Abstract—We propose a new regularization technique, named Hybrid Spatio-Spectral Total Variation (HSSTV), for hyperspectral (HS) image denoising and compressed sensing. Regularization techniques based on total variation (TV) focus on local differences of an HS image to model its underlying smoothness and have been recognized as a popular approach to HS image restoration. However, existing TVs do not fully exploit underlying spectral correlation in their designs and/or require a high computational cost in optimization. Our HSSTV is designed to simultaneously evaluate two types of local differences: direct local spatial differences and local spatio-spectral differences in a unified manner with a balancing weight. This design resolves the said drawbacks of existing TVs. Then, we formulate HS image restoration as a constrained convex optimization problem involving HSSTV and develop an efficient algorithm based on the alternating direction method of multipliers (ADMM) for solving it. In the experiments, we illustrate the advantages of HSSTV over several state-of-the-art methods.

Index Terms—hyperspectral image restoration, total variation, ADMM, denoising, compressed sensing.

I. INTRODUCTION

HYPERSONTICAL (HS) images are required with many applications in a wide range of field, e.g., earth observation, agriculture, and medical and biological imaging [1]–[5]. An HS image has 1D spectral information including invisible light and narrow wavelength interval in addition to 2D spatial information and thus can visualize invisible intrinsic characteristics of scene objects and environmental lighting.

Observed HS images are often affected by noise because of the small amount of light in narrow wavelength and/or sensor failure. Also, in compressive HS imaging [3], [5], we have to estimate a full HS image from incomplete measurements. Thus, we need some methods for restoring clean HS images in HS applications.

Most HS image restoration methods are established based on optimization: a restored HS image is characterized as a solution to some optimization problem, which consists of regularization and data-fidelity terms. The regularization term plays a role to incorporate a-priori knowledge about underlying properties on HS images, and so the design is very important to get a reasonable result under ill-posed or ill-conditioned scenarios, which is typical in HS image restoration.

Regularization techniques based on total variation (TV) have been recognized as a powerful approach to image restoration. Since natural images have piecewise-smooth structures, the total magnitude of local spatial differences tends to be small. Hyperspectral TV (HTV) [6], its band-group-wise extension [8], and collaborative TV (CTV) [7] are relatively simple TV regularization techniques for HS images. These TVs strongly promote spatial piecewise smoothness, but they do not or a little care about spectral correlation, which is a typical property of HS images. As a result, these TVs cause spatial oversmoothing, as will be detailed in Sec. III. Some TVs rely on semilocal [9], [10] or nonlocal [11], [12] similarity of HS images. Although they can remove noise while keeping sharp edges, the use of them often results in spectral distortion because of ignoring spectral correlation.

To overcome the drawback in the above TVs, there have been proposed many methods that explicitly take spectral correlation into account. Such methods can be roughly classified into two approaches: (i) combining the above TVs with low-rank modeling (LRM), and (ii) incorporating spectral correlation into the design of TV.

The methods proposed in [10], [13]–[15] are categorized as the approach (i), which characterize spectral correlation via low-rankness. They can restore better HS images than the methods only using the above TVs. However, their computational cost is expensive because one has to compute iterative singular value decomposition in optimization involving LRM. Moreover, since TV and LRM are different types of regularization terms, tuning of hyperparameters is difficult in their models.

Arranged structure tensor TV (ASTV) [16], spatio-spectral TV (SSTV) [17], and anisotropic SSTV (ASSTV) [18] are categorized as (ii), which explicitly evaluate spectral correlation in addition to spatial piecewise smoothness. ASTV models spectral correlation as the low-rankness of semilocal gradient matrices. However, the use of ASTV requires singular value decomposition as with LRM, so the computational

1In addition, TVs based on nonlocal similarity require a chicken-and-egg self-similarity evaluation for finding similar patches.
cost becomes high. On the other hand, in the design of SSTV and ASSTV, spatial correlation is interpreted as spectral piecewise smoothness and is evaluated by the $\ell_1$ norm of local differences along a spectral direction, resulting in computationally efficient optimization. Specifically, to evaluate spatial and spectral piecewise smoothness simultaneously, SSTV focuses on local spatio-spectral differences. However, it ignores direct local spatial differences, and so restored HS images by SSTV tend to have artifacts, especially in highly noisy scenarios. Meanwhile, ASSTV directly handles both local spatial and spectral differences, but it often produces spectral oversmoothing because it strongly suppresses the $\ell_1$ norm of direct spectral differences (SSTV and ASSTV are detailed in Sec. III). Some methods [18]–[23] combine SSTV/ASSTV with LRM, but their computational cost is expensive due to LRM. Table I summarizes the pros and cons of the methods reviewed in this section.

Based on the above discussion, we propose a new TV regularization technique for HS image restoration, named Hybrid Spatio-Spectral Total Variation (HSSTV). To effectively utilize a-priori knowledge on HS images, HSSTV is designed to evaluate two types of local differences: direct local spatial differences and local spatio-spectral differences in a unified manner with a balancing weight. This design resolves the drawbacks of the existing TV regularization techniques mentioned above. HSSTV consists of local difference operators and the $\ell_1$/mixed $\ell_{1,2}$ norm, and thus optimization problems involving HSSTV can be efficiently solved by nonsmooth convex optimization methods like proximal splitting algorithms. Specifically, we develop an algorithm based on the alternating direction method of multipliers (ADMM) [24]–[27].

