Robust Hyperspectral Image Fusion with Simultaneous Guide Image Denoising via Constrained Convex Optimization

Saori Takeyama Member, IEEE and Shunsuke Ono, Member, IEEE

Abstract—The paper proposes a new high spatial resolution hyperspectral (HR-HS) image estimation method based on convex optimization. The method assumes a low spatial resolution HS (LR-HS) image and a guide image as observations, where both observations are contaminated by noise. Our method simultaneously estimates an HR-HS image and a noiseless guide image, so the method can utilize spatial information in a guide image even if it is contaminated by heavy noise. The proposed estimation problem adopts hybrid spatio-spectral total variation as regularization and evaluates the edge similarity between HR-HS and guide images to effectively use apriori knowledge on an HR-HS image and spatial detail information in a guide image. To efficiently solve the problem, we apply a primal-dual splitting method. Experiments demonstrate the performance of our method and the advantage over several existing methods.

Index Terms—hyperspectral image fusion, pansharpening, multispectral image, total variation, primal-dual splitting method.

I. INTRODUCTION

HIGH-RESOLUTION spectral information helps reveal the intrinsic characteristics of objects and environment lighting. Hyperspectral (HS) imaging can capture spectral information from 400nm to 2500nm in the 5-10nm interval and is expected to resolve many problems in a wide range of fields, e.g., remote sensing, biomedical, industrial, and food safety [1–3]. Owing to high-resolution spectral information, HS imaging is limited to the amount of incident energy in each band, so it has critical tradeoffs between the spatial resolution, the spectral resolution, and the signal-to-noise ratio. Many applications require high spatial and spectral resolution HS (HR-HS) images without noise, so it is an essential task to resolve the problem.

In general, observed HS images have a high spectral resolution but a low spatial resolution (called LR-HS images) because the spatial detail information is lost through an imaging process. Therefore, fusion techniques have been widely studied to estimate HR-HS images [4, 5]. This technique fuses two observations of an LR-HS image and a guide image that has high spatial resolution but low spectral one. Here, the guide image is assumed a panchromatic (PAN) image or a multispectral (MS) image, and the fusion technique with a PAN image is especially called HS pansharpening. This fusion technique has been known to effectively estimate HR-HS images because it can utilize both the high spatial resolution information of the guide image and the high spectral resolution information of the LR-HS image.

Many fusion methods have been proposed, which can be roughly classified into three groups: component analysis-based methods, optimization-based methods, and deep learning-based methods. Component analysis-based methods [6–12] separate spatial details from a guide image using a linear transformation, replace the spatial information of an LR-HS image with the extracted spatial details, and then apply the inverse transformation to the fused image. They are easy implementation and have low computational cost but the performance is limited in especially noisy cases because they cannot directly handle apriori knowledge on HR-HS images.

In contrast, optimization-based approaches can evaluate apriori knowledge on HR-HS images using a regularization function. Many methods [13–20] assume the low rankness of HR-HS images and model an HR-HS image as the product of endmember spectra and abundance maps. Here, endmember and abundance have spectral and spatial information of an HR-HS image, respectively. Then, the methods estimate the endmember and the abundance by independently evaluating spatial and spectral apriori knowledge. The methods achieve an estimation preserving spectral information but produce spatial artifacts especially when noisy cases. This is because they do not directly evaluate spatial edge information in a guide image even though a guide image has high-resolution spatial detailed information. Besides, the performance relies on the rank number of the endmember that cannot be known, and the setting error sufficiently affects their estimation performance. In other words, it is an essential and a difficult tasks for the methods to decide the suitable rank number.

Some optimization-based methods [21, 22] evaluate spatial detail information to estimate an HR-MS image, so they can estimate a spatial detail-preserving HR-MS image. In [21], [22], they can estimate a spatial detail-preserving HR-MS image from an MS image with less than 10 bands. The two methods assume a noiseless guide image, so the performance would be degraded for a noisy one. Moreover, they use a nonconvex function or do not directly evaluate spectral information, so they require improvement to achieve the estimation unaffected by the initial value and preserving spectral information if they are used for HR-HS image estimation.
Deep learning has been recently known as a powerful approach to image processing, so some deep learning-based fusion methods have been proposed [23]–[27]. The methods train the intrinsic characteristics of HR-HS images and use the characteristics for HR-HS image estimation. For the training process, they extract latent properties from a large number of HR-HS images, thus they achieve high estimation performance. However, capturing a large number of HR-HS images is very difficult, and in fact, the number of HS datasets is limited. Moreover, the training HR-HS images are required to have the same number of bands as an observed LR-HS image, so they need re-training when the methods are used on another LR-HS image with a different number of bands. To solve the dilemma, CNN-Fus [28] applies a convolution neural network (CNN) pre-trained with grayscale images. CNN-Fus can evaluate a spatial underlying characteristics of an HR-HS image and achieves high performance. However, the performance depends on the iteration number as in Fig. 1, so the setting is an essential task. In [29], the authors manually determine the iteration number at which the quality measure is stable, but one cannot know the suitable iteration number in a real estimation. Hence, CNN-Fus faces the problem of how determining the iteration number in real observations.

In this paper, we propose a new HR-HS image estimation method from LR-HS and guide images based on optimization, which preserves spatial edge and is robust to noise included in observations. Our proposed method estimates both an HR-HS image and a noiseless guide image, so it is rarely affected by noise in a guide image. Then, our method evaluates the edge similarity between HR-HS and guide images to take over the spatial edge in a guide image to an HR-HS image. To effectively evaluate apriori knowledge on an HR-HS image, the method adopts hybrid spatio-spectral total variation (HSSTV) [29] as a regularization function. The optimization problem is convex, and it can be efficiently solved by a primal-dual splitting method. In experiments, we verified the performance of our proposed method and confirm the advantage over several existing fusion methods.

Our paper makes the following contributions:

- The method simultaneously estimates an HR-HS image and a noiseless guide image to achieve robust HR-HS image estimation to noise included in an LR-HS image and an observed guide image. The experiments illustrate that our method can estimate an HR-HS image without spatial artifacts when observations are contaminated by noise.

- To effectively utilize the spatial detailed information in a guide image, our method evaluates the edge similarity between an HR-HS image and a guide image. Thanks to the setting, the method achieves high performance for HS pansharpening and HS and MS image fusion.

The rest of this article is organized as follows. Section II reviews several existing optimization-based fusion methods in detail. In Section III, we introduce mathematical ingredients. Our proposed method is described in Section IV. Experimental results and comparisons on HS pansharpening and HS and MS image fusion are presented in Section V. Section VI concludes this article.

II. RELATED WORKS

We explain several optimization-based existing HR-HS image estimation methods. Tab. I shows the features of several existing methods.

Many optimization-based fusion methods assume that an HS image has low-rank properties and can be decomposed into some bases and coefficients. In HS unmixing, one of the low-rank modeling (LRM)-based approaches, the bases and the coefficients are handled as endmember spectra and abundance maps, respectively. Here, the endmember and the abundance represent the intrinsic spectral response of materials and the existing ratio of each material. From the characteristics, the abundance and the endmember are considered to have spatial and spectral information of an HS image, respectively. Based on the assumption, the methods estimate them from a pair of LR-HS and guide images.

All methods in Tab. I consider that a guide image is an MS image and HySure and Lanaras’s also consider an HS pansharpening case. The methods [14]–[16], [18], [20] only consider an MS guide image, but they only use a guide image to evaluate the data fidelity in abundance estimation. Therefore, by setting the suitable spectral response operator, they can estimate an HR-HS image in the HS pansharpening case. Coupled NMF (CNMF) [17] decomposes an MS image into an LR-endmember and an HR-abundance using nonnegative matrix factorization, and use an HR-abundance to generate an HR-HS image. PAN images do not have spectral information, so estimating an LR-endmember is difficult. However, according to [30], the method can be adopted for HS pansharpening by only estimating HR abundance from a PAN image using an estimated HR-endmember and a spectral response operator of a PAN image.

All methods consider that both observations are contaminated by additive Gaussian noises. Here, in some papers [14], [16], [19], [20], the authors do not clearly model the noise, but all methods evaluate the data fidelity between observations and an HR-HS image. Therefore, we regard them as methods considering noisy observations.

Some methods [13]–[16] separately estimate the endmember and the abundance. The methods first decompose an LR-HS image into HR-endmember and LR-abundance and use HR-endmember. Then, they estimate abundance from the estimated endmember and observations. Since an LR-HS image is generally contaminated by noise, the endmember estimation
TABLE I
THE FEATURES OF EXISTING HR-HS IMAGE ESTIMATION METHODS BASED ON OPTIMIZATION

| method            | guide image | noise  | spatial evaluation | spectral evaluation |
|-------------------|-------------|--------|--------------------|---------------------|
| CNMF [13]         | MS image    | both images | low-rankness       | low-rankness        |
| HySure [14]       | MS or PAN image | both images | piecewise smoothness | low-rankness        |
| NED [16]          | MS image    | both images | low-rankness       | low-rankness        |
| Yang’s [18]       | MS image    | both images | low-rankness       | low-rankness        |
| FUMI [19]         | MS image    | both images | low-rankness       | low-rankness        |
| NCTRF [20]        | MS image    | both images | low-rankness       | low-rankness        |
| Lanaras’s [19]    | MS or PAN image | both images | low-rankness       | low-rankness        |
| LTMK [14]         | MS image    | both images | low-rankness       | low-rankness        |
| proposed          | MS or PAN image | both images | edge similarity    | piecewise smoothness |
|                   |             |         |                    |                     |
| HR-MS image estimation methods
| PSSGDNSP [21]     | PAN image   | only MS image | edge similarity    | edge similarity     |
| VPECT [22]        | PAN image   | only MS image | piecewise smoothness |                  |

is required to be robust. To reduce the influence of noise, the methods [13–15] use singular value decomposition for pre-denosing and/or endmember estimation, and the method [16] estimates the endmember by solving a nonnegative matrix factorization (NMF)-based problem considering noise. For the abundance estimation step, they formulate an optimization problem including regularization and data fidelity and effectively solve it by an iterative algorithm. They adopt different regularization functions, i.e., they utilize the different spatial characteristics on an HR-HS image (See Tab. I).

