Spatio-Spectral Structure Tensor Total Variation for Hyperspectral Image Denoising and Destriping

Shingo Takemoto, *Student Member, IEEE,* Kazuki Naganuma, Shunsuke Ono, *Member, IEEE,*

***Abstract*—We propose a novel regularization function, Graph-Aided Spatio-Spectral Total Variation (GASSTV), for hyperspectral (HS) image denoising. Spatio-Spectral Total Variation (SSTV), defined using second-order spatio-spectral differences, is a widely used regularization approach that effectively captures the spatio-spectral piecewise smoothness specific to HS images, enabling robust noise removal. However, SSTV uniformly computes adjacent differences in spatial and spectral dimensions, leading to the corruption of complex spatial structures and abrupt spectral variations in HS images. To address these limitations, we incorporate spatial and spectral graphs that explicitly reflect the spatial and spectral structures into the regularization model. Furthermore, we design a framework to construct these graphs from the noisy HS image. In addition, we formulate the mixed noise removal problem as a convex optimization problem involving GASSTV and develop an efficient algorithm based on the preconditioned primal-dual splitting method to solve it. Experimental results on mixed noise removal demonstrate the superior performance of GASSTV compared with existing HS image regularization models.**

***Index Terms*—Hyperspectral image, denoising, destriping, spatio-spectral regularization, total variation, structure tensor**

I. Introduction

HYPERSPECTRAL (HS) images, with their rich spectral information over 100 bands, are widely applied in diverse fields such as agriculture, mineralogy, astronomy, and biotechnology [1]–[4]. Despite their potential, HS images are contaminated with noise during acquisition, which adversely affects the performance of subsequent analyses, including anomaly detection [5], [6], classification [7], and unmixing [8], [9]. Therefore, HS image denoising is an essential preprocessing step [10]–[13].

For HS image denoising tasks, Spatio-Spectral Total Variation (SSTV) [14] is widely known as a powerful regularization method applied in state-of-the-art methods [**?**], [15]–[21]. SSTV effectively captures the both spatial and spectral piecewise smoothness of HS images by leveraging the ℓ_1-norm of spatial differences computed after spectral ones. However, SSTV evaluates neighborhood differences uniformly, which is limited to preserve 1) complex spatial structures and 2) spectral jumps.

As a promising extension of SSTV, Graph Spatio-Spectral Total Variation (GSSTV) [20] integrates a spatial graph-based weighted difference operator into SSTV, addressing challenge 1) i.e., preserving complex spatial structures. However, GSSTV leaves residual noise in low-weight regions of the spatial graph, and does not address the spectral limitations of SSTV. Now a natural question arises: *can we design a regularization method that enhances the noise removal capabilities of SSTV while capturing complex spatial structures and spectral jumps?*

In this paper, we propose a novel method that incorporates both spatial and spectral graph-based TVs into SSTV, called Graph-Aided Spatio-Spectral Total Variation (GASSTV). The main contributions of this article are listed below:

- We propose a novel regularization method, Graph-Aided Spatio-Spectral Total Variation (GASSTV), which incorporates spatial and spectral graph-based weighted TVs separately from SSTV. These graph-based TVs enhance the noise removal capabilities of SSTV while preserving complex spatial structures and spectral jumps. Furthermore, we robustly construct the spectral graph from the noisy HS image (Refer to GSSTV for constructing the spatial graph).
- We formulate the HS image denoising problem as a constrained convex optimization problem involving GASSTV. To simplify parameter tuning, we impose data fidelity and sparse noise characterization terms as constraints rather than adding them to the objective function.
- We develop an efficient algorithm based on Preconditioned Primal-Dual Splitting Method (P-PDS) [22], [23] to solve our optimization problem. Unlike standard PDS [24], [25], P-PDS automatically determines appropriate stepsizes.

Finally, we demonstrate the effectiveness of GASSTV by comparing it with state-of-the-art HS image regularization methods through HS image denoising experiments.