The remainder of the paper is organized as follows. Section II introduces notation and mathematical ingredients. Section III reviews existing work related to our method. In Section IV, we establish HSSTV, formulate HS image restoration as a nonsmooth convex optimization problem involving HSSTV, and develop an algorithm for solving the problem. Extensive experiments on denoising and compressed sensing (CS) reconstruction of HS images are given in Section V, where we illustrate the advantages of HSSTV over several state-of-the-art methods. Section VI concludes the paper. The preliminary version of this work, without mathematical details, deeper discussion, new applications, and comprehensive experiments has appeared in conference proceedings [28].

### II. Preliminaries

#### A. Notation and Definitions

In this paper, 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 $U \in \mathbb{R}^{N_b \times N_h}$ as a vector $u \in \mathbb{R}^{NB} (N := N_cN_h$ 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 $i$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 $f : \mathbb{R}^L \to (-\infty, \infty]$ is called proper lower semicontinuous convex if $\text{dom}(f) := \{x \in \mathbb{R}^L | f(x) < \infty \} \neq \emptyset$, $\text{lev}_{\leq c}(f) := \{x \in \mathbb{R}^L | f(x) \leq c \}$ is closed for every $c \in \mathbb{R}$, and $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ for every $x, y \in \mathbb{R}^L$ and $\lambda \in (0, 1)$, respectively. Let $\Gamma_0(\mathbb{R}^L)$ be the set of all proper lower semicontinuous convex functions on $\mathbb{R}^L$.

The proximity operator $[29]$ plays a central role in convex optimization based on proximal splitting. The proximity operator of $f \in \Gamma_0(\mathbb{R}^L)$ with an index $\gamma > 0$ is defined by

$$\text{prox}_{\gamma f}(x) := \arg\min_y f(y) + \frac{1}{2\gamma}||y - x||^2.$$

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

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

Then, for any $\gamma > 0$, its proximity operator is given by

$$\text{prox}_{\gamma \iota_C}(x) = P_C(x) := \arg\min_{y \in C} ||x - y||,$$

where $P_C(x)$ is the metric projection onto $C$.  

| Methods         | Feature                        | Spatial Smoothness | Spectral Correlation | Gradient Locality | Hyperparameter Setting | Computational Cost |
|-----------------|--------------------------------|--------------------|----------------------|-------------------|------------------------|--------------------|
| HTV [6], CTV [7] and SRBFuse [8] |                           | ×                  | ×                    | local easy          | easy                   | ○                  |
| SSTV [9]        |                           | ×                  | ×                    | semilocal easy     | ×                      | ●                  |
| LRWT [10]       |                           | ×                  | △                    | semilocal easy     | △                      | ×                  |
| NLSTTV [11]     |                           | ×                  | ×                    | nonlocal easy      | ×                      | ×                  |
| NLTV [12]       |                           | ×                  | △                    | local difficult    | ×                      | ×                  |
| HTV + LRM [13]  |                           | ×                  | △                    | local difficult    | ×                      | ×                  |
| LRWT + LRM [10] |                           | ×                  | △                    | nonlocal difficult | ×                      | ×                  |
| ASTV [16]       |                           | ×                  | △                    | local easy         | ○                      | ●                  |
| SSTV [17]       |                           | △                  | ×                    | local difficult    | ○                      | ○                  |
| ASSTV [18]      |                           | △                  | ×                    | local difficult    | ○                      | ○                  |
| ASSTV + LRM [18]|                           | ∆                  | △                    | local difficult    | ○                      | ○                  |
| SSTV + LRM [19] |                           | ∆                  | △                    | local difficult    | ○                      | ○                  |
| HSSTV           |                           | △                  | △                    | local easy         | ○                      | ○                  |
where $d_{v,ij}$ and $d_{h,ij}$ are vertical and horizontal differences for $i$th pixel of $j$th band in an HS image, respectively. From this definition, one can see that HTV evaluates spatial piecewise smoothness but does not consider spectral correlation, resulting in spatial oversmoothing. This will be empirically shown in Sec. V.

Adesso et al. proposed to use CTV \cite{51} for HS image inpainting \cite{7}. CTV is defined by

$$\text{CTV}(u) := \left( \sum_{i=1}^{N} \left( \sum_{j=1}^{B} \left( |d_{v,i,j}|^p + |d_{h,i,j}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \right)^{\frac{1}{q}}. $$

It evaluates spatial piecewise smoothness by using $\ell_{p,q,r}$ norm. In addition, the method can also use the Schatten-p norm as

$$\text{CTV}(u) := \left( \sum_{i=1}^{N} \left( \left\| \begin{array}{c} d_{v,1,i} \\ \vdots \\ d_{v,B,i} \\ d_{h,1,i} \\ \vdots \\ d_{h,B,i} \end{array} \right\|_{Sp}^{r} \right) \right)^{\frac{1}{r}}. $$

CTV can be seen as a generalization of HTV, which is equivalent to HTV when $p = 2, q = 2$ and $r = 1$. In \cite{7}, the authors experimentally show that CTV with $\ell_{2,2,1}$ norm achieves the best performance, which means that the limitation of CTV in HS image restoration is the same as HTV.

He et al. proposed ASSTV \cite{18} for HS image denoising. ASSTV simultaneously evaluates direct spatial and spectral differences, which is defined by

$$\text{ASSTV}(u) := \sum_{i=1}^{NB} \left\| \left( \tau_v d_{v,i} \right) + \left( \tau_h d_{h,i} \right) \right\|_1, $$

where $d_{v,i}$, $d_{h,i}$, and $d_{b,i}$ are vertical, horizontal, and spectral differences for the $i$th pixel of an HS image, respectively, and $\tau_v$, $\tau_h$ and $\tau_b > 0$ are balancing parameters for each difference (Fig. 1 blue lines). Although the parameters play a very important role, their suitable values are changed for each HS image and noise intensity. Therefore, their settings are very difficult.