The methods [17–19] simultaneously estimate the endmember and the abundance. CNMF decomposes LR-HS and guide images into the endmembers and the abundances by NMF and generates an HR-HS image from the endmember of the LR-HS image and the abundance of the guide one. This is because LR-HS and guide images have full spectral and spatial information, respectively. CNMF uses the relationship between the observations for the decomposition and alternatively and iteratively estimates the endmember and the abundance. The two methods [18], [19] formulate and solve optimization problems to simultaneously estimate the endmember and the abundance. They evaluate the sparsity of the abundance by sum-to-one and nonnegative constraints. The method [20] assumes the low-rankness of HS images and considers three low-rank factors. It simultaneously estimates the factors from observations by solving an optimization problem that evaluates the data fidelity and spectral correlation.

The LRM-based methods achieve an HR-HS image estimation preserving spectral information, but they have problems. Although a guide image has spatial edge information, they do not directly evaluate it. Specifically, the methods use spatial information in a guide image when they estimate abundance or low-rank factors, but they only evaluate the data fidelity. Therefore, the methods produce spatial artifacts or oversmoothing. Besides, they require to set the number of the endmember, and the setting is affected by the estimation performance. However, since the suitable number is unknown and different by objects, it is an essential and difficult task to search for it.

In [21], [22], the authors propose an HR-MS image estimation based on other approaches. Pansharpening with spatial and spectral gradient difference-induced nonconvex sparsity priors (PSSGDNSP) [21] uses $\ell_{1/2}$ norm to evaluate the spatial gradient similarity between a guide image and an HR-MS image and the spectral gradient similarity between an LR-HS image and an HR-MS image. This is because the authors experimentally confirm that the similarities are close to hyper-Laplacian. Since the spatial gradient is almost the same as the spatial edge information, PSSGDNSP can take over the spatial information in a guide image to an HR-MS image. However, the method assumes that a guide image is not contaminated by noise, so the performance would significantly decrease when real observations. Moreover, the $\ell_{1/2}$ norm can measure more sparsity than the $\ell_1$ norm but is non-convex. Therefore, the convergence is not guaranteed, and the performance of PSSGDNSP would depend on the initial value. Tian et al. propose a cartoon-texture-based method for MS pansharpening in [22], named a variational pansharpening method by exploiting cartoon-texture similarities (VPECT). VPECT decomposes a PAN image into the cartoon and the texture components, and then VPECT estimates an HR-MS image using the components. From the design, the method can estimate an HR-MS image with spatial detail, but it cannot directly evaluate spectral information. Therefore, the performance would be limited for HR-MS image estimation. Moreover, VPECT assumes a noiseless PAN image to effectively extract the spatial texture component from a PAN image, but the situation is not realistic.

### III. Preliminaries

#### A. Notations and Definitions

Let $\mathbb{R}$ be the set of real numbers. We shall use boldface lowercase and capital to represent vectors and matrices, respectively, and := to define something. We denote the transpose of a vector/matrix by $(\cdot)^\top$ and the Euclidean norm (the $\ell_2$ norm) of a vector by $\| \cdot \|$.

For notational convenience, we treat an HS image $\mathbf{U} \in \mathbb{R}^{N_x \times N_y \times B}$ as a vector $\mathbf{u} \in \mathbb{R}^{N \times B}$ ($N := N_x N_y$ is the number of the pixels of each band, and $B$ is the number of the bands) by stacking its columns on top of one another, i.e., the index of the component of the $r$th pixel in $k$th band is $i + (k-1)N$ for $i = 1, \ldots, N$ and $k = 1, \ldots, B$. 
B. Proximal Tools

A function \( f : \mathbb{R}^N \to (-\infty, \infty] \) is called proper lower semicontinuous convex if \( \text{dom}(f) := \{ x \in \mathbb{R}^N \mid f(x) < \infty \} \neq \emptyset \), \( \text{lev}_{\leq \alpha}(f) := \{ x \in \mathbb{R}^N \mid f(x) \leq \alpha \} \) is closed for every \( \alpha \in \mathbb{R} \), and \( f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \) for every \( x, y \in \mathbb{R}^N \) and \( \lambda \in (0, 1) \), respectively. Let \( \Gamma_0(\mathbb{R}^N) \) be the set of all proper lower semicontinuous convex functions on \( \mathbb{R}^N \).

The proximity operator \cite{rockafellar1970} plays a central role in convex optimization based on proximal splitting. The proximity operator of \( f \in \Gamma_0(\mathbb{R}^N) \) with a index \( \gamma > 0 \) is then defined by

\[
\text{prox}_{\gamma f}(x) := \text{argmin}_y \left\{ f(y) + \frac{1}{2\gamma} \| y - x \|^2 \right\},
\]

where the existence and uniqueness of the minimizer are guaranteed respectively by the coercivity\footnote{A function \( f \in \Gamma_0(\mathbb{R}^N) \) is called coercive if \( \|x\| \to \infty \Rightarrow f(x) \to \infty \). In this case, the existence of a minimizer of \( f \) is guaranteed, that is, there exists \( x^* \in \text{dom}(f) \) such that \( f(x^*) = \inf_{x \in H} f(x) \) \cite{rockafellar1970}.} and the strict convexity of \( f(\cdot) + \frac{1}{2\gamma} \| \cdot \|^2 \).

We introduce the indicator function of a nonempty closed convex set \( C \subset \mathbb{R}^N \), which is defined as follows:

\[
\iota_C(x) := \begin{cases} 
0, & \text{if } x \in C, \\
\infty, & \text{otherwise}. 
\end{cases}
\]

Then, for any \( \gamma > 0 \), the proximity operator is defined by

\[
\text{prox}_{\gamma \iota_C}(x) = P_C(x) := \text{argmin}_y \| x - y \|,
\]

where \( P_C(x) \) is a metric projection onto \( C \), and this proximity operator is reduced to the metric projection onto \( C \).

C. Primal-Dual Splitting Method

The primal-dual splitting method \cite{tseng2008} is based on proximal splitting, and it is an algorithm that can solve convex optimization problems of the form:

\[
\min_x g(x) + h(Lx),
\]

where \( g \in \Gamma_0(\mathbb{R}^a) \), \( h \in \Gamma_0(\mathbb{R}^b) \), and \( L \) is a linear operator. Here, we assume that \( g \) and \( h \) are proximable, i.e., the proximity operators of \( g \) and \( h \) are computable. For arbitrarily chosen \( y^{(0)} \) and \( \gamma_1, \gamma_2 > 0 \) satisfying \( \gamma_1 \gamma_2 \| L \|_{op} \leq 1 \) (\( \| \cdot \|_{op} \) is the operator norm), the primal-dual splitting method iterates the following steps:

\[
\begin{align*}
\lambda^{(n+1)} &= \text{prox}_{\gamma_1 g}(\lambda^{(n)} - \gamma_1 L^\top y^{(n)}), \\
y^{(n+1)} &= \text{prox}_{\gamma_2 h}(y^{(n)} + \gamma_2 L(2 \lambda^{(n+1)} - x^{(n)})). 
\end{align*}
\]

The function \( h^* \) is the convex conjugate of \( h \), and the proximity operator \( h^* \) is available via that of \( h \) \cite{rockafellar1970} Theorem 14.3 (ii)] as follows:

\[
\text{prox}_{\gamma h^*}(x) = x - \gamma \text{prox}_{\frac{1}{\gamma} h} \left( \frac{1}{\gamma} x \right).
\]

IV. PROPOSED METHOD

We propose a new HR-HS image estimation method from a pair of an LR-HS image and a guide image. Let \( u \in \mathbb{R}^{NB} \) be an HR-HS image, and a LR-HS image \( v \) and a guide image \( g \) are given by

\[
\begin{align*}
v &= SBu + n_v \in \mathbb{R}^{NB}, \\
g &= Ru + n_g \in \mathbb{R}^{Nb},
\end{align*}
\]

where \( S \) is a down-sampling matrix with the ratio of \( r^2 \), \( B \) is a blur matrix, \( n_v \) and \( n_g \) are additive white Gaussian noises, and \( b \) is the band number of a guide image. We assume that \( NB \) is divisible by \( r^2 \) because the number of pixels in \( v \) must be a positive integer. In the paper, the guide image is supposed to be an MS image or a PAN image, where the PAN image is a single-channel image and is equal to a grayscale image. The matrix \( R \) represents the spectral response of the guide image and depends on the imaging system. This model considers both observations contaminated by noise, which is a natural situation. In general, since HS images are more affected by noise than MS and PAN images, the noise intensity \( \sigma_v \) and \( \sigma_g \) of \( n_v \) and \( n_g \) satisfy \( \sigma_v > \sigma_g \). This model is based on Wald’s protocol \cite{wald1947}, and existing methods are also based on the same or very similar model.