II. Preliminaries

A. Notations

Throughout this paper, we denote vectors and matrices by boldface lowercase letters (e.g., $\mathbf{x}$) and boldface capital letters (e.g., $\mathbf{X}$), respectively. We treat an HS image, denoted by $\mathbf{u}$ with N_1 vertical pixels, N_2 horizontal pixels, and N_3 bands. We denote the total number of elements in the HS image by $N = N_1N_2N_3$. For a matrix data $\mathbf{X} \in \mathbb{R}^{N_1 \times N_2}$, the value at the location (i, j) is denoted by $[\mathbf{X}]_{i,j}$. The ℓ_1-norm and the ℓ_2-norm of a vector $\mathbf{x} \in \mathbb{R}^N$ are defined as $\|\mathbf{x}\|_1 := \sum_{n=1}^{N} |x_n|$

S. Takemoto is with the Department of Computer Science, Institute of Science Tokyo, Yokohama, 226-8501, Japan (e-mail: takemoto.s.e908@m.isct.ac.jp).

K. Naganuma is with the Institute of Engineering of Tokyo University of Agriculture and Technology, Tokyo, 184-8588, Japan (e-mail: k-naganuma@go.tuat.ac.jp).

S. Ono is with the Department of Computer Science, Institute of Science Tokyo, Yokohama, 226-8501, Japan (e-mail: ono.s.5af2@m.isct.ac.jp).

This work was supported in part by Grant-in-Aid for JSPS Fellows Grant Number JP24KJ1068 and 23KJ0912, in part by JSPS KAKENHI under Grant 22H03610, 22H00512, 23H01415, 23K17461, 24K03119, and 24K22291, and in part by JST PRESTO under Grant JPMJPR21C4, JST AdCORP under Grant JPMJKB2307, and JST ACT-X Grant JPMJAX23CJ.

and $\|\mathbf{x}\|_2 := \sqrt{\sum_{n=1}^{N} x_n^2}$, respectively, where x_n represents the n-th entry of $\mathbf{x}$. The nuclear norm of a marix, which is the sum of all the singular values, is denoted by $\|\cdot\|_*$. For an HS image $\mathbf{u} \in \mathbb{R}^N$, let $\mathbf{D}_{\mathrm{v}} \in \mathbb{R}^{N\times N}$, $\mathbf{D}_{\mathrm{h}} \in \mathbb{R}^{N\times N}$, and $\mathbf{D}_{\lambda} \in \mathbb{R}^{N\times N}$ be the forward difference operators along the horizontal, vertical, and spectral directions, respectively, with the periodic boundary condition. Here, a spatial difference operator is denoted by $\mathbf{D}_{\mathrm{vh}} := \begin{pmatrix}\mathbf{D}_{\mathrm{v}}^\top & \mathbf{D}_{\mathrm{h}}^\top\end{pmatrix}^\top \in \mathbb{R}^{2N\times N}$. Using $\mathbf{D}_{\mathrm{v}}$, $\mathbf{D}_{\mathrm{h}}$, and $\mathbf{D}_{\lambda}$, we denote the second-order spatio-spectral differences by $\mathbf{D}_{\mathrm{v}}\mathbf{D}_{\lambda}\mathbf{u} \in \mathbb{R}^N$ and $\mathbf{D}_{\mathrm{h}}\mathbf{D}_{\lambda}\mathbf{u} \in \mathbb{R}^N$. Other notations will be introduced as needed.

B. Proximal Tools

In this chapter, we introduce basic proximal tools that play a central role in the optimization part of our method. Let f be a *proper lower semi-continuous convex function*.[1] Then, for $\gamma > 0$, the *proximity operator* of f is defined by

$$\mathrm{prox}_{\gamma,f}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathbb{R}^N} f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{x}-\mathbf{y}\|_2^2. \tag{1}$$

The *Fenchel–Rockafellar conjugate function* f^* of the function f is defined by

$$f^*(\mathbf{x}) := \sup_{\mathbf{y}} \langle \mathbf{x}, \mathbf{y}\rangle - f(\mathbf{y}), \tag{2}$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. Thanks to a generalization of Moreau's identity [26], the proximity operator of f^* is calculated as

$$\mathrm{prox}_{\gamma,f^*}(\mathbf{x}) = \mathbf{x} - \gamma\,\mathrm{prox}_{\frac{1}{\gamma}f}\left(\frac{1}{\gamma}\mathbf{x}\right). \tag{3}$$

The indicator function of a set $C \subset \mathbb{R}^N$, denoted by ι_C, is defined as

$$\iota_C(\mathbf{x}) := \begin{cases} 0, & \text{if } \mathbf{x} \in C, \\ \infty, & \text{otherwise.} \end{cases} \tag{4}$$