Aggarwal and Majumder proposed SSTV \cite{17} for HS image denoising, which is defined as follows:

$$\text{SSTV}(u) := \sum_{i=1}^{NB} \left\| \left( d_{vb,i} - d_{hb,i} \right) \right\|_1, $$

where $d_{vb,i}$ and $d_{hb,i}$ are vertical-spectral and horizontal-spectral differences for $i$th pixel of the HS image, respectively. SSTV considers spectral piecewise smoothness of HS images together with spatial piecewise smoothness. Specifically, in \cite{8}, local spectral differences of an HS image are calculated before the calculation of local spatial differences (Fig. 1 yellow lines). SSTV is an effective and computationally efficient regularization technique for HS image restoration, and outperforms several popular regularization methods that are not limited to TVs \cite{6, 12, 13}. However, SSTV does not evaluate direct spatial differences, and so it often causes undesirable artifacts, as will be demonstrated in Sec. V.
IV. PROPOSED METHOD

A. Hybrid Spatio Spectral Total Variation

Now, we propose a new regularization technique for HS image restoration, named HSSTV. Our TV simultaneously handles both direct local spatial differences and local spatio-spectral differences of an HS image. Here, $D_v$, $D_h$, and $D_b$ are vertical, horizontal, and spectral difference operators, respectively, and we denote a local spatial difference operator as $D = (D_v^t D_h^t)^t \in \mathbb{R}^{2NB \times NB}$. Then, HSSTV is defined by

$$\text{HSSTV}(u):= \|A_v u\|_{1,p} \text{ with } A_v:=\left(DD_b \omega D\right),$$

where $\|\cdot\|_{1,p}$ is the mixed $\ell_{1,p}$ norm, and $\omega \geq 0$. We assume $p = 1$ or $2$, i.e., the $\ell_1$ norm ($\|\cdot\|_{1,1} = \|\cdot\|_1$) or the mixed $\ell_{1,2}$ norm, respectively.

In [4], $DD_u u$ and $D u$ correspond to local spatio-spectral and direct local spatial differences, respectively, as shown in Fig. [1] (red lines). The weight $\omega$ adjusts the relative importance of direct spatial piecwise smoothness to spatio-spectral piecwise smoothness. HSSTV evaluates two kinds of smoothness by taking the $\ell_p$ norm ($p = 1$ or $2$) of these differences associated with each pixel and then summing up for all pixels, i.e., calculating the $\ell_1$ norm. Thus, it can be defined via the mixed $\ell_{1,p}$ norm. When we set $\omega = 0$ and $p = 1$, HSSTV recovers SSTV as [4], meaning that HSSTV can be seen as a generalization of SSTV.

Handling direct spatial differences in HSSTV is intended to suppress artifacts produced by only evaluating spatio-spectral differences as in SSTV, i.e., $\omega D u$ is supplemental to $D D_b u$. In addition, suppressing the norm of $D u$ too much would cause spatial oversmoothing, which is the limitation of HTV. Thus, the weight $\omega$ should be set to less than one, as will be demonstrated in Sec. XV.

B. HS Image Restoration by HSSTV

We consider to restore a clean HS image $\bar{u} \in \mathbb{R}^{NB}$ from an observation $v \in \mathbb{R}^M$ ($M \leq NB$) contaminated by an additive white Gaussian noise $n \in \mathbb{R}^{NB}$ with the standard deviation $\sigma$. The observation model is given by the following form:

$$v = \Phi u + n,$$

where $\Phi \in \mathbb{R}^{M \times NB}$ is a matrix representing a linear observation process, e.g., random sampling.

Based on the above model, we formulate HS image restoration using HSSTV as the following optimization problem:

$$\min_u \text{HSSTV}(u)$$

$$\text{s.t. } \begin{align*}
\Phi u &\in B_{\varepsilon,v}^2 := \{x \in \mathbb{R}^{NB}||v - x|| \leq \varepsilon\}, \\
u &\in [\mu_{\min}, \mu_{\max}]^{NB},
\end{align*}$$

where $B_{\varepsilon,v}^2$ is a $v$-centered $\ell_2$-norm ball with the radius $\varepsilon > 0$, and $[\mu_{\min}, \mu_{\max}]^{NB}$ is a dynamic range of an HS image ($\mu_{\min} < \mu_{\max}$). The first constraint manages data fidelity to the observation $v$. In many methods, data fidelity is included in an objective function, e.g., [4] (17) and [7] (4). On the other hand, as mentioned in [13], [18], [28], [35]–[42], such a constraint-type formulation facilitates the parameter setting because one can intuitively set $\varepsilon$ via observed information.

Since both constraints are closed convex sets, and HSSTV is a convex function, Prob. (6) is a constrained convex optimization problem.

Since HSSTV is not differentiable, and there are multiple hard constraints, we require a suitable iterative algorithm to solve it. In this paper, we adopt ADMM (see Sec. II-C). In what follows, we reformulate Prob. (6) into Prob. (1) to solve it by ADMM.