A. HR-HS Image Estimation

Our proposed method estimates an HR-HS image by solving a convex optimization problem. For robust estimation, the problem simultaneously estimates an HR-HS image \( u \in \mathbb{R}^{NB} \) and a noiseless guide image \( q \in \mathbb{R}^{Nb} \). Our method evaluates apriori knowledge on \( u \) and \( q \) and the edge similarity between the two images, so it can effectively use spatial and spectral information in observations. We formulate the HR-HS image estimation problem as follows:

\[
\begin{align*}
\min_{u, q} & \quad \text{HSSTV}(u) + \lambda \| DM_u u - DM_q \|_{1,2} + \rho \| Dq \|_{1,2} \\
\text{s.t.} & \quad SBu \in B_{\epsilon, u} := \{ x \in \mathbb{R}^{NB} \mid \| x - v \| \leq \epsilon \}, \\
& \quad q \in B_{\gamma, q} := \{ x \in \mathbb{R}^{Nb} \mid \| x - g \| \leq \eta \}, \\
& \quad u \in [\mu_u, \bar{u}], \\
& \quad q \in [\mu_q, \bar{q}].
\end{align*}
\]

where \( D = (D_u^\top D_h^\top)^\top \) is a spatial difference operator with \( D_u \) and \( D_h \) being vertical and horizontal difference operators, respectively. A linear operator \( M_u \in \mathbb{R}^{NB \times NB'} \) extracts the same spectral range information as a guide image. Specifically, when a guide image captures information with spectral range \([\lambda_{g, st}, \lambda_{g, en}]\) \((\lambda_{g, st} < \lambda_{g, en})\), and from \( B_{\lambda_{g, st}, \lambda_{g, en}} \) to \( B_{\lambda_{g, st}, \lambda_{g, en}} \) bands of \( v \) have the spectral range information, \( M_u \) only maintains the information in \([B_{\lambda_{g, st}}, B_{\lambda_{g, en}}] \) and removes the other information. Moreover, \( M \in \mathbb{R}^{NB \times Nb} \) is a normalized transpose matrix of \( R \) in the spectral range included in a guide image. The matrix is calculated by transposing \( R \), normalizing it in the row direction, and extracting a part of the spectral range. This is because the evaluation of the edge similarity in other than the spectral range is not essential and would produce spatial over-smoothing. Since the method assumes a noisy guide image, it estimates both the HR-HS image and the desirable guide image for a noise-robust estimation.
Algorithm 1: Primal-dual splitting algorithm for solving Prob. (9)

\begin{algorithm}
\begin{algorithmic}[1]
\State \textbf{input}: \(u^{(0)}, q^{(0)}, y_1^{(0)}, y_2^{(0)}, y_3^{(0)}, y_4^{(0)}, y_5^{(0)}\)
\While {A stopping criterion is not satisfied}
\State \(u^{(n+1)} = \text{prox}_{\gamma_1 \frac{1}{\mu} D y_1^{(n)} + M} y_1^{(n)} - \gamma_1 (A_y y_1^{(n)} + M y_2^{(n)} + B S y_4^{(n)});\)
\State \(q^{(n+1)} = \text{prox}_{\gamma_2 \frac{1}{\mu} M} q^{(n)} - \gamma_2 (\gamma y_3^{(n)} + M y_2^{(n)} + y_5^{(n)}));\)
\State \(y_1^{(n+1)} = y_1^{(n)} + \gamma_2 A y_2^{(n)} (2 u^{(n+1)} - u^{(n)});\)
\State \(y_2^{(n+1)} = y_2^{(n)} + \gamma_2 (D M y_2^{(n)} - (u^{(n+1)} - u^{(n)}) - (2 q^{(n+1)} - q^{(n)}));\)
\State \(y_3^{(n+1)} = y_3^{(n)} + \gamma_2 D y_2^{(n)} (q^{(n+1)} - q^{(n)});\)
\State \(y_4^{(n+1)} = y_4^{(n)} + \gamma_2 (B u^{(n+1)} - u^{(n)}) - 2 y_5^{(n+1)});\)
\State \(y_5^{(n+1)} = y_5^{(n)} + 2 y_5^{(n+1)} - q^{(n)});\)
\EndWhile
\State \(n \leftarrow n + 1;\)
\end{algorithmic}
\end{algorithm}

The first term in Prob. (9) is a hybrid regularization technique for HS image restoration, named HSSTV [29]. HSSTV can effectively utilize a-priori knowledge of HS images, and according to [29, 35–37], it has achieved high performance for several HS image restorations. HSSTV is defined by

\[ \text{HSSTV}(u) := \|A_y u\|_{1,p} \text{ s.t. } A_y := \begin{bmatrix} D y_1 & \omega D \\ y_2 & y_3 \end{bmatrix}, \] (10)

where \(\|\cdot\|_{1,p}\) is a mixed \(\ell_{1,p}\) norm, and \(D y_1\) is a spectral difference operator. In [38, 39], \(p\) is assumed to be 1 or 2. Since \(\omega > 0\) adjusts the relative importance of \(D y_1\) to \(D y_2\), HSSTV effectively utilizes a-priori knowledge on HS images by a suitable \(\omega\).

The second term in Prob. (9) evaluates the edge similarity between the HR-HS image and the guide image, which compares the edges in the range of the common spectral range of the two images. The technique is based on [40] and assumes that the non-zero differences of the HR-HS and guide images are sparse and correspond to edges and that their edge positions are similar. Therefore, minimizing the mixed \(\ell_{1,2}\) norm of the difference between the edge images of \(M_y u\) and \(M_y q\) helps to exploit spatial information on the guide image. The third term in Prob. (9) is a TV for the guide image. The parameter \(\lambda\) and \(\rho\) balance the three terms.

The first and second constraints serve as data-fidelity of \(u\) and \(q\), respectively. The sets \(B^*_v\) and \(B^*_\eta\) are defined as the \(v\)-centered \(\ell_2\)-norm ball with radius \(v > 0\) and the \(\eta\)-centered \(\ell_2\)-norm ball with radius \(\eta > 0\), respectively. As mentioned in [33–39], [41–44], such hard constraints facilitate the parameter settings because \(v\) and \(\eta\) can decide from the observation information. The third and fourth constraints in Prob. (9) are the dynamic range of \(u\) and \(q\) with \(\mu < \mu\), respectively.

B. Optimization

Prob. (9) is a non-smooth constrained convex optimization, and so a suitable iterative algorithm is required to solve it. In this paper, we adopt the primal-dual splitting method, which is introduced in Sec. III-C. By the definition of the primal-dual splitting method, we reformulate Prob. (9) into Prob. (4).

Since Prob. (9) has four constraints, we put them into the objective function using the indicator function. The problem is rewritten as follows:

\[ \min \text{HSSTV}(u) + \lambda \|D y_1 u - M y_1 q\|_{1,2} + \rho \|D y_2 q\|_{1,2} + \lambda g_{B^*_v}(S y u) + \rho g_{B^*_\eta}(q) + \mu g_{B^*_v}(q) \] (11)

Note that Prob. (11) exactly equals Prob. (9) because of the definition of the indicator function (see (2)).

By letting

\[ g : \mathbb{R}^{N(B+b)} \rightarrow \mathbb{R}^2 : (u, q) \mapsto (\|u\|_{B+y}, \|q\|_{B+y}) \] (12)

\[ h : \mathbb{R}^{(6+\varepsilon)N(B+b)+N(B+b)} \rightarrow \mathbb{R} \cup \{\infty\} : (y_1, y_2, y_3, y_4, y_5) \] (13)

\[ L : \mathbb{R}^{N(B+b)} \rightarrow \mathbb{R}^{(6+\varepsilon)N(B+b)+N(B+b)} : (u, q) \mapsto (A_y u, D y_1 u - M y_1 q, D y_2 q, S y u, q), \] (14)

Prob. (5) is reduced to Prob. (4). The resulting algorithm based on the primal-dual splitting method is summarized in Alg. 1 with (6).

