009	0.492 ± 0.002	1.423 ± 0.004	$0.989 \pm 6.836 \times 10^{-5}$	36.255 ± 0.025	35.010 ± 0.028	34.034 ± 0.025	185.59 ± 32.49
CSTF [14]	3.192 ± 0.096	0.510 ± 0.001	1.412 ± 0.003	$0.988 \pm 3.539 \times 10^{-4}$	35.987 ± 0.016	34.687 ± 0.012	33.919 ± 0.017	16.63 ± 4.02
CMS [37]	3.195 ± 0.005	0.519 ± 0.001	1.434 ± 0.003	$0.988 \pm 4.027 \times 10^{-5}$	35.960 ± 0.014	34.532 ± 0.013	33.936 ± 0.015	437.48 ± 67.33
STEREO [38]	3.220 ± 0.005	0.522 ± 0.001	1.459 ± 0.003	$0.987 \pm 3.399 \times 10^{-5}$	35.892 ± 0.014	34.464 ± 0.014	33.803 ± 0.017	$\underline{40.13} \pm 5.19$
SCOTT [39]	3.369 ± 0.002	0.544 ± 0.001	1.604 ± 0.001	$0.987 \pm 1.299 \times 10^{-5}$	35.500 ± 0.005	34.235 ± 0.006	33.222 ± 0.006	$\mathbf{3.49 \pm 0.1}$
SC-LL1 [22]	3.531 ± 0.085	0.559 ± 0.010	1.704 ± 0.046	$0.985 \pm 6.267 \times 10^{-4}$	35.095 ± 0.208	34.026 ± 0.116	32.712 ± 0.222	238.60 ± 23.84
I-BM3D	2.722 ± 0.004	0.427 ± 0.001	1.263 ± 0.003	$0.991 \pm 2.186 \times 10^{-5}$	37.353 ± 0.012	36.244 ± 0.013	35.307 ± 0.023	4412 ± 159.5
R1BM3D	$\mathbf{2.544} \pm 0.0030$	$\mathbf{0.402} \pm 0.001$	$\mathbf{1.097} \pm 0.002$	$\mathbf{0.992} \pm 1.813 \times 10^{-5}$	$\mathbf{37.944} \pm 0.010$	$\mathbf{36.738} \pm 0.012$	$\mathbf{36.741} \pm 0.014$	1216.7 ± 20.34

Table III
SEMI-BLIND FUSION - AVERAGE QUALITY AND STANDARD DEVIATION OVER 10 REPLICATIONS CHANGING THE RANDOM NOISE.
HSI-MSI FUSION METRICS: RMSE, ERGAS, SAM (IN DEGREES), UIQI, PSNR (IN DB) AND RUNNING TIME (IN SECONDS)

Pavia University								
Methods	RMSE ↓	ERGAS ↓	SAM ↓	UIQI ↑	Global PSNR ↑	Spatial PSNR ↑	Spectral PSNR ↑	Time [s] ↓
BSTEREO [38]	3.159 ± 0.006	1.148 ± 0.002	2.128 ± 0.006	$0.985 \pm 6.842 \times 10^{-4}$	36.877 ± 0.018	37.184 ± 0.016	30.092 ± 0.022	$\underline{6.375} \pm 0.075$
BSCOTT [39]	4.121 ± 0.002	1.583 ± 0.001	2.461 ± 0.002	$0.973 \pm 2.874 \times 10^{-5}$	34.569 ± 0.004	35.196 ± 0.006	29.554 ± 0.008	$\mathbf{0.198 \pm 0.005}$
R1BM3D	$\mathbf{2.768} \pm 0.063$	$\mathbf{0.971} \pm 0.021$	$\mathbf{2.011} \pm 0.049$	$\mathbf{0.989} \pm 1.252 \times 10^{-4}$	$\mathbf{38.028} \pm 0.198$	$\mathbf{38.708} \pm 0.141$	$\mathbf{30.752} \pm 0.226$	1022 ± 95.6

during the first stages of recovery when the filtering suppresses spurious and erroneous structures from the HRI estimate.