By using the indicator functions of the constraints, Prob. (6) can be rewritten as

$$\min_u \|A_v u\|_{1,p} + \tau_B^\varepsilon(\Phi u) + \tau_{[\mu_{\min}, \mu_{\max}]^{NB}}(u).$$

Note that from the definition of the indicator function, Prob. (7) exactly equals to Prob. (6). By letting

$$f : \mathbb{R}^{NB} \rightarrow \mathbb{R} : u \mapsto 0,$$

$$g : \mathbb{R}^{5NB+M} \rightarrow \mathbb{R} \cup \{\infty\} : (z_1, z_2, z_3) \mapsto \|z_1\|_{1,p} + \tau_B^\varepsilon(z_2) + \tau_{[\mu_{\min}, \mu_{\max}]^{NB}}(z_3),$$

$$G : \mathbb{R}^{NB} \rightarrow \mathbb{R}^{5NB+M} : u \mapsto (A_v u, \Phi u, u),$$

Prob. (7) is reduced to Prob. (1). The resulting algorithm based on ADMM is summarized in Alg. 1.

The update of $u$ in Alg. 1 derives from the following strictly-convex quadratic minimization:

$$u^{(n+1)} = \arg\min_u \frac{1}{2\gamma} (\|z_1 - A_v u - d_1\|^2 + \|z_2 - \Phi u - d_2\|^2 + \|z_3 - u - d_3\|^2).$$

One can obtain this quadratic minimization by substituting (8) and (10) for (2). Here, we should consider the structure of $\Phi$ because it affects the matrix inversion step (Step 2) in Alg. 1. If $\Phi$ is a block-circulant-with-circulant-blocks (BCCB) matrix, we can leverage 3DFFT to efficiently solve the inversion in Step 2 with the difference operators having periodic boundary, i.e., $A_v^\top A_v + \Phi^\top \Phi + I$ can be diagonalized by the 3D FFT matrix and its inverse. If $\Phi$ is a semi-orthogonal matrix, i.e., $\Phi \Phi^\top = \alpha I$ ($\alpha > 0$), we leave it to the update of $z_2$, which means that we replace $\tau_B^\varepsilon$ by $\tau_B^\varepsilon \circ \Phi$ in (2) and $\Phi u$ by $u$ in (10). This is because the proximity operator

```
Algorithm 1: ADMM method for Prob. (6)
input : $z_1^{(0)}, z_2^{(0)}, z_3^{(0)}, d_1^{(0)}, d_2^{(0)}, d_3^{(0)}$
while A stopping criterion is not satisfied do
1. $u^{(n+1)} = (A_v^\top A_v + \Phi^\top \Phi + I)^{-1} (A_v^\top (z_1 - d_1) + \Phi^\top (z_2 - d_2) + (z_3 - d_3))$
2. $z_1^{(n+1)} = \text{prox}_{\gamma_1 \|\cdot\|_{1,p}}(A_v u^{(n+1)} + d_1^{(n)})$
3. $z_2^{(n+1)} = \text{prox}_{\gamma_2 \Phi^\top \|\cdot\|_{1,p}}(\Phi u^{(n+1)} + d_2^{(n)})$
4. $z_3^{(n+1)} = \text{prox}_{\gamma_3 \mu_{\min}, \mu_{\max}}(\mu_{\min}, \mu_{\max})^\top (u^{(n+1)} + d_3^{(n)})$
5. $d_1^{(n+1)} = d_1^{(n)} + A_v u^{(n+1)} - z_1^{(n+1)}$
6. $d_2^{(n+1)} = d_2^{(n)} + \Phi u^{(n+1)} - z_2^{(n+1)}$
7. $d_3^{(n+1)} = d_3^{(n)} + u^{(n+1)} - z_3^{(n+1)}$
8. $n \leftarrow n + 1$
end while
```
of \( \ell_{2/n} \circ \Phi \) in this case can be computed by using \cite{44} Table 1.1-x as follows:

\[
\text{prox}_{\gamma \ell_{2/n} \circ \Phi}(x) = x + \alpha^{-1} \left( P_{\ell_{2/n}}(\Phi x) - \Phi x \right).
\]

If \( \Phi \) is a sparse matrix, we offer to use a preconditioned conjugate gradient method \cite{45} for approximately solving the inversion, or to apply primal-dual splitting methods \cite{46, 48} instead of ADMM\cite{49}. Otherwise, randomized image restoration methods using stochastic proximal splitting algorithms \cite{49}–\cite{52} might be useful for reducing the computational cost.

For the update of \( z_1 \), the proximity operators are reduced to simple soft-thresholding type operations: for \( \gamma > 0 \) and for \( i = 1, \ldots, 4N_B \), (i) in the case of \( p = 1 \),

\[
[\text{prox}_{\gamma \ell_{1}}(x)]_i = \text{sgn}(x_i) \max\{ |x_i| - \gamma, 0 \},
\]

where \( \text{sgn} \) is the sign function, and (ii) in the case of \( p = 2 \),

\[
[\text{prox}_{\gamma \ell_{2}}(x)]_i = \max\left\{ 1 - \gamma \left( \sum_{j=0}^{3} x_{i+jN_B}^2 \right)^{-\frac{1}{2}}, 0 \right\} x_i,
\]

where \( i := ((i - 1) \mod NB) + 1 \).