We explain how to calculate the proximity operator of some functions. First, in step 2 and 3, the proximity operator of the indicator function of box constraint is required and defined as follows: for \(i = 1, \ldots, N B\)

\[ [P_{\mu B+y,N B}(x)]_i = \min\{\max\{x_i, \mu\}, \mu\}. \]

Second, step 9, 10, and 11 require the proximity operators of \(\ell_1\) and the mixed \(\ell_{1,2}\) norm, which are reduced to a simple softthresholding type operation: for \(\gamma > 0\) and for \(i = 1, \ldots, 4 N B\), (i) in the case of the \(\ell_1\) norm,

\[ [\text{prox}_{\gamma \cdot \|\cdot\|_1}(x)]_i = \text{sgn}(x_i) \max\{|x_i| - \gamma, 0\}, \]

where \(\text{sgn}\) is the sign function, and (ii) in the case of the \(\ell_{1,2}\) norm,

\[ [\text{prox}_{\gamma \cdot \|\cdot\|_{1,2}}(x)]_i = \max\left\{1 - \gamma \left(\sum_{j=0}^{\gamma} x_{i+jN B}^2\right)^{-\frac{1}{2}}, 0\right\} x_i, \]

where \(i := ((i-1) \mod N B) + 1\). Finally, we introduce the proximity operators of \(g_{B^*_v}\) and \(g_{B^*_\eta}\). Since the two sets \(B^*_v\) and \(B^*_\eta\) are \(\ell_2\)-norm ball, the above proximity operators can be calculated by following forms replacing the center and the radius:

\[ P_{B^*_v}(x) = \begin{cases} x, & \text{if } x \in B^*_v, \\ v + \frac{v(x-v)}{\|x-v\|}, & \text{otherwise.} \end{cases} \]

where \(\|\cdot\|\) is the \(\ell_2\)-norm.
TABLE II  
PARAMETER SETTINGS FOR EXPERIMENTS (TOP: HS PANSHARPENING EXPERIMENTS, BOTTOM: HS AND MS IMAGE FUSION EXPERIMENTS)

| method | parameters $(r, \sigma_g)$ | HS pansharpening                | HS and MS image fusion |
|--------|-----------------------------|--------------------------------|------------------------|
| HySure | $(\lambda, \omega)$        | (2.0, 0.01)                    | (2.0, 0.01)            |
| LTMR   | $(\alpha, \lambda)$        | (2.5, 10^{-1}, 4.5, 10^{-1})   | (2.5, 10^{-1}, 4.5, 10^{-1}) |
| CNN-Fus| $(\lambda, \gamma)$       | (3.0, 1.10, 3.0, 1.50)         | (3.0, 1.10, 3.0, 1.50)  |
| proposed | $\omega$, $\alpha$      | 0.001, 0.001                    | 0.001, 0.001           |
| proposed | $\omega$, $\alpha$      | 0.001, 0.001                    | 0.001, 0.001           |
| Reno   | 256 × 256 × 128            |                                 |                        |
| Pavia U| 256 × 256 × 98             |                                 |                        |
| Salinas| 216 × 216 × 100            |                                 |                        |
| Moffett field | 128 × 128 × 176  |                                 |                        |

The algorithm requires fast Fourier transform and sort function to calculate the blur operator and the $\ell_1$-ball projection. Therefore, the computational complexity is $O(NB \log NB)$.

V. EXPERIMENTS

We demonstrate the advantages of our proposed method on HS pansharpening and HS and MS image fusion experiments. In the experiments, we artificially generated an LR-HS image $v$ and a noisy guide image $g$ from a test HS image based on (7) and (8), and then we estimated an HR-HS image from them. Here, the down-sampling rate of $S$ was set as $r = 2, 4, 8,$ and $16$, and $B$ was a blur matrix with the size of the kernel $2r + 1$. For test HS images, we took four HS images from SpecTIR [43], GIC [46], and Moffett field dataset and normalized all test HS images. Fig. 2 is the test images with their size, where they were depicted as RGB images with $(R, G, B) = (16th, 32nd, 64th)$ bands.

We adopted several existing methods as compared methods: GSA [8], MTF-GLP [9], GFPCA [12], CNMF [17], HySure [13], Lanaras’s [19], LTMR [14], and CNN-Fus [28]. GSA, MTF-GLP, and GFPCA are component analysis-based methods, CNMF, HySure, Lanaras’s, and LTMR are optimization-based methods, and CNN-Fus is a deep learning-based method. We used the MATLAB code presented by the authors of [4], [5], [13], [19], [28]. GSA and MTF-GLP do not consider noisy observations, so we estimated an HR-HS image after pre-denoising. Since GFPCA assumes a noisy LR-HS image and a noiseless guide image, we denoised an observed guide image, and then we estimate an HR-HS image. We adopted FGSLR [47] and FFDNet [48] as pre-denoising methods for an LR-HS image and a guide image, respectively. HySure, Lanaras’s, LTMR, and CNN-Fus require the hyper-parameter settings for each situation, so we experimentally set the value that achieves the highest performance in several settings. In addition, CNN-Fus requires setting the number of iterations T, so we decide it in the same way as the original paper. Tab. III shows the suitable parameters for each situation.

In the proposed method, $\varepsilon$ and $\eta$ in Prob. (9) were set as $\varepsilon = \|SBu - v\|$ and $\eta = \|Ru - g\|$. The hyperparameters $\omega$ in (10), $\lambda$, and $\rho$ in (9) were set like Tab. II where the values achieve the highest performance in $\omega \in [0, 0.1]$, $\lambda \in [0.01, 1.5]$, and $\rho \in [0.5, 1.5]$. To quantitatively evaluate the estimation accuracy, we used four quality measures: the peak signal-to-noise ratio (PSNR) [dB] index, spectral angle mapper (SAM) [49], error relative globale adimensionelle de synthèse (ERGAS) [50], and Q2n [51]. PSNR is defined by $10 \log_{10}(NB/\|u - \hat{u}\|^2)$. It evaluates the pixel-wise similarity between $u$ and $\hat{u}$, and the larger the PSNR value is, the more similar the two HS images are. SAM measures the angle between spectral vectors of an estimated HS image and a ground truth, and it is defined by $\cos^{-1}(u_{spe,j} \cdot \hat{u}_{spe,j}/\|u_{spe,j}\|\|\hat{u}_{spe,j}\|)$. Where $u_{spe,j}$ and $\hat{u}_{spe,j}$ are the $j$th spectral vectors in an estimated HS image and a ground-truth ($j = 1, 2, \ldots, N$). For the definition, SAM is not affected by gain, so it can evaluate inherent spectral similarity. If SAM is equal to $0$, the two spectral vectors are the same, and the small value means that they are similar. ERGAS is a global statistical measure for evaluating fuse data with the best value at $0$ and is defined by

$$ERGAS(u, \hat{u}) = \frac{100}{r} \sqrt{\frac{1}{B} \sum_{i=1}^{B} \frac{\|u_{spa,i} - \hat{u}_{spa,i}\|^2}{(\frac{1}{B} \sum_{i=1}^{B} u_{spa,i}^2)^2}},$$

where $u_{spa,i} \in \mathbb{R}^N$ and $\hat{u}_{spa,i} \in \mathbb{R}^N$ ($i = 1, \ldots, B$) are the images of $i$th band in $u$ and $\hat{u}$, respectively, and $1 \in \mathbb{R}^N$ is all-one vector. ERGAS mainly calculates the band-wise normalized root-mean-square error (RMSE) and considers the difficulty in the fusion problem by multiplying.
| image | method | PSNR[dB] | SAM | ERGAS | Q2+ | PSNR[dB] | SAM | ERGAS | Q2+ | PSNR[dB] | SAM | ERGAS | Q2+ |
|-------|--------|---------|------|-------|-----|---------|------|-------|-----|---------|------|-------|-----|
| Reno | CNN-Fus [28] | 31.42 | 5.409 | 7.347 | 0.834 | 31.04 | 4.595 | 6.109 | 0.809 | 31.05 | 4.624 | 5.622 | 0.764 |
| PaviaU | CNN-Fus [28] | 30.95 | 5.700 | 8.037 | 0.463 | 30.57 | 5.719 | 8.461 | 0.474 | 30.24 | 5.716 | 8.847 | 0.443 |
| Salinas | MT-GlP [9] | 25.87 | 7.283 | 5.740 | 0.933 | 25.68 | 7.069 | 5.974 | 0.903 | 25.67 | 7.052 | 5.977 | 0.903 |
| Moffett field | GFPCA [12] | 36.30 | 5.191 | 2.966 | 0.083 | 36.30 | 5.191 | 2.966 | 0.083 | 36.30 | 5.191 | 2.966 | 0.083 |
| Average | CNN-Fus [28] | 31.54 | 5.142 | 7.622 | 0.549 | 31.54 | 5.204 | 8.091 | 0.554 | 30.97 | 5.204 | 8.091 | 0.554 |

**TABLE III**

The quality measures of the results on HS pansharpening experiments with \( r = 2 \) (boldface: the highest performance, underline: the second performance).

---

The ground sampling distance ratio between LR-HS and guide images, \( Q_{2+} \) is a generalization of the universal image quality index \([52]\) and evaluates three factors: correlation coefficient, the change of the mean bias, and the change in contrast. The value is from 0 to 1, and the two images are similar when it is close to 1.

We set the max iteration number and the stopping criterion of the primal-dual splitting method to 10000 and \( \|u(n) - u(n+1)\|/\|u(n+1)\| < 10^{-4} \), respectively.

---

### A. HS Pansharpening

We conducted HS pansharpening experiments, where the guide image \( g \in \mathbb{R}^N \) has only single-channel information, i.e., a grayscale image. In the experiments, the spectral response operator \( R \in \mathbb{R}^{N \times N} \) in \([8]\) averages information of the visible-light band, and we set the standard deviations of \( n_r \) in \( \mathbb{Z} \) and \( n_g \) in \( \mathbb{R} \) as \( \sigma_r = 0.1 \) and \( \sigma_g = 0.02 \) and 0.04 in \( r = 2 \) and 4 cases, respectively. For \( r = 8 \) and 16, \( \sigma_r \) and \( \sigma_g \) are set as 0.1 and 0.02, and we use *Reno* and *Pavia U* as ground-truth. To satisfy the convergence condition of the primal-dual splitting method, the stepizes were set to \( \gamma_1 = 0.005 \) and \( \gamma_2 = 0.1818 \).