V. DISCUSSION

The presented fusion approach takes advantage of a rank-one similarity prior, under which the dissimilarity maps of high-contrast reference blocks are assumed to be proportional to each other across the spectral domain. Nonetheless, the analysis reported in Fig. 7 shows that the quality gap in favour of the proposed method can be very significant also for low-contrast blocks, which is perhaps unexpected. However, low-contrast areas are more impacted by noise, and their reconstruction can be challenging for the comparison methods. As illustrated in Fig. 3, the strategy of determining the dissimilarity of blocks on the first PC is particularly effective under noise and is especially important for the recovery of low-contrast blocks. Using the rank-one prior in R1BM3D helps with guiding the block-matching under heavy noise, and therefore the quality is improved. This is confirmed in Fig. 7, where M-BM3D outperforms independent filtering by I-BM3D over the first r PCs.

In connection to our work, it is interesting to mention the hyperspectral denoising method FastHyDe [45]. While FastHyDe is also based on a low-rank approximation of the HRI and self-similarity via BM3D, the combination of these principles with the rank-one similarity prior proposed here is stronger and more efficient than the one adapted by FastHyDe. Specifically, in FastHyDe it is assumed that each PC is internally self-similar and the denoising is applied separately on each PCs, implying that the block-matching is operated independently on each PC, which leads to increased computation and can also be impaired by the low SNR of some PCs. R1BM3D is instead a multichannel approach, where through the rank-one self-similarity prior we operate the block-matching only once, on the PC with the highest SNR, and reuse the matching positions for all PCs; this leads to computational savings and substantial benefit for content with low SNR and low-contrast.