The update of \( z_2 \) and \( z_3 \) require the proximity operators of the indicator functions of \( B_{2,\tilde{}} \) and \( [\mu_{\min}, \mu_{\max}]^{NB} \), respectively, which equal to the metric projections onto them (see Sec. II-B). Specifically, the metric projection onto \( B_{2,\tilde{}} \) is given by

\[
P_{B_{2,\tilde{}}}(x) = \begin{cases} x, & \text{if } x \in B_{2,\tilde{}} \vspace{1mm} \\ v + \frac{\varepsilon(x-v)}{\|x-v\|}, & \text{otherwise}
\end{cases}
\]

and onto \( [\mu_{\min}, \mu_{\max}]^{NB} \) is given, for \( i = 1, \ldots, NB \), by

\[
(P_{[\mu_{\min}, \mu_{\max}]^{NB}}(x))_i = \min\{ \max\{ x_i, \mu_{\min} \}, \mu_{\max} \}.
\]

V. EXPERIMENTS

We demonstrate the advantages of HSSTV by applying it to two specific HS image restoration problems: denoising and CS reconstruction. In these experiments, we use 13 HS images taken from the SpecTIR \cite{53}, MultiSpec \cite{54}, and GIC \cite{55}, where their dynamic range were normalized into \([0, 1]\).

HSSTV is compared with HTV \cite{6}, SSTV \cite{17}, and ASSTV \cite{18}. For a fair comparison, we replaced HSSTV in Prob. 6 with HTV, SSTV or ASSTV, and solved the problem by ADMM. In the denoising experiment, we also compare HSSTV with BM4D \cite{56}, which is known to be a state-of-the-art denoising method for 3D signals. Note that BM4D and a recent CNN-based HS image denoising method \cite{57} cannot be represented as explicit regularization functions and are fully customized to denoising tasks. In contrast, our HSSTV can be used as a building block in various HS image restoration methods based on optimization. In addition, BM4D is a nonlocal method, and thus, it has several limitations mentioned in Sec. I. Meanwhile, CNN-based methods strongly depend on what training data are used, which means that they cannot adapt to a wide range of noise intensity. Thus, the design concepts of these methods are different from TVs, and the comparison of HSSTV and BM4D is just a reference. We also remark that we do not compare HSSTV with LRM-based techniques. This is because our focus is to evaluate the potential of local TV regularization techniques including HTV, SSTV, and ASSTV, and our HSSTV. It should be noted that these TVs can be combined with LRM.

To quantitatively evaluate restoration performance, we use the peak signal-to-noise ratio (PSNR) [dB] index and the structural similarity (SSIM) \cite{58} index between a true HS image and a restored HS image \( \hat{u} \). PSNR is defined by \( 10 \log_{10}(NB/\|\hat{u} - u\|^2) \), and the higher the value is, the more similar the two images are. SSIM is an image quality assessment index based on the human vision system, which is defined as follows:

\[
\text{SSIM}(u, \hat{u}) = \frac{1}{P} \sum_{i=1}^{P} \text{SSIM}_i(u, \hat{u})
\]

\[
\text{SSIM}_i(u, \hat{u}) = \frac{(2\mu_u \mu_{\hat{u}} + C_1)(2\sigma_{u,\hat{u}} + C_2)}{\left( \mu_u^2 + \mu_{\hat{u}}^2 + C_1 \right)\left( \sigma_u^2 + \sigma_{\hat{u}}^2 + C_2 \right)},
\]

where \( u_i \) and \( \hat{u}_i \) are the \( i \)th pixel-centered local patches of a restored HS image and a true HS image, respectively, \( P \) is the number of patches, \( \mu_u \) and \( \mu_{\hat{u}} \) is the average values of the local patches of the restored and true HS images, respectively, \( \sigma_u \) and \( \sigma_{\hat{u}} \) represent the variances of \( u_i \) and \( \hat{u}_i \), respectively, and \( \sigma_{u,\hat{u}} \) denotes the covariance between \( u_i \) and \( \hat{u}_i \). Moreover, \( C_1 \) and \( C_2 \) are two constants, which avoid the numerical instability when either \( \sigma_u^2 + \sigma_{\hat{u}}^2 \), or \( \sigma_u^2 + \sigma_{\hat{u}}^2 \), is very close to zero. SSIM gives a normalized score between zero and one, where the maximum value means that \( u \) equals to \( \hat{u} \).

We set the max iteration number, the stepsize \( \gamma \) and the stopping criterion of ADMM to 5000, 0.1 and \( \|u^{(n)} - u^{(n+1)}\| < 0.01 \), respectively.

A. Gaussian Noise Removal

First, we conducted an experiment on Gaussian noise removal of HS images. We generated noisy HS images by adding Gaussian noise with \( \sigma \) to true HS images. In this case, \( \Phi = I \) in \cite{5}, and the radius \( \varepsilon \) in Prob. 6 was set to \( \sqrt{NB}\sigma^2 \). In the ASSTV case, we set the weights \( \tau_{v} = \tau_{h} = \tau_{b} = 1 \).

In Tab. \cite{1} we show PSNR and SSIM of the denoised HS images by each method for various \( \sigma \) and HS images. The balancing weight \( \omega \) in HSSTV is set to 0.08 for the \( \ell_{1} \)-norm case and 0.06 for the \( \ell_{1/2} \)-norm case. For all HS images, \( \sigma \) and quality measures, HSSTV outperforms HTV, SSTV, and ASSTV. In addition, even though HSSTV does not utilize nonlocal information, the denoising ability is better than BM4D for most cases. We also found that SSTV does not work well when \( \sigma \) is large. On the other hand, HSSTV is effective for a wide range of noise intensity. This would be because HSSTV evaluates direct spatial piecewise smoothness.

Fig. \cite{2} shows the resulting images on Scuwanne \( (\sigma = 0.1, \text{top}) \) and DC \( (\sigma = 0.2, \text{bottom}) \) with their PSNR (left) and SSIM (right). Here, we depict these HS images as RGB images (R = 8th, G = 16th and B = 32nd bands). One can see that the results by HTV and ASSTV lose spacial details, and noise remains in the results by SSTV. In the case of \( \sigma = 0.1 \),
BM4D can restore a high quality HS image, but it produces spectral artifacts in the case of \( \sigma = 0.2 \). This means that BM4D cannot preserve spectral information when an observed image is heavily contaminated by noise. In contrast, HSTV can restore HS images preserving both details and spectral information without artifacts.