Tab. \( \text{III} \) and \( \text{IV} \) are the results of the quality measures for \( r = 2 \) and \( r = 4 \), respectively. Here, the tables show the average SAM values, and boldface and underline are the best and the second performances, respectively. In many \( r = 2 \) cases, GSA and MT-GlP achieve high performance. The input images are pre-denosed by state-of-the-art denoising methods, so they can effectively extract spatial and spectral characteristics from the images. However, since for the *Moffett field*, GSA is significantly low performance, one can see that the performance is affected by scene objects. Moreover, from Tab. \( \text{III} \) and \( \text{IV} \) the performance of GSA, MTF-GlP, and GFPCA is more sensitive to down-sampling ratio \( r \) than...
other methods. Existing optimization-based methods can estimate a high-performance HR-HS image, but the performance significantly decreases with increasing the noise intensity of a guide image, even though the methods consider noisy real observations. This would be because they do not denoise a guide image and do not sufficiently use spatial information in the image. The tables show that CNN-Fus achieves a guide image and do not sufficiently use spatial information in the image. The tables show that CNN-Fus achieves the highest-estimation performance of all methods in some cases and quality measures, so one can see that CNN-Fus is an effective estimation method. However, CNN-Fus requires setting the iteration number because it cannot converge and the performance is affected by the iteration number such as Fig 1. The suitable iteration number is different by noise intensity and down-sampling ratio like Tab. 1 so the setting is a very difficult task when HR-HS image estimation from real observations. One can see that our proposed methods achieve the highest and/or second estimation performance in some cases and comparable performance in other situations. Specifically, for Pavia U and Moffett field that have more spatial details, our proposed methods can estimate more desirable HS images than compared methods, but for Reno and Salinas, SAM and ERGAS show that some existing methods achieve great performance.

Tab. 1 shows the results in strong down-sampling situations, i.e., $r = 4$ and $16$. From the table, one can see that some methods and our proposed methods can effectively estimate an HR-HS image. Like the $r = 2$ and $4$ cases, our methods can estimate a more desirable HR-HS image for Pavia U, but quality measures especially SAM and ERGAS show that some existing methods achieve more high performance than our methods for Reno.

To verify SAM performance in detail, we plot SAM maps of the results by GSA, MTF-GLP, GFPCA, HySure, CNN-Fus, and proposed (Reno, $r = 4$, and $\sigma_g = 0.02$) in Fig. 3 that
achieve high SAM performance. The SAM maps show that GSA, MTF-GLP, GFPCA, and HySure can estimate desirable spectral information in smooth areas, but the performance in high-variation areas is low. The methods represent an HR-HS image by several bases (endmember) less than the number of the bands and evaluate the smoothness of the coefficient (abundance) in the basis space. Therefore, they can promote low rankness of HS images but would lose spectral information that cannot be represented by the basis. On the other hand, in Fig. 3, one can see that SAM maps of the results by CNN-Fus and our methods have more spectral error than GSA, MTF-GLP, GFPCA, and HySure in smooth areas, but the methods suppress the outstanding error in high-variation areas. One of the applications of HS images is the detection of extraordinary objects, so preserving information on high-variation areas is important. Based on the above discussion, one can see that CNN-Fus and our methods can remain desirable information.

Both ERGAS and PSNR are based on MSE, but in Tab. III and IV the two measures show different evaluations for some situations. Specifically, in Salinas, \( r = 4 \), and \( \sigma_g = 0.04 \), one can see that the proposed method \( (p = 2) \) achieves the highest PSNR for all methods but is inferior to GSA, GFPCA, HySure, and MTF-GLP, GFPCA, and HySure in smooth areas, but the methods suppress the outstanding error in high-variation areas. One of the applications of HS images is the detection of extraordinary objects, so preserving information on high-variation areas is important. Based on the above discussion, one can see that CNN-Fus and our methods can remain desirable information.

Table V

| method | \( r = 8 \), \( \sigma_v = 0.1 \), \( \sigma_g = 0.02 \) | \( r = 16 \), \( \sigma_v = 0.1 \), \( \sigma_g = 0.02 \) |
|---|---|---|
| **GSA** | **PSNR**(dB) | **SAM** | **ERGAS** | **Q2** | **PSNR**(dB) | **SAM** | **ERGAS** | **Q2** |
| MTF-GLP | 28.72 | 8.009 | 3.085 | 0.6988 | 25.08 | 8.064 | 3.140 | 0.5874 |
| GFPCA | 26.84 | 10.28 | 3.516 | 0.6586 | 24.68 | 12.61 | 2.146 | 0.6791 |
| HySure | 30.51 | 6.849 | 2.374 | 0.7645 | 29.45 | 8.026 | 1.544 | 0.7841 |
| Lanaras's | 29.54 | 8.439 | 2.630 | 0.8066 | 28.24 | 10.55 | 1.943 | 0.7603 |
| CNN-Fus | 30.31 | 6.702 | 2.386 | 0.8740 | 28.56 | 9.795 | 3.091 | 0.5142 |

Fig. 3. SAM map on HS pansharpening experiments (Reno, \( r = 4 \), and \( \sigma_g = 0.02 \)).

Fig. 4. MSE and band-wise normalized MSE of the HS pansharpening results by GSA, MTF-GLP, GFPCA, HySure, CNN-Fus, and proposed \( (p = 1) \) and \( (p = 2) \) and mean luminance of the test HS image (Salinas, \( r = 4 \), and \( \sigma_g = 0.04 \)).
and CNN-Fus for ERGAS and almost the same to MTF-GLP. To verify the reason, we confirmed MSE, band-wise normalized MSE, and the band-wise mean luminance in the above situation. This is because PSNR averages MSE by all pixels, and in contrast, ERGAS normalizes MSE by the mean of the spatial vector $\bar{u}_{\text{spa},i}$, which is included in (15). For MSE, one can see that our proposed method almost improves the estimation performance from the 1st to the 75th bands, but over the 76th bands, the proposed method is inferior to other methods in many bands. In the bands of the higher MSE by proposed than others, the mean luminance is relatively low, so one can see that the proposed method is a little weak in the estimation of the low luminance bands. The graph of band-wise normalized MSE shows that the value by the proposed method is significantly high in the low luminance bands. Especially, from 75th to 80 bands, the mean luminance is less than 0.1, and the band-wise normalized MSE is very high. From the results, ERGAS regards MSE in low luminance bands as more important, and our proposed method is weak in the estimation of the bands including low spatial information. However, one can see that our method achieves high performance in most bands.

We show some results in Fig. 5 which depict the resulting HS images as RGB images with $R = 16$, $G = 32$, and $B = 64$ bands. One can see that (i) GFPCA produces spatial artifacts, (ii) CNMF, Lanaras’s, and LTMR remain noise included in the observed PAN images, (iii) GSA, MTF-GLP, CNMF, HySure, and LTMR cannot fully estimate spectral information, (iv) CNN-Fus produces oversmoothing, and (v) the band-wise mean luminance of Salinas. Here, band-wise normalized MSE means $\|u_{\text{spa},i} - u_{\text{spa},i}\|^2/(1^T 1_{\text{spa},i}/N)^2$ included in (15). For MSE, one can see that our proposed method almost improves the estimation performance from the 1st to the 75th bands, but over the 76th bands, the proposed method is inferior to other methods in many bands. In the bands of the higher MSE by proposed than others, the mean luminance is relatively low, so one can see that the proposed method is a little weak in the estimation of the low luminance bands. The graph of band-wise normalized MSE shows that the value by the proposed method is significantly high in the low luminance bands. Especially, from 75th to 80 bands, the mean luminance is less than 0.1, and the band-wise normalized MSE is very high. From the results, ERGAS regards MSE in low luminance bands as more important, and our proposed method is weak in the estimation of the bands including low spatial information. However, one can see that our method achieves high performance in most bands.

![Fig. 5. Results on HS pansharpening experiments (top: Reno, $r = 4$, and $\sigma_g = 0.02$, bottom: PaviaU, $r = 8$, and $\sigma_g = 0.02$).](image)

| method              | computational time (s) | iteration numbers | endmember (average time of each iteration) | abundance (average time of each iteration) |
|---------------------|------------------------|-------------------|-------------------------------------------|------------------------------------------|
| GSA                 | 0.3483                 | -                 | -                                         | -                                        |
| MTF-GLP             | 0.3273                 | -                 | -                                         | -                                        |
| GFPCA               | 0.4234                 | -                 | -                                         | -                                        |
| CNMF                | 23.09                  | 2                 | 0.5375 (0.2688)                           | -                                        |
| HySure              | 26.50                  | 383               | 0.4431 (0.0642)                           | 23.82 (0.1642)                           |
| Lanaras’s           | 19.02                  | 219               | 18.84 (0.08684)                           | -                                        |
| LTMR                | 7.342                  | 11                | 0.02488 (0.6001)                          | 6.601 (0.6001)                           |
| CNN-Fus             | 12.13                  | 25                | 0.1611 (0.4659)                           | 11.68 (0.4659)                           |
| proposed ($p = 1$)  | 94.58                  | 1702              | 94.54 (0.05555)                           | 94.54 (0.05555)                          |
| proposed ($p = 2$)  | 98.38                  | 1697              | 98.37 (0.05796)                           | 98.37 (0.05796)                          |
our proposed methods can effectively estimate HR-HS images. From the enlarged area of the top of Fig. 5 and the top center area in the bottom, the results of our proposed methods have more similar colors to ground-truth than that by HySure (especially blue and yellow architectures), so our methods can preserve spectral information of small areas.