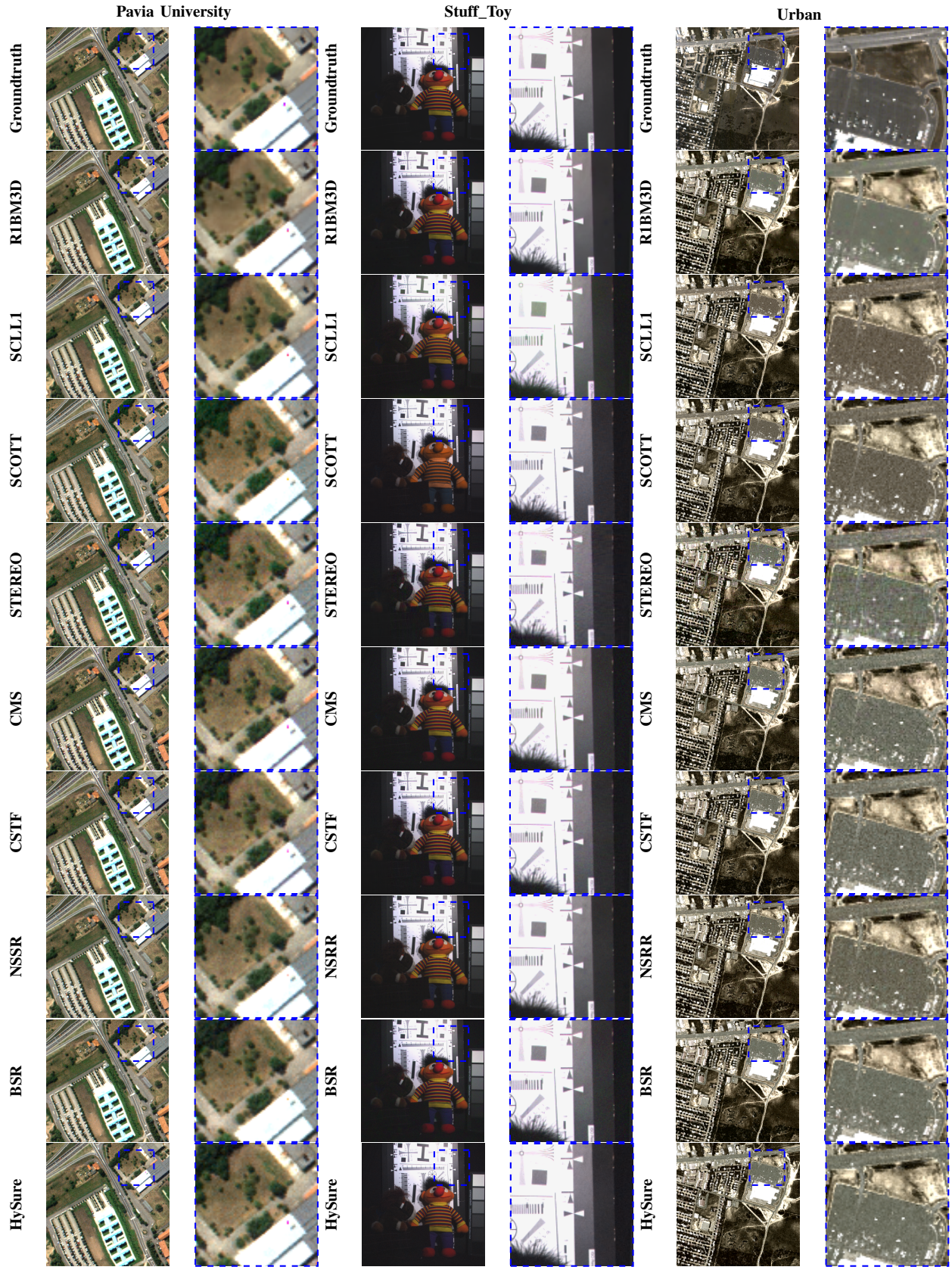

Figure 6. RGB mapping of the recovered Pavia University, Stuff_Toy, and Urban datasets by using the R1BM3D and comparative benchmarking. Notice in the zoomed versions that the fused image obtained with R1BM3D presents a better reconstruction than the other methods, in terms of both detail and color preservation of fine structures and noise reduction at smooth regions.

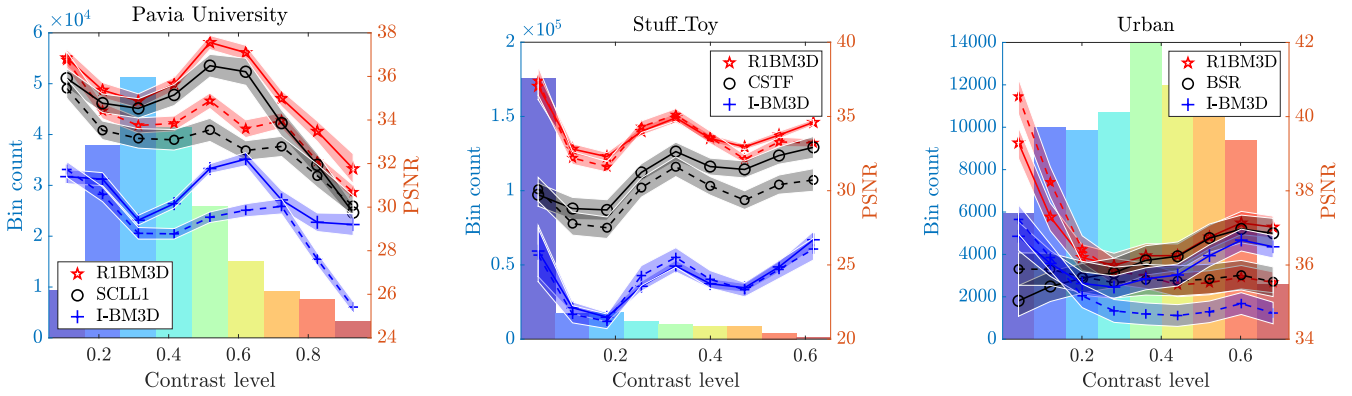

Figure 7. Analysis of fusion quality vs block contrast. The PSNR is computed for each block of size $p\times p=8\times 8$ within a fused image and collected into nine bins according to the contrast (3) of the block at the corresponding position in the ground-truth image. The bar histogram shows the block bin count, confirming that high-contrast blocks are not frequent. The plots show the average PSNR over each contrast bin, where the solid lines correspond to the global PSNR and the dashed lines correspond to the spectral PSNR; the shaded areas around each plot visualize the standard deviation of the average PSNR over ten independent noise realizations for each case. We compare the proposed R1BM3D with the best comparison method for each dataset (SCLL1 for Pavia University, CSTF for Stuff_Toy, and BSR for Urban), and with the I-BM3D algorithm. Although the rank-one similarity prior is primarily verified on high-contrast blocks, R1BM3D shows its superiority also for bins corresponding to low-contrast blocks, where the relative impact of noise is higher and the reconstruction can be challenging for the comparison methods. This advantage can be observed in terms of both global PSNR and spectral PSNR.