Fig. 3 plots PSNR or SSIM of the results by HSTV versus \( \sigma \) for various values of \( \omega \) changed from 0.01 to 0.2, where the values of PSNR and SSIM are averaged over the 13 HS images. One can see that \( \omega \in [0.05, 0.1] \) is a good choice in most cases.

Fig. 4 plots bandwise PSNR and SSIM (left) and spatial and spectral responses (right) of the denoised Sauknee image, where \( \sigma = 0.1 \). The graphs regarding bandwise PSNR and SSIM show that HSTV achieves higher-quality restoration than HTV, SSTV, and ASSTV for all bands and BM4D for most bands. The graph (c) plots the spatial response of the 243rd row of the 30th band. In the same way, the graph (d) plots the spectral response of the 243rd row and 107th col. We can see that the spatial response of the results by HSTV and ASSTV are too smooth compared with the true one. On the other hand, there exist undesirable variations in the spatial response of the result by SSTV. In contrast, BM4D and HSTV restore similar responses to the true one. In the graph (d), one can see (i) the shape of the spectral responses of the results by HSTV and SSTV are similar to the that of true one, but the mean values are larger than the true one, (ii) the spectral response of the results by ASSTV and BM4D are too smooth and different from the true one, and (iii) HSTV can restore a spectral response very similar to the true one.
Fig. 2. Resulting HS images with their PSNR (left) and SSIM (right) in the Gaussian denoising experiment (top: Suwannee, \( \sigma = 0.1 \), bottom: DC, \( \sigma = 0.2 \)).

Fig. 3. PSNR (top) or SSIM (bottom) versus \( \omega \) in (4) in the Gaussian denoising experiment.

Fig. 4. Bandwise PSNR and SSIM and spatial and spectral responses in the Gaussian denoising experiment (Suwannee).

B. Real Noise Removal

We also examine HTV, SSTV, ASSTV, BM4D, and HSSTV on an HS image with real noise. We selected noisy 16 bands from Suwannee and use it as a real observed HS image \( v \). Here, we experimentally selected \( \sigma = 0.04 \), and set the radius parameter \( \varepsilon \) and the weights in ASSTV like Sec. V-A.

Fig. 5 shows the results, where the HS images are depicted as RGB images (R = 2nd, G = 6th and B = 13rd bands). The results by HTV and ASSTV have spatial oversmoothing, and SSTV, ASSTV, and BM4D produce artifacts. In contrast, the result by HSSTV can restore a detail-preserved HS image without artifacts.

C. Compressed Sensing Reconstruction

We conducted an experiment on compressed sensing (CS) reconstruction [59], [60]. The CS theory says that high-dimensional signal information can be reconstructed from incomplete random measurements by exploiting some sparsity. In general, HS imaging captures an HS image by scanning 1D spatial and spectral information, because it senses spectral one
by dispersing the incident light. Therefore, shooting an HS image spends much time, and so capturing moving objects is very difficult in HS imaging. To overcome the drawback, one-shot HS imaging based on CS has been actively studied [4, 5].

In this experiment, we assume that \( \Phi \in \mathbb{R}^{M \times N B} \) in (5) is a random sampling matrix with the sampling rate \( r \) (0 < \( r < 1 \) and \( M = r NB \)). Here, since \( \Phi \) is a semi-orthogonal matrix, we can efficiently solve the problem as explained in Sec. V-B. We set \( r = 0.2 \) or 0.4, \( \sigma = 0.1 \) and \( \varepsilon = \sqrt{r NB \sigma^2} \) in (6). In the ASSTV case, we set the parameters \( (\tau_v, \tau_h, \tau_r) = (1, 1, 0.5) \).

Tab. III shows PSNR and SSIM of the reconstructed HS images. For all \( r \) and HS images, both PSNR and SSIM of the results by HSSTV are higher than that by HTV, SSTV, and ASSTV.

Fig. 6 is the reconstructed results on KSC and Reno with the random sampling ratio \( r = 0.4 \) and 0.2, respectively. Here, the HS images are depicted as RGB images (R = 8th, G = 16th and B = 32nd bands). One can see that (i) HTV causes spatial oversmoothing, (ii) SSTV produces artifacts and spectral distortion, where it appears as the difference from the color of the true HS images, and (iii) the results by ASSTV have spatial oversmoothing and spectral distortion. On the other hand, HSSTV reconstructs meaningful details without both artifacts and spectral distortion.

Fig. 7 plots PSNR or SSIM of the results by HSSTV versus \( \omega \) averaged over the 13 HS images for each \( r \). The graphs show that \( \omega \in [0.05, 0.1] \) is a good choice in most cases. In comparison with Fig. (5), the suitable range of \( \omega \) in CS reconstruction is almost the same as that in denoising.