Fig. 6 and 7 show $\omega$ and $\lambda$ versus four quality measures, respectively, where the graphs plot the average. In Fig. 6 one can see that the suitable $\omega$ is rarely affected by $r$ and $\sigma_g$. In PSNR, ERGAS, and $Q^2n$, it is good that $\omega$ is set as less than 0.03, and For SAM, $\omega \in [0.070, 1]$ is suitable. The two results are contrastive, so $\omega$ requires to be adjusted by what information one requires. In Fig. 7, one can see that the suitable $\lambda$ changes between SAM and the other quality measures. Since the hyperparameter $\lambda$ is the weight of the edge-similarity evaluation, the results are appropriate.

We plot the spatial and spectral response of the results in a Moffett field, $r = 2$, and $\sigma_g = 0.04$ case in Fig. 8. Here, we picked up the spatial vectors at the 43rd row and the 30th band of the results and the spectral vectors at the 43rd row and the 107th column of them. The graph of the spatial response shows that many methods produce artifacts, and the result by MTF-GLP has oversmoothing. Especially, the result by CNMF has some less than 0 value that cannot be actually observed value.
| Image method | GSA [8] | MTF-GLP [9] | GFPCA [10] | CNMF [11] | HySure | GSA [8] | MTF-GLP [9] | GFPCA [10] | CNMF [11] | HySure | GSA [8] | MTF-GLP [9] | GFPCA [10] | CNMF [11] | HySure | GSA [8] | MTF-GLP [9] | GFPCA [10] | CNMF [11] | HySure |
|---------------|----------|-------------|-------------|------------|--------|----------|-------------|-------------|------------|--------|----------|-------------|-------------|------------|--------|----------|-------------|-------------|------------|--------|----------|-------------|-------------|------------|
| PSNR[dB] | 32.73 | 35.36 | 6.815 | 7.6589 | 0.3847 | 32.73 | 35.36 | 6.815 | 7.6589 | 0.3847 | 32.73 | 35.36 | 6.815 | 7.6589 | 0.3847 | 32.73 | 35.36 | 6.815 | 7.6589 | 0.3847 |
| EREMS | 8.105 | 7.6589 | 7.6589 | 7.6589 | 7.6589 | 8.105 | 7.6589 | 7.6589 | 7.6589 | 7.6589 | 8.105 | 7.6589 | 7.6589 | 7.6589 | 7.6589 | 8.105 | 7.6589 | 7.6589 | 7.6589 |
| Q2 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 |
| PSNR[dB] | 30.36 | 5.624 | 8.783 | 0.6988 | 30.36 | 5.624 | 8.783 | 0.6988 | 30.36 | 5.624 | 8.783 | 0.6988 | 30.36 | 5.624 | 8.783 | 0.6988 | 30.36 | 5.624 | 8.783 | 0.6988 |
| EREMS | 7.978 | 10.59 | 6.259 | 6.259 | 7.978 | 10.59 | 6.259 | 6.259 | 7.978 | 10.59 | 6.259 | 6.259 | 7.978 | 10.59 | 6.259 | 6.259 | 7.978 | 10.59 | 6.259 |
| Q2 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 | 0.7859 |

We measured computational time, where we use Pavia U as a test HS image and set \( r = 2 \) and \( \sigma_g = 0.02 \). We conducted the experiments with MATLAB 2022a and a computer equipped with 64-GB random access memory and 12th Gen Intel(R) Core(TM) i9-12900K CPU with 3.2 GHz or NVIDIA GeForce RTX 3090. In Tab. [VI] we align the computational time of all processing. In terms of the optimization-based methods and CNN-Fus, the table also shows the iteration number and endmember/abundance estimation time with the average time of each iteration. In this table, our methods spend much time compared with existing methods, but the iteration number is changed by image, noise intensity, degradation operator, and so on. From the time of each iteration, one can see that our methods achieve less computational time than existing optimization-based methods and CNN-Fus. Moreover, GSA, GFPCA, and HySure suppress spatial artifacts compared with other existing methods, but the value of some pixels is significantly different, e.g., 13th, 53rd, and 120th elements. On the other hand, CNN-Fus and our proposed method can estimate spectral response without artifacts and oversmoothing compared with other methods. From the spectral response graph, one can see that CNMF suppresses intrinsic spectral response, and the results by GFPCA, CNMF, and CNN-Fus have spectral artifacts. In addition, one can also see that the values estimated by GSA, GFPCA, HySure, LTMR, and the proposed \((p = 2)\) method are far apart with the large of the band number, and they are not good at estimating information of large wavelength. MTF-GLP, LTMR, and Lanaras’ produce the increase of spectral response that does not exist in the original in small bands. In contrast, the proposed \((p = 1)\) method can estimate a more similar spectral response to the original one than the other methods. Based on the results of the spatial and spectral response, one can see that our method \((p = 1)\) achieves effective spatial and spectral estimation.
TABLE VIII
The quality measures of the results on HS and MS image fusion experiments with \( r = 4 \) (boldface: the highest performance, underline: the second performance).

| image       | method       | \( \sigma_F = 0 \) | PSNR[DB] | SAM | ERGAS | Q2*   |
|-------------|--------------|--------------------|---------|-----|-------|-------|
| Reno        | GSA [8]      | 28.52              | 7.943   | 6.391| 0.6115|
|             | MTF-GLP [9]  | 26.36              | 7.943   | 7.235| 0.4737|
|             | GPCA [10]    | 27.22              | 8.218   | 6.353| 0.4586|
|             | CNMF [11]    | 30.20              | 6.008   | 5.541| 0.3864|
|             | HySure [12]  | 31.29              | 8.316   | 4.574| 0.7605|
|             | Lanaras’s [13]| 38.78              | 3.375   | 2.102| 0.8914|
|             | LTMB [14]    | 32.85              | 6.810   | 4.778| 0.8485|
|             | CNN-Fas [15] | 33.31              | 8.534   | 3.528| 0.9377|
| proposed    | \( r = 1 \)   | 33.17              | 6.866   | 4.588| 0.9217|
| proposed    | \( r = 2 \)   | 33.22              | 6.967   | 4.512| 0.9220|
|             | GSA [8]      | 26.95              | 9.777   | 6.433| 0.3997|
| Pavia U     | MTF-GLP [9]  | 24.56              | 10.54   | 8.988| 0.2310|
|             | GPCA [10]    | 25.06              | 8.769   | 7.950| 0.2190|
|             | CNMF [11]    | 30.63              | 4.884   | 3.697| 0.6466|
|             | HySure [12]  | 27.84              | 10.66   | 5.556| 0.4548|
|             | Lanaras’s [13]| 35.78              | 4.058   | 2.192| 0.7412|
|             | LTMB [14]    | 28.71              | 8.279   | 5.089| 0.5697|
|             | CNN-Fas [15] | 29.07              | 7.564   | 5.046| 0.9063|
| proposed    | \( r = 1 \)   | 30.48              | 7.025   | 3.621| 0.9418|
| proposed    | \( r = 2 \)   | 30.58              | 7.059   | 3.607| 0.9416|
| Salinas      | GSA [8]      | 32.14              | 4.427   | 2.672| 0.5144|
|             | MTF-GLP [9]  | 29.88              | 5.189   | 3.286| 0.4559|
|             | GPCA [10]    | 30.46              | 4.723   | 3.228| 0.4277|
|             | CNMF [11]    | 28.26              | 9.156   | 4.928| 0.5030|
|             | HySure [12]  | 32.25              | 4.022   | 2.791| 0.5115|
|             | Lanaras’s [13]| 36.28              | 4.481   | 1.834| 0.7838|
|             | LTMB [14]    | 31.54              | 4.431   | 2.625| 0.4253|
|             | CNN-Fas [15] | 32.26              | 4.471   | 2.507| 0.7233|
| proposed    | \( r = 1 \)   | 32.10              | 4.426   | 2.729| 0.5756|
| proposed    | \( r = 2 \)   | 31.95              | 4.571   | 3.219| 0.6230|
| Moffett field| GSA [8]      | 25.47              | 10.22   | 7.575| 0.4821|
|             | MTF-GLP [9]  | 25.89              | 10.15   | 7.401| 0.6449|
|             | GPCA [10]    | 26.21              | 9.417   | 6.504| 0.5408|
|             | CNMF [11]    | 24.96              | 9.901   | 7.045| 0.2180|
|             | HySure [12]  | 25.48              | 11.17   | 6.935| 0.3952|
|             | Lanaras’s [13]| 27.13              | 8.303   | 5.710| 0.5863|
|             | LTMB [14]    | 26.55              | 10.86   | 6.567| 0.2155|
|             | CNN-Fas [15] | 27.55              | 9.766   | 5.870| 0.8172|
| proposed    | \( r = 1 \)   | 26.49              | 9.577   | 6.147| 0.7704|
| proposed    | \( r = 2 \)   | 26.67              | 9.521   | 6.033| 0.7774|
| Average      | GSA [8]      | 28.27              | 8.902   | 5.846| 0.5109|
|             | MTF-GLP [9]  | 26.68              | 8.456   | 6.728| 0.4514|
|             | GPCA [10]    | 27.24              | 7.782   | 6.058| 0.4115|
|             | CNMF [11]    | 28.51              | 7.506   | 5.177| 0.5372|
|             | HySure [12]  | 29.27              | 8.542   | 4.964| 0.5314|
|             | Lanaras’s [13]| 34.49              | 5.455   | 2.960| 0.7507|
|             | LTMB [14]    | 29.92              | 7.495   | 4.549| 0.5136|
|             | CNN-Fas [15] | 30.55              | 6.533   | 4.238| 0.8466|
| proposed    | \( r = 1 \)   | 30.56              | 6.972   | 3.228| 0.8212|
| proposed    | \( r = 2 \)   | 30.61              | 7.030   | 4.264| 0.8165|

one can see that GSA, MTF-GLP, and GPCA achieve less computational time than other methods, but they require a pre-denosing step, so one needs to consider the time of the step. In the pre-denosing process, FFDNet and FSGLR spend 0.1210s and 246.4s for denoising a guide image and an LR-HS image, respectively.