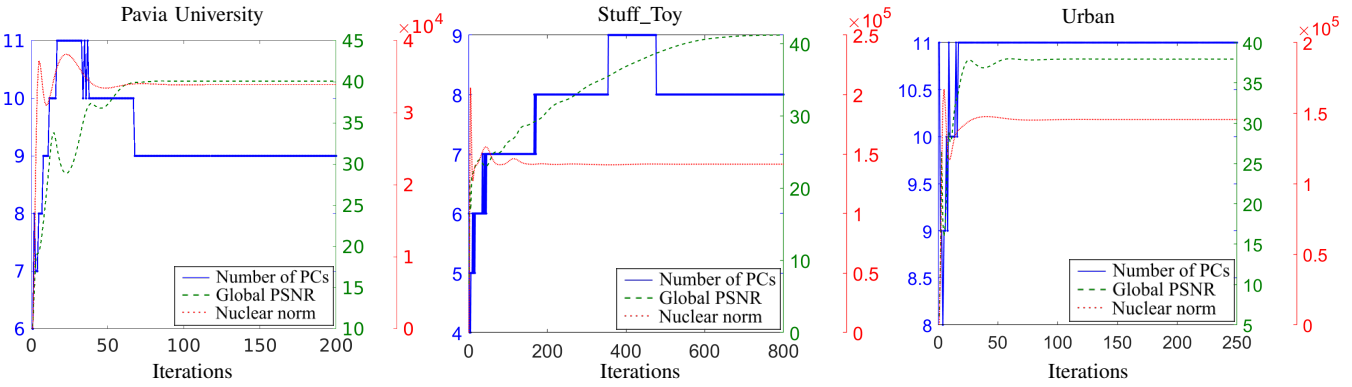

Figure 8. Analysis of the number of preserved PCs during the iterations of R1BM3D. The plots show the behaviour of the number of preserved PCs that are considered significant, the nuclear-norm of the estimated HRI, and the obtained global PSNR across the iterations. After some iterations, the number of PCs converges to a small number for the three datasets. This behavior was expected because of the low-rank property. Furthermore, it can be observed a counter-relation between the nuclear norm and the global PSNR during the first few iterations, in which a high value in the nuclear norm implies low quality. These results indicate that the information removed during the first iterations effectively corresponds to noise.

Furthermore, it is interesting to discuss the differences between our approach and recent model-based tensor approaches that maintain the three-dimensional natural structure of spectral scenes. For instance, [38] and [39] present formulations based on the canonical polyadic decomposition (CPD) and the Tucker decomposition, respectively. These methods take advantage of the tensor-based model to provide identifiability and recoverability guarantees under mild conditions. Meanwhile, our approach is not based only on leveraging the natural 3D structure/regularity of the images, but also incorporates the nonlocal self-similarity within each PC and the rank-one similarity prior across all PCs. We wish to mention also [22], which uses a block-term decomposition with a multi-linear rank model, where the latent factors can be interpreted as the HRI endmembers and abundances of the spectral linear mixture model. Differently, we formulate a regularization function that considers jointly the spectral, the nonlocal self-similarity, and the introduced rank-one similarity priors, inducing the low-rank property over different dimensions.

Various methods such as [46]–[50] have integrated low-dimensional reduction with the filtering of 3D cubes to promote the low-rank property. All these methods fundamentally differ from ours. Firstly, in the definition of the rank-one similarity prior, the proposed method defines non-locality with respect to 2D patches instead of 3D cubes, and specifically M-BM3D operates shrinkage independently on groups of 2D patches over different channels, which is more flexible than assuming that nonlocal 4D groups composed of 3D cubes are *jointly* sparse. Secondly, in the way the prior is leveraged by the algorithm, the proposed R1BM3D is more efficient, as by the rank-one similarity prior the positions of similar patches are propagated across the spectral bands, avoiding the computational burden of testing similarity by computing differences and norms over the entire 3D cubes.