Fig. 8 plots bandwise PSNR or SSIM (left) and spatial and spectral responses (right) (Suwannee, \( r = 0.2 \)). According to bandwise PSNR and SSIM, one can see that HSSTV achieves higher-quality reconstruction for all bands than HTV, SSTV and ASSTV. The graphs (c) and (d) plot the spatial and spectral responses of the same position in Sec. V-A. The graph (c) shows that (i) the spatial response of the results by HTV and ASSTV are oversmoothing, (ii) SSTV produces undesirable variation, and (iii) the spatial response reconstructed by HSSTV is similar to the true one. In the graph (d), HTV generates undesirable variation, and ASSTV causes oversmoothing. Thanks to the evaluation of spatio-spectral piecewise smoothness, SSTV reconstructs similar

---

### TABLE III

| \( \sigma \) | \( r \) | PSNR | SSIM |
|---|---|---|---|
| | | HTV | SSTV | ASSTV | \( \ell_1 \) -HSSTV | \( \ell_1,2 \) -HSSTV |
| | | HTV | SSTV | ASSTV | \( \ell_1 \) -HSSTV | \( \ell_1,2 \) -HSSTV |
| Beltville | 0.1 | 27.96 | 28.35 | 29.35 | 31.15 | 30.71 | 0.0829 | 0.0013 | 0.8836 |
| | | 26.23 | 24.34 | 24.12 | 29.63 | 30.38 | 0.6363 | 0.4348 | 0.6108 |
| | | 26.47 | 25.69 | 25.39 | 31.37 | 31.80 | 0.6810 | 0.5739 | 0.6633 |
| Suwannee | 0.1 | 24.69 | 27.33 | 24.71 | 29.70 | 29.29 | 0.6096 | 0.7522 | 0.6245 |
| | | 23.31 | 24.16 | 22.69 | 27.98 | 27.59 | 0.5215 | 0.6120 | 0.5037 |
| | | 29.94 | 28.21 | 28.59 | 34.36 | 34.34 | 0.7665 | 0.6826 | 0.7852 |
| | | 28.77 | 25.79 | 26.38 | 32.97 | 32.95 | 0.7368 | 0.5057 | 0.7207 |
| | | 26.99 | 27.82 | 26.49 | 31.80 | 31.61 | 0.6769 | 0.7414 | 0.6730 |
| | | 25.57 | 25.57 | 24.52 | 30.22 | 30.04 | 0.6202 | 0.6276 | 0.5940 |
| DC | 0.1 | 26.10 | 27.81 | 25.13 | 30.32 | 30.15 | 0.6683 | 0.7551 | 0.6460 |
| | | 26.66 | 24.79 | 22.86 | 28.79 | 28.63 | 0.6014 | 0.6225 | 0.5519 |
| | | 30.54 | 27.63 | 29.33 | 31.36 | 31.04 | 0.7497 | 0.3066 | 0.7187 |
| | | 29.99 | 25.11 | 28.19 | 30.71 | 30.46 | 0.7366 | 0.3488 | 0.7465 |
| | | 28.30 | 28.34 | 28.63 | 34.10 | 34.03 | 0.7660 | 0.6814 | 0.7544 |
| | | 28.11 | 27.00 | 26.59 | 32.67 | 32.60 | 0.7318 | 0.6008 | 0.7032 |
| | | 25.66 | 29.66 | 25.41 | 31.96 | 31.83 | 0.6082 | 0.8386 | 0.5932 |
| | | 24.26 | 27.17 | 23.24 | 30.20 | 30.08 | 0.5103 | 0.7319 | 0.4434 |
| | | 25.83 | 29.86 | 25.61 | 32.45 | 32.21 | 0.6357 | 0.7962 | 0.6275 |
| | | 24.50 | 27.54 | 23.61 | 30.56 | 30.38 | 0.5502 | 0.6917 | 0.5069 |
| | | 24.95 | 30.02 | 26.98 | 32.88 | 32.70 | 0.6067 | 0.7830 | 0.6318 |
| | | 31.19 | 27.69 | 30.18 | 35.43 | 35.21 | 0.8877 | 0.6153 | 0.8566 |
| | | 29.94 | 25.28 | 28.09 | 34.05 | 34.10 | 0.8404 | 0.4620 | 0.8302 |
| | | 30.67 | 27.93 | 28.19 | 34.45 | 34.14 | 0.8647 | 0.6595 | 0.8489 |
| | | 28.68 | 24.15 | 24.94 | 32.71 | 32.36 | 0.8387 | 0.4810 | 0.8005 |

---

Fig. 5. Resulting HS images on the real noise removal experiment.
spectral response to the true one, but the mean values are larger than the true one. HSSTV achieves the most similar reconstruction of spectral response among all the TVs.

VI. CONCLUSION

We have proposed a new TV regularization technique for HS image restoration, named HSSTV. It evaluates both the direct spatial and spatio-spectral piecewise smoothness of an HS image, leading to an effective regularization for HS images. We also formulate HSSTV-regularized HS restoration as a constrained convex optimization problem and develop an efficient algorithm based on ADMM for solving the problem. Experimental results on Gaussian denoising, real noise removal, and CS reconstruction demonstrate the effectiveness and utility of HSSTV.

In this paper, we have focused on a single use of HSSTV. Finally, we would like to note that our HSSTV can be combined with other regularization techniques, for example, LRM and texture regularization [61], [62].

ACKNOWLEDGMENT

The work was partially supported by JSPS Grants-in-Aid (18J20290, 17K12710, 16K12457) and JST-PRESTO.

REFERENCES

[1] C. I. Chang, Hyperspectral imaging: techniques for spectral detection and classification. Springer Science & Business Media, 2003, vol. 1.
[2] A. P. 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. Licciardi, and J. Chanussot, “Noise reduction in hyperspectral imagery: Overview and application,” Remote Sens., vol. 10, no. 3, p. 482, 2018.
[4] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, “Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection,” IEEE Signal Process. Magazine, vol. 31, no. 1, pp. 116–126, 2014.