The function ι_C is proper lower semi-continuous convex when C is nonempty and closed convex. The proximity operator of ι_C is equivalent to the projection onto C, as given by

$$\mathrm{prox}_{\iota_C}(\mathbf{x}) = P_C(\mathbf{x}) := \operatorname*{argmin}_{\mathbf{y}\in C} \|\mathbf{y}-\mathbf{x}\|_2. \tag{5}$$

C. Preconditoned Primal-Dual Splitting Method (P-PDS)

The standard PDS [24], [25] and P-PDS [22], on which our algorithm is based, are efficient algorithms for solving the following generic form of convex optimization problems:

$$\min_{\substack{\mathbf{x}_1,\dots,\mathbf{x}_N,\\ \mathbf{y}_1,\dots,\mathbf{y}_M}} \sum_{i=1}^{N} f_i(\mathbf{x}_i) + \sum_{j=1}^{M} g_j(\mathbf{y}_j) \quad \text{s.t.} \begin{cases} \mathbf{y}_1 = \sum_{i=1}^{N} \mathbf{A}_{1,i}\mathbf{x}_i, \\ \vdots \\ \mathbf{y}_M = \sum_{i=1}^{N} \mathbf{A}_{M,i}\mathbf{x}_i, \end{cases} \tag{6}$$

[1]A function $f : \mathbb{R}^N \to (-\infty, \infty]$ is called a proper lower semi-continuous convex function if $\{\mathbf{x} \in \mathbb{R}^N | f(\mathbf{x}) < \infty\}$ is nonempty, $\{\mathbf{x} \in \mathbb{R}^N | f(\mathbf{x}) \le \alpha\}$ is closed for every $\alpha \in \mathbb{R}$, and $f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and $\lambda \in (0, 1)$.

where $f_i (i = 1, \dots, N)$ and $g_j (j = 1, \dots, M)$ are lower semi-continuous proper convex functions, $\mathbf{x}_i \in \mathbb{R}^{n_i}$ $(i = 1, \dots, N)$ are primal variables, $\mathbf{y}_j \in \mathbb{R}^{m_j}$ $(j = 1, \dots, M)$ are dual variables, and $\mathbf{A}_{j,i} \in \mathbb{R}^{m_j \times n_i}$ $(i = 1, \dots, N,\ j = 1, \dots, M)$ are linear operators.

These methods solve Prob. (6) by the following iterative procedures:

$$\begin{cases} \mathbf{x}_1^{(t+1)} \leftarrow \mathrm{prox}_{\gamma_{1,1},f_1}\left(\mathbf{x}_1^{(t)} - \gamma_{1,1}\left(\sum_{j=1}^{M} \mathbf{A}_{j,1}^\top \mathbf{y}_j^{(t)}\right)\right), \\ \vdots \\ \mathbf{x}_N^{(t+1)} \leftarrow \mathrm{prox}_{\gamma_{1,N},f_N}\left(\mathbf{x}_N^{(t)} - \gamma_{1,N}\left(\sum_{j=1}^{M} \mathbf{A}_{j,N}^\top \mathbf{y}_j^{(t)}\right)\right), \\ \mathbf{x}_i^{'} = 2\mathbf{x}_i^{(t+1)} - \mathbf{x}_i^{(t)}\ (\forall i = 1, \dots, N), \\ \mathbf{y}_1^{(t+1)} \leftarrow \mathrm{prox}_{\gamma_{2,1},g_1^*}\left(\mathbf{y}_1^{(t)} - \gamma_{2,1}\left(\sum_{i=1}^{N} \mathbf{A}_{1,i}\mathbf{x}_i^{'}\right)\right), \\ \vdots \\ \mathbf{y}_M^{(t+1)} \leftarrow \mathrm{prox}_{\gamma_{2,M},g_M^*}\left(\mathbf{y}_M^{(t)} - \gamma_{2,M}\left(\sum_{i=1}^{N} \mathbf{A}_{M,i}\mathbf{x}_i^{'}\right)\right), \end{cases} \tag{7}$$

where $\gamma_{1,i} (i = 1, \dots, N)$ and $\gamma_{2,j} (j = 1, \dots, M)$ are the stepsize parameters.