VI. Conclusion

We presented an HSI-MSI fusion termed R1BM3D that combines the spectral low-rank, the nonlocal self-similarity,

and an introduced rank-one similarity prior. The rank-one similarity prior is introduced as an intrinsic HRI characteristic under which the structural similarities of small blocks are shared across all spectral bands. We conducted an extensive empirical study over several real-world objects and remote-sensing and satellite datasets, finding support for the rank-one similarity prior. We developed an effective algorithm that uses a nonlocal patch-based denoiser to promote the proposed low-rank priors. The experiments show that R1BM3D can significantly improve the HSI-MSI fusion, with the most remarkable improvement observed in the recovery of the spectral information at low-contrast locations.

REFERENCES

[1] H. Rueda-Chacon, F. Rojas, D. R. Molina, and H. Arguello, "Demonstration of a compressive hyperspectral image fusion optical imager," in *Computational Optical Sensing and Imaging*. Optical Society of America, 2020, pp. CW4B–6.

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

[3] J. M. Ramirez, J. I. M. Torre, and H. Arguello, "Feature fusion via dual-resolution compressive measurement matrix analysis for spectral image classification," *Signal Processing: Image Communication*, vol. 90, p. 116014, 2021.

[4] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5345–5355, 2018.

[5] X. Fu, W. Wang, Y. Huang, X. Ding, and J. Paisley, "Deep multiscale detail networks for multiband spectral image sharpening," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 2090–2104, 2020.

[6] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1585–1594.

[7] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sensing*, vol. 10, no. 5, p. 800, 2018.

[8] W. Wang, W. Zeng, Y. Huang, X. Ding, and J. Paisley, "Deep blind hyperspectral image fusion," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4150–4159.

[9] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-level and multi-scale spatial and spectral fusion cnn for hyperspectral image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[10] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Information Fusion*, vol. 69, pp. 40–51, 2021.

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

[12] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2014.

[13] I. Ortiz, S. Rivera, T. Gelvez, F. Rojas, and H. Arguello, "Hyperspectral-multispectral image fusion with rank estimation by using a joint-sparse regularizer," in *2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE, 2021, pp. 1–6.

[14] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.

[15] R. Wu, W.-K. Ma, X. Fu, and Q. Li, "Hyperspectral super-resolution via global local low-rank matrix estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7125–7140, 2020.

[16] B. Du, Z. Huang, N. Wang, Y. Zhang, and X. Jia, "Joint weighted nuclear norm and total variation regularization for hyperspectral image denoising," *International Journal of Remote Sensing*, vol. 39, no. 2, pp. 334–355, 2018.

[17] J. Huang, T.-Z. Huang, L.-J. Deng, and X.-L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2419–2438, 2018.

[18] C. Yi, Y.-Q. Zhao, J. C.-W. Chan, and S. G. Kong, "Joint spatial-spectral resolution enhancement of multispectral images with spectral matrix factorization and spatial sparsity constraints," *Remote Sensing*, vol. 12, no. 6, p. 993, 2020.

[19] K. Wang, Y. Wang, X.-L. Zhao, J. C.-W. Chan, Z. Xu, and D. Meng, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor decomposition and spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[20] X. Li, Y. Yuan, and Q. Wang, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor approximation and sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[21] Y. Chang, L. Yan, B. Chen, S. Zhong, and Y. Tian, "Hyperspectral image restoration: Where does the low-rank property exist," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[22] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral super-resolution via interpretable block-term tensor modeling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 641–656, 2020.

[23] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[24] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.

[25] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play admm for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2016.

[26] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 408–423, 2016.

[27] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. Figueiredo, "A convergent image fusion algorithm using scene-adapted gaussian-mixture-based denoising," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 451–463, 2018.

[28] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[29] T. Gelvez-Barrera, "Hyperspectral datasets ONLINE," https://github.com/TatianaGelvez/Datasets.git.

[30] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum," https://www.cs.columbia.edu/CAVE/databases/multispectral/, Tech. Rep., Nov 2008.

[31] B. A. M Graña, MA Veganzons, "Hyperspectral remote sensing scenes ONLINE," http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

[32] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[33] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[34] H. J. Kramer, *Observation of the Earth and its Environment: Survey of Missions and Sensors*. Springer Science & Business Media, 2002.

[35] R. Feng, L. Wang, and Y. Zhong, "Least angle regression-based constrained sparse unmixing of hyperspectral remote sensing imagery," https://rslab.ut.ac.ir/data, p. 1546, 09 2018.