B. HS and MS Image Fusion
To verify the performance of our proposed method for HS and MS image fusion, we did an experiment that estimates an HR-HS image from a pair of LR-HS and MS images. Here, we assumed a 4-bands MS image, and \( \mathbf{R} \in \mathbb{R}^{N \times N} \) in (8) was the spectral response of the IKONOS satellite. In addition, we set the standard deviations of \( \mathbf{v} \) in (7) and \( \mathbf{g} \) in (8) as \( \sigma_v = 0.2 \) and \( \sigma_g = 0, 0.05 \) and 0.1, respectively, for \( r = 2 \) and 4 cases. In the experiments on large downsampling ratio \( r = 8 \) and 16, we set \( \sigma_v = 0.1 \) and \( \sigma_g = 0.05 \) and use only Reno and Pavia U because of the spatial size of the test image. We set the stepizes \( \gamma_1 \) and \( \gamma_2 \) to 0.01 and 0.5, respectively, as satisfied with the convergence condition.

In Tab. VII and VIII we show four quality measures of the results on \( r = 2 \) and \( r = 4 \), respectively. When the noiseless guide image (\( \sigma_g = 0 \)), Lanaras’s achieves the highest performance of all methods. However, the performance of Lanaras’s significantly decrease as the noise intensity is large. Even though the method considers noisy observations, one can see that the result is sensitive to noise in a guide image. In some cases, the results of CNN-Fus are better than the other methods. As mentioned in Sec. V.A, CNN-Fus has a serious problem, so in a real situation, the achievement of the performance is very difficult. GSA, MTF-GLP, and HySure outperform our methods in some cases, but the performance is affected by image. From the average of the results, one can see that our methods achieve the highest performance and the
stable estimation of all methods from a noisy guide image.

Table IX shows the quality measures of the results, and one can see that CNN-Fus achieves the best performance. However, from the discussion in the above experiments, the setting of the iteration number is very difficult but essential. Therefore, in fact, it is difficult that CNN-Fus realizes the performance. For Reno, SAM of GSA, MTF-GLP, and HySure is superior to that of our methods about 0.3–0.7, but our methods greatly outperform them about from 1 to 5 in Pavia U cases. In PSNR, ERGAS, and \( Q^2 \), the table shows that our methods achieve the second performance, i.e., one can see that our methods can estimate a desirable HR-HS image in the stable estimation.

From the results of quality measures, our methods can estimate a higher-performance HR-HS image on average, but SAM shows that some methods achieve the higher of almost the same performance as our methods for Reno and Salinas. To deeply analyze the reason, we depict the SAM map of the results by some existing methods and proposed (\( p = 1 \) and \( p = 2 \)) on HS and MS image fusion experiments (Reno, \( r = 4 \), and \( \sigma_g = 0.1 \)).

![SAM map of the results by some existing methods and proposed](image)

Fig. 9. SAM map of the results by some existing methods and proposed (\( p = 1 \) and \( p = 2 \)) on HS and MS image fusion experiments (Reno, \( r = 4 \), and \( \sigma_g = 0.1 \)).

The following table shows the quality measures of the results on HS and MS image fusion experiments with \( r = 8 \) and 16 (boldface: the highest performance, underline: the second performance).

| Image    | Method     | \( r = 8 \), \( \sigma_v = 0.1 \), \( \sigma_g = 0.05 \) | \( r = 16 \), \( \sigma_v = 0.1 \), \( \sigma_g = 0.05 \) |
|----------|------------|--------------------------------------------------|--------------------------------------------------|
|          |            | PSNR\( [\text{dB}] \) | SAM | ERGAS | \( Q^2 \) | PSNR\( [\text{dB}] \) | SAM | ERGAS | \( Q^2 \) |
| Reno     | GSA [8]    | 27.87 | 8.009 | 3.340 | 0.5810 | 24.06 | 8.064 | 2.237 | 0.2813 |
|          | MTF-GLP [9]| 24.62 | 8.009 | 4.273 | 0.3432 | 21.86 | 8.664 | 3.007 | 0.2813 |
|          | GFPCA [12]| 23.55 | 7.604 | 4.633 | 0.2939 | 20.97 | 9.960 | 3.105 | 0.2813 |
|          | CNMF [17] | 27.15 | 10.79 | 3.633 | 0.5824 | 26.87 | 10.44 | 1.943 | 0.7390 |
|          | HySure [15]| 29.27 | 7.387 | 2.717 | 0.7587 | 27.55 | 9.887 | 1.727 | 0.7487 |
|          | Lanaras's [19]| 28.49 | 9.876 | 3.009 | 0.7818 | 27.68 | 11.61 | 1.624 | 0.6918 |
|          | LTMR [14] | 27.29 | 11.44 | 3.655 | 0.6514 | 26.39 | 13.18 | 2.103 | 0.5485 |
|          | CNN-Fus [28]| 32.69 | 6.053 | 1.837 | 0.9388 | 31.27 | 7.872 | 1.113 | 0.9253 |

**Table IX**

The quality measures of the results on HS and MS image fusion experiments with \( r = 8 \) and 16 (boldface: the highest performance, underline: the second performance).
We show two results in Fig. [11] where they are depicted by RGB images with $R = 16$th $G = 32$nd $B = 64$th bands. The upper two rows are the results in Moffett field, $r = 2$, and $\sigma_g = 0.1$ case, and the lower two are that in Pavia U, $r = 4$, and $\sigma_g = 0.05$ case. One can see that GSA, MTF-GLP, GFPCA, CNMF, Lanaras’s, and LTMR produce spatial and/or spectral artifacts, and the results by CNN-Fus are spatial oversmoothing or artifacts. One can also see that HySure can estimate spatial information, but the results by HySure have a few different colors, e.g. square architecture of Moffett field and orange architecture in the top center of Pavia U. Therefore, one can see that HySure produces a little distortion in the spectral direction. In contrast, our methods can estimate an HR-HS image without spatial and spectral artifacts compared with existing methods.

In Fig. [12] and [13] we plot $\omega$ and $\lambda$ versus four quality measures, where the top and bottom are the results in $r = 2$ and 4 cases, respectively. Fig. [12] shows that the suitable $\omega$ rarely changes by $r$ and $\sigma_g$ in most cases. In PSNR, ERGAS, and $Q^2$, one can see that it is good that $\omega$ is set as less than 0.03. On the other hand, the graph shows that in SAM, the suitable $\omega$ value is from 0.03 to 0.05 different from the above quality measures. Fig. [13] shows that the suitable $\lambda$ is almost the same in different $\sigma_g$ and $r$ cases, and $\lambda \in [0.07, 0.1]$ is good in most cases.

In Fig. [14] we plot the spectral and spatial response of the results in a Moffett field, $r = 2$, and $\sigma_g = 0.1$ case in like manner to HS pan sharpening experiments. From the spatial response graph, one can see that other than GFPCA, HySure, and the proposed methods produce spatial artifacts. GFPCA, HySure, and our proposed method can reduce spatial artifacts and estimate some rapid spatial change. Especially, around the 40th elements and from the 45th to the 50th elements, the proposed methods can estimate a similar response to the original one and other methods. The spectral response graph shows that GFPCA, CNMF, HySure, Lanaras’s, and LTMR produce huge spectral artifacts. Moreover, one can see that the values estimated by GSA and MTF-GLP are far apart from the original after the 95th band. CNN-Fus can estimate similar spectral vectors to the original, but from the 10th to 40th bands, the spectral value significantly increases from the original. In contrast, our proposed methods can estimate more similar spectral responses to the original without spectral artifacts than other methods. From the above results, one can see that our method can estimate desirable spatial and spectral information.

VI. CONCLUSION

We have proposed a new HR-HS estimation method based on convex optimization robust to noise in observations. Our method simultaneously estimates an HR-HS image and a noiseless guide image to suppress the effect of noise. The method can utilize the spatial detailed information in a guide image because it evaluates the edge-similarity between an HR-HS image and an estimated guide image. Moreover, it utilizes HSSTV to evaluate spatial and spectral piecewise smoothness of an HS image, which is TV-based regularization and achieves high performance for HS image denoising and compressed sensing reconstruction. Thanks to the design, the proposed method achieves robust HR-HS image estimation. To efficiently solve the problem, we adopt the primal-dual splitting method. In the experiments, we verify the performance of our method and illustrate the advantages over several existing methods.