Here, we introduce the convergence property of P-PDS. For the convergence analysis, we define the diagonal matrices of the stepsize parameters as follows:

$$\begin{aligned} \mathbf{\Gamma}_1 &= \mathrm{diag}(\gamma_{1,1}\mathbf{I}_{n_1}, \dots, \gamma_{1,N}\mathbf{I}_{n_N}), \\ \mathbf{\Gamma}_2 &= \mathrm{diag}(\gamma_{2,1}\mathbf{I}_{m_1}, \dots, \gamma_{2,M}\mathbf{I}_{m_M}), \end{aligned} \tag{8}$$

where $\mathbf{I}_{n_i} \in \mathbb{R}^{n_i \times n_i}$ and $\mathbf{I}_{m_j} \in \mathbb{R}^{m_j \times m_j}$ are the identity matrices. We define the linear operator including $\mathbf{A}_{j,i}$ as

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{1,1} & \cdots & \mathbf{A}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{M,1} & \cdots & \mathbf{A}_{M,N} \end{bmatrix}. \tag{9}$$

Then, we state the convergence property of P-PDS.

Theorem 1. *[64, Theorem 1] Let $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ be symmetric and positive definite matrices satisfying*

$$\|\mathbf{\Gamma}_1^{\frac{1}{2}} \circ \mathbf{A} \circ \mathbf{\Gamma}_2^{\frac{1}{2}}\|_{\mathrm{op}}^2 < 1. \tag{10}$$

Then, the sequence $\{\mathbf{x}_1^{(t)}, \dots, \mathbf{x}_N^{(t)}, \mathbf{y}_1^{(t)}, \dots, \mathbf{y}_M^{(t)}\}$ generated by the procedure in (7) *converges to an optimal solution of Prob.* (6)*.*

The standard PDS needs to adjust the appropriate stepsize parameters to satisfy the convergence conditions (10). On the other hand, P-PDS can automatically determine the stepsize parameters that guarantee convergence [22], [23]. According to [22], we summarize the stepsize design and their convergence property.

Lemma 1. *[64, lemma 2] Let the diagonal matrices $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2$ and the block matrix $\mathbf{A}$ be set as Eq.* (8) *and* (9)*, respectively. In paticular,*

$$\begin{aligned} \gamma_{1,i} &= \frac{1}{\sum_{j=1}^{M}\sum_{l=1}^{m_j} |[\mathbf{A}_{j,i}]_{l,1}|},\ (\forall i = 1, \dots, N), \\ \gamma_{2,j} &= \frac{1}{\sum_{i=1}^{N}\sum_{k=1}^{n_i} |[\mathbf{A}_{j,i}]_{1,k}|},\ (\forall j = 1, \dots, M), \end{aligned} \tag{11}$$

then the inquality in (10) *holds.*

III. Proposed Method

A. Spectral Graph-based Weighted Difference Operator

In constructing the spectral graph from a noisy HS image, we assume that when the HS image is segmented, each segment contains similar spectral characteristics. Based on this assumption, we compute the guide image $\mathbf{v}'$ that reflects the spatial structure of the target HS image according to Eq. (**??**), and divide the guide image into segments $\mathcal{S}_k (k = 1, \dots, K)$ using the k-means clustering method [**?**]. Let $\mathbf{L} \in \mathbb{R}^{N_1 \times N_2}$ denote the resulting label map, where $\mathbf{L}(i, j) = k$ indicates that the pixel at position (i, j) belongs to segment $\mathcal{S}_k$.

Next, for each segment, all pixels of the noisy HS image $\mathbf{v}$ belonging to $\mathcal{S}_k$ are averaged along the spatial direction to compute a representative spectrum $\mathbf{r}_k$ as

$$\mathbf{r}_k := \frac{1}{|\mathcal{S}_k|} \sum_{(i,j)\in\mathcal{S}_k} \mathbf{v}(i, j, :), \tag{12}$$

where $\mathbf{v}(i, j, :)$ denotes the spectral vector (of length N_3) at location (i, j). By averaging the spectral vectors within each segment, the spectra of the segment can be effectively characterized while reducing noise.