[36] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5412–5427, Dec 2014.

[37] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5969–5982, 2018.

[38] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6503–6517, 2018.

[39] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled tucker approximation: Recoverability and svd-based algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 931–946, 2020.

[40] Junjun-Jiang, "Hyperspectral image super resolution benchmark ONLINE," https://github.com/junjun-jiang/Hyperspectral-Image-Super-Resolution-Benchmark.
[41] C. Kanatsoulis, "Hsr via tensor decomposition ONLINE," https://github.com/marhar19/HSR_via_tensor_decomposition.
[42] C. Prévost, "Software for hyperspectral super-resolution with coupled tucker approximation ONLINE," https://github.com/cprevost4/HSR_Software.
[43] M. Ding, "Code for scll1 hyperspectral super-resolution ONLINE," https://github.com/MengDing56/Code_SCLL1_HSR.
[44] T. Gelvez and H. Arguello, "Nonlocal low-rank abundance prior for compressive spectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
[45] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 730–742, 2018.
[46] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang, "Non-local meets global: An iterative paradigm for hyperspectral image restoration," *arXiv preprint arXiv:2010.12921*, 2020.
[47] G. Chen, T. D. Bui, K. G. Quach, and S.-E. Qian, "Denoising hyperspectral imagery using principal component analysis and block-matching 4d filtering," *Canadian Journal of Remote Sensing*, vol. 40, no. 1, pp. 60–66, 2014.
[48] P. Xu, B. Chen, L. Xue, J. Zhang, L. Zhu, and H. Duan, "A new mnf–bm4d denoising algorithm based on guided filtering for hyperspectral images," *ISA transactions*, vol. 92, pp. 315–324, 2019.
[49] L. Sun and B. Jeon, "Hyperspectral mixed denoising via subspace low rank learning and bm4d filtering," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 8034–8037.
[50] R. Kurihara, S. Ono, K. Shirai, and M. Okuda, "Hyperspectral image restoration based on spatio-spectral structure tensor regularization," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 488–492.

Tatiana Gelvez-Barrera (S'17) received the B.S. degree in industrial engineering and systems engineering from the Universidad Industrial de Santander, Colombia, in 2016, and the Ph.D. degree in Engineering from the Universidad Industrial de Santander, Colombia, in 2022. During the second semester of 2019 and 2020 she was an intern at Tampere University, Tampere, Finland. Her research interests include numerical optimization, high-dimensional signal processing, spectral imaging, computational imaging, deep-learning for image processing, and compressive sensing.

Henry Arguello (S'11–M'13–SM'17) received the B.Sc. Eng. degree in electrical engineering and the M.Sc. degree in electrical power from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from the University of Delaware, Newark, DE, USA, in 2013. He is currently an Associate Professor with the Department of Systems Engineering, Universidad Industrial de Santander. In first semester 2020, he was a Visiting Professor with Stanford University, Stanford, CA, USA, funded by Fulbright. His research interests include high-dimensional signal processing, optical imaging, compressed sensing, hyperspectral imaging, and CI.

Alessandro Foi received the M.Sc. degree in Mathematics from the Università degli Studi di Milano, Italy, in 2001, the Ph.D. degree in Mathematics from the Politecnico di Milano in 2005, and the D.Sc.Tech. degree in Signal Processing from Tampere University of Technology, Finland, in 2007. He is Professor of Signal Processing at Tampere University, Finland, and the Director of Tampere University Imaging Research Platform. His research interests include mathematical and statistical methods for signal processing, functional and harmonic analysis, and computational modeling of the human visual system. His work focuses on spatially adaptive (anisotropic, nonlocal) algorithms for the restoration and enhancement of digital images, on noise modeling for imaging devices, and on the optimal design of statistical transformations for the stabilization, normalization, and analysis of random data. He is the Editor-in-Chief of the IEEE Transactions on Image Processing. He previously served as a Senior Area Editor for the IEEE Transactions on Computational Imaging and as an Associate Editor for the IEEE Transactions on Image Processing, the SIAM Journal on Imaging Sciences, and the IEEE Transactions on Computational Imaging. He is a Fellow of the IEEE.