ACKNOWLEDGMENT

This work was supported in part by JST PRESTO under Grant JPMJPR21C4, and in part by JSPS KAKENHI under Grant 22H03610, 22H00512, 21K21312, 20H02145, and 18H05413.

REFERENCES

[1] C. I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003, vol. 1.
[2] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Braile, L. Bruzzone, G. Camps-Valls, I. Chaussot, M. Fauvel, P. Gamba, A. Gualtieri et al., “Recent advances in techniques for hyperspectral image processing,” *Remote sensing of environment*, vol. 113, pp. S110–S122, 2009.
[3] B. Rasti, P. Scheunders, P. Ghamisi, G. Liciardi, and J. Chanussot, “Noise reduction in hyperspectral imagery: Overview and application,” *Remote Sens.*, vol. 10, no. 3, p. 482, 2018.
[4] L. Loncan, L. B. de Almeida, J. M. Boucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Liciardi, M. Simeos et al., “Hyperspectral pansharpening: A review,” *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015.
Fig. 11. Results on HS and MS image fusion experiments (top: Reno, $r = 2$, and $\sigma_g = 0.1$, bottom: Pavia U, $r = 4$, and $\sigma_g = 0.05$).

[5] N. Yokoya, C. Grohfeldt, and J. Chanussot, “Hyperspectral and multispectral data fusion: A comparative review of the recent literature,” IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 2, pp. 29–56, 2017.

[6] P. Chavez, S. C. Sides, J. A. Anderson et al., “Comparison of three different methods to merge multispectral and hyperspectral data-landsat tm and spot panchromatic,” Photogrammetric Engineering and Remote Sensing, vol. 57, no. 3, pp. 295–303, 1991.

[7] C. A. Laben and B. V. Brower, “Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening,” Jan. 4 2000, uS Patent 6,011,875.

[8] B. Aiazzi, S. Baronti, and M. Selva, “Improving component substitution pansharpening through multivariate regression of ms + pan data,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 10, pp. 3230–3239, 2007.

[9] B. Aiazzi, L. Aliparone, S. Baronti, A. Garzelli, and M. Selva, “Mtf-tailored multiscale fusion of high-resolution ms and pan imagery,” Photogrammetric Engineering & Remote Sensing, vol. 72, no. 5, pp. 591–596, 2006.

[10] G. Vivone, R. Restaino, M. D. Mura, G. Licciardi, and J. Chanussot, “Contrast and error-based fusion schemes for multispectral image pansharpening,” IEEE Geosci. and Remote Sens. Lett., vol. 11, no. 5, pp. 930–934, 2014.

[11] J. G. Liu, “Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details,” International Journal of Remote Sensing, vol. 21, no. 18, pp. 3461–3472, 2000.

[12] W. Liao, X. Huang, F. V. Coillie, S. Gautama et al., “Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 ieee gss data fusion contest,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2984–2996, 2015.

[13] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, “A convex formulation for hyperspectral image superresolution via subspace-based regularization,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 6, pp. 3373–3388, 2015.

[14] R. Dian and S. Li, “Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization,” IEEE Trans. Image Process., vol. 28, no. 10, pp. 5135–5146, 2019.

[15] Y. Yang, C. Wang, Y. Feng, J. Zhang, Y. Zheng, and S. Chen, “Regularizing subspace representation for fusing hyperspectral and multispectral images,” IEEE Selected Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 12,273–12,286, 2021.

[16] J. Ying, H. Shen, and S. Cao, “Unaligned hyperspectral image fusion via registration and interpolation modeling,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2021.

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

[18] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J. Tournet, M. Chen, and S. Godsill, “Multiband image fusion based on spectral unmixing,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 12, pp. 7236–7249, 2016.

[19] C. Lanaras, E. Baltsavias, and K. Schindler, “Hyperspectral super-resolution by coupled spectral unmixing,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2015, pp. 3586–3594.

[20] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, “Hyperspectral super-resolution via coupled tensor ring factorization,” Pattern Recognition, vol. 122, p. 108280, 2022.

[21] P. Liu and L. Xiao, “A nonconvex pansharpening model with spatial and spectral gradient difference-induced nonconvex sparsity priors,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2021.

[22] X. Tian, Y. Chen, C. Yang, and J. Ma, “Variational pansharpening by exploiting cartoon-texture similarities,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2021.

[23] Y. Qu, H. Qi, C. Kwan, N. Yokoya, and J. Chanussot, “Unsupervised and unregistered hyperspectral image super-resolution with mutual dirichlet-net,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–18, 2021.
Fig. 12. Four quality measures versus $\omega$ in (10) on HS and MS image fusion (top: $r = 2$, bottom: $r = 4$).

Fig. 13. Four quality measures versus $\lambda$ in (9) on HS and MS image fusion (top: $r = 2$, bottom: $r = 4$).

Fig. 14. Spatial (left) and spectral (right) response of the results on HS and MS image fusion experiments (Moffett field, $r = 2$, and $\sigma_g = 0.1$).

[24] R. Dian, S. Li, A. Guo, and L. Fang, “Deep hyperspectral image sharpening,” IEEE Trans. Neural Networks and Learning Systems, vol. 29, no. 11, pp. 5345–5355, 2018.
[25] J. Yang, L. Xiao, Y. Zhao, and J. C. Chan, “Variational regularization network with attentive deep prior for hyperspectral–multispectral image fusion,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–17, 2021.
[26] Z. Wu, T. Huang, L. Deng, G. Vivone, J. Miao, J. Hu, and X. Zhao, “A new variational approach based on proximal deep injection and gradient intensity similarity for spatio-spectral image fusion,” IEEE Selected Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 6277–6290, 2020.
[27] J. Qu, Y. Shi, W. Xie, Y. Li, X. Wu, and Q. Du, “Mssl: Hyperspectral and panchromatic images fusion via multiresolution spatial–spectral feature learning networks,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, 2021.
[28] R. Dian, S. Li, and X. Kang, “Regularizing hyperspectral and panchromatic data fusion by cnn denoiser,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 3, pp. 1124–1135, 2020.
[29] S. Takeyama, S. Ono, and K. I, “A constrained convex optimization approach to hyperspectral image restoration with hybrid spatio-spectral regularization,” Remote Sensing, vol. 12, no. 21, p. 3541, 2020.
[30] N. Yokoya, T. Yairi, and A. Iwasaki, “Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization,” in Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011, pp. 1–4.
[31] J. J. Moreau, “Fonctions convexes duales et points proximaux dans un espace hilbertien,” C. R. Acad. Sci. Paris Ser. A Math., vol. 255, pp. 2897–2899, 1962.
[32] H. H. Bauschke and P. L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. New York: Springer, 2011.
[33] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for
convex problems with applications to imaging,” *J. Math. Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2010.

[34] L. Wald, T. Ranchin, and M. Mangolini, “Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images,” *Photogrammetric Engineering & Remote Sensing*, vol. 65, no. 6, pp. 691–699, 1999.

[35] S. Takeyama, S. Ono, and I. Kumazawa, “Mixed noise removal for hyperspectral images using hybrid spatio-spectral total variati,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 3128–3132.

[36] S. Takeyama and S. Ono, “Joint mixed-noise removal and compressed sensing reconstruction of hyperspectral images via convex optimization,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2020, pp. 1492–1495.

[37] ——, “Compressed hyperspectral pansharpening,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2020, pp. 2855–2859.

[38] S. Takeyama, S. Ono, and I. Kumazawa, “Hyperspectral image restoration by hybrid spatio-spectral total variation,” *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 4586–4590, 2017.

[39] ——, “Hybrid spatio-spectral total variation: A regularization technique for hyperspectral image denoising and compressed sensing,” *arXiv preprint arXiv:1907.13357*, 2019.

[40] C. Chen, Y. Li, W. Liu, and J. Huang, “Image fusion with local spectral consistency and dynamic gradient sparsity,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 2760–2765.

[41] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, “An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems,” *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[42] G. Cherchiglia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, “Epigraphical projection and proximal tools for solving constrained convex optimization problems,” *Signal, Image and Video Process.*, vol. 9, no. 8, pp. 1737–1749, 2015.

[43] S. Ono and I. Yamada, “Signal recovery with certain involved convex data-fidelity constraints,” *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6149–6163, 2015.

[44] S. Ono, “Efficient constrained signal reconstruction by randomized epigraphical projection,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2019, pp. 4993–4997.

[45] “SpecTIR,” http://www.spectir.com/free-data-samples/.

[46] “GIC,” http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

[47] Y. Chen, T. Z. Huang, W. He, X. L. Zhao, H. Zhang, and J. Zeng, “Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.

[48] K. Zhang, W. Zuo, and L. Zhang, “Fidnet: Toward a fast and flexible solution for cnn-based image denoising,” *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.

[49] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Burloon, and A. Goetz, “The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data,” *Remote Sensing of Environment*, vol. 44, no. 2, pp. 145–163, 1993.

[50] L. Wald, “Quality of high resolution synthesised images: Is there a simple criterion?” in *Third conference Fusion of Earth data: merging point measurements, raster maps and remotely sensed images*, 2000, pp. 99–103.

[51] A. Garzelli and F. Nencini, “Hypercomplex quality assessment of multi/hyperspectral images,” *IEEE Geosci. and Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, 2009.

[52] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.