[5] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, “Compressive coded aperture spectral imaging: An introduction,” IEEE Signal Process. Magazine, vol. 31, no. 1, pp. 105–115, 2014.

[6] Q. Yuan, L. Zhang, and H. Shen, “Hyperspectral image denoising employing a spectral–spatial adaptive total variation model,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 10, pp. 3660–3677, 2012.

[7] P. Adesso, D. M. D. Mora, L. Condat, R. Restaino, G. Vivone, D. Picone, and J. Chanussot, “Hyperspectral image inpainting based on collaboratively total variation,” in Proc. IEEE Int. Conf. Image Process. (ICIP), 2016, pp. 4382–4386.

[8] X. Zheng, Y. Yuan, and X. Lu, “Hyperspectral image denoising by fusing the selected related bands,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 5, pp. 2596 – 2609, 2018.

[9] S. Lefkimmiatis and A. R. P. M. Unser, “Structure tensor total variation,” SIAM J. Imag. Sci., vol. 8, no. 2, pp. 1090–1122, 2015.

[10] H. Liu, P. Sun, Q. Du, Z. Wu, and Z. Wei, “Hyperspectral image restoration based on low-rank recovery with a local neighborhood weighted spectral-spatial total variation model,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 3, pp. 1–13, 2018.

[11] S. Lefkimmiatis and S. Osher, “Nonlocal structure tensor functionalities for image regularization,” IEEE Trans. Comput. Imag., vol. 1, no. 1, pp. 16–29, 2015.

[12] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, “A nonlocal structure tensor-based approach for multicomponent image recovery problems,” IEEE Trans. Image Process., vol. 23, no. 12, pp. 5531–5544, 2014.

[13] W. He, H. Zhang, L. Zhang, and H. Shen, “Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 178–188, 2016.

[14] H. K. Aggarwal and A. Majumdar, “Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation,” IEEE Selected Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 9, pp. 4257–4266, 2016.

[15] H. Li, P. Sun, H. Liu, Z. Wu, and Z. Wei, “Non-convex low-rank approximation for hyperspectral image recovery with weighted total variation regularization,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 07 2018, pp. 2733–2736.

[16] S. Ono, K. Shirai, and M. Okuda, “Vectorial total variation based on arranged structure tensor for multichannel image restoration,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 2016, pp. 4528–4532.

[17] H. K. Aggarwal and A. Majumdar, “Hyperspectral image denoising using spatio-spectral total variation,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 3, pp. 442–446, 2016.

[18] W. He, H. Zhang, H. Shen, and L. Zhang, “Hyperspectral image denoising using explicit low-rank matrix recovery and global spatial-spectral total variation,” IEEE Selected Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 3, pp. 713–729, 2018.

[19] W. Cao, K. Wang, G. Han, J. Yao, and A. Cichocki, “A robust pca approach with noise structure learning and spatial–spectral low-rank modeling for hyperspectral image restoration,” IEEE Selected Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 10, pp. 3863–3879, 2018.

[20] Y. Wang, J. Feng, Q. Zhao, Y. Leung, X. Zhao, and D. Meng, “Hyperspectral image restoration via total variation regularized low-rank tensor decomposition,” IEEE Selected Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 4, pp. 1227–1234, 2017.

[21] Q. Wang, Z. Wu, J. Jin, T. Wang, and Y. Shen, “Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising,” Signal Processing, vol. 142, pp. 11–26, 2018.

[22] L. Sun, T. Zhan, Z. Wu, L. Xiao, and B. Jeon, “Hyperspectral mixed denoising via spectral difference induced total variation and low-rank approximation,” Remote Sens., vol. 10, no. 12, p. 1956, 2018.

[23] T. Ince, “Hyperspectral image denoising using group low-rank and spatial-spectral total variation,” IEEE Access, vol. 7, pp. 52095–52109, 2019.

[24] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite elements approximations,” Comput. Math. Appl., vol. 2, pp. 17–40, 1976.

[25] J. Eckstein and D. Bertsekas, “On the Douglas-Rachford splitting method and proximal point algorithm for maximal monotone operators,” Math. Program., vol. 55, pp. 293–318, 1992.
[50] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb, “Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications,” SIAM J. Optim., vol. 28, no. 4, pp. 2783–2808, 2018.

[51] P. L. Combettes and J.-C. Pesquet, “Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration,” in Proc. Eur. Signal Process. Conf. (EUSIPCO), Aug 2016, pp. 1813–1817.

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

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

[54] “MultiSpec,” https://engineering.purdue.edu/biehl/MultiSpec.

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

[56] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, “Nonlocal transform-domain filter for volumetric data denoising and reconstruction,” IEEE Trans. Image Process., vol. 22, no. 1, pp. 119–133, 2013.

[57] W. Liu and J. Lee, “A 3-d atrous convolution neural network for hyperspectral image denoising,” IEEE Trans. Geosci. Remote Sens., 2019.

[58] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004.

[59] R. G. Baraniuk, “Compressive sensing,” IEEE Signal Process. Magazine, vol. 24, no. 4, 2007.

[60] E. Candès and M. Wakin, “An introduction to compressive sampling,” IEEE Signal Process. Magazine, vol. 25, no. 2, pp. 21–30, 2008.

[61] H. Schaeffer and S. Osher, “A low patch-rank interpretation of texture,” SIAM J. Imag. Sci., vol. 6, no. 1, pp. 226–262, 2013.

[62] S. Ono, T. Miyata, and I. Yamada, “Cartoon-texture image decomposition using blockwise low-rank texture characterization,” IEEE Trans. Image Process., vol. 23, no. 3, pp. 1128–1142, 2014.