Here, the representative spectra $\{\mathbf{r}_1, \dots, \mathbf{r}_K\}$ are regarded as line graphs of length N_3. Each representative spectrum is assigned to the pixels of its corresponding segment, and by doing so, we construct a spectral graph $\mathcal{G}_\lambda(\mathbf{x}_\lambda, \mathcal{E}_\lambda, \mathbf{W}_\lambda)$ that consists of $N_1 \times N_2$ line graphs. The node vectors $\mathbf{x}_\lambda$ are defined as follows: for all i, j

$$\mathbf{x}_\lambda(i, j, :) := \mathbf{r}_{\mathbf{L}(i,j)}, \tag{13}$$

where $\mathbf{x}_\lambda(i, j, :)$ represents the spectral vector of length N_3 at location (i, j). The weight matrix $\mathbf{W}_\lambda \in \mathbb{R}^{N\times N}$ is a diagonal matrix whose entries are the weights $w \in (0, 1]$ defined by

$$w_i := \exp\left(-[\mathbf{D}_\lambda \mathbf{x}_\lambda]_i^2 / 2\sigma_\lambda^2\right), \tag{14}$$

where $[\mathbf{D}_\lambda \mathbf{x}_\lambda]_i^2$ represents the i-th element of the product, and squaring is applied to each element. Then, the weighted spectral difference operator $\mathbf{D}_{\mathcal{G}_\lambda}$ is defined via the graph $\mathcal{G}_\lambda$ as $\mathbf{D}_{\mathcal{G}_\lambda} := \mathbf{W}_\lambda \mathbf{D}_\lambda$.

B. GASSTV

Incorporating both the spatial graph $\mathcal{G}_{\text{sp}}$ and the spectral graph $\mathcal{G}_\lambda$, our *Graph-Aided Spatio-Spectral Total Variation* (GASSTV) is defined as follows:

$$\text{GASSTV}(\mathbf{u}) := \|\mathbf{D}_{\text{vh}} \mathbf{D}_\lambda \mathbf{u}\|_1 + \omega_1 \|\mathbf{D}_{\mathcal{G}_{\text{sp}}} \mathbf{u}\|_1 + \omega_2 \|\mathbf{D}_{\mathcal{G}_\lambda} \mathbf{u}\|_1, \tag{15}$$

where $\mathbf{D}_{\text{vh}} := \begin{pmatrix} \mathbf{D}_{\text{v}}^\top & \mathbf{D}_{\text{v}}^\top \end{pmatrix}^\top \in \mathbb{R}^{2N\times N}$ is the spatial difference operator, ω_1 and ω_2 are the parameters. The first term represents Spatio-Spectral Total Variation (SSTV) [14], while the second and third terms represent the spatial and spectral graph-based TV (spatial graph TV and spectral graph TV), respectively. SSTV effectively removes noise and preserves spatio-spectral continuity but struggles to capture complex spatial structures and sharp spectral jumps, resulting in the loss of important details. GASSTV addresses these issues by incorporating spatial and spectral graph TVs. By capturing complex spatial structures and spectral jumps with spatial and spectral graph TVs, respectively, GASSTV enhances noise removal while preserving important spatial and spectral details.

C. Problem Formulation

An observed HS image $\mathbf{v} \in \mathbb{R}^N$ contaminated by mixed noise is modeled by

$$\mathbf{v} = \bar{\mathbf{u}} + \bar{\mathbf{s}} + \mathbf{n}, \tag{16}$$

where $\bar{\mathbf{u}}$, $\bar{\mathbf{s}}$, and $\mathbf{n}$ represent a clean HS image, sparse noise such as outliers, and random noise, respectively.

Based on the observation model (16), we formulate HS image denoising problem by GASSTV as a constrained convex optimization problem with the following form:

$$\min_{\mathbf{u},\mathbf{s}} \text{GASSTV}(\mathbf{u}) \quad \text{s.t.} \begin{cases} \mathbf{u} + \mathbf{s} \in B_{2,\varepsilon}^{\mathbf{v}}, \\ \mathbf{s} \in B_{1,\alpha}, \\ \mathbf{u} \in R_{\underline{\mu},\bar{\mu}}, \end{cases} \tag{17}$$

where

$$B_{2,\varepsilon}^{\mathbf{v}} := \{\mathbf{x} \in \mathbb{R}^N |\ \|\mathbf{x} - \mathbf{v}\|_2 \le \varepsilon\}, \tag{18}$$

$$B_{1,\alpha} := \{\mathbf{x} \in \mathbb{R}^N |\ \|\mathbf{x}\|_1 \le \alpha\}, \tag{19}$$

$$R_{\underline{\mu},\bar{\mu}} := \{\mathbf{x} \in \mathbb{R}^N |\ \underline{\mu} \le x_i \le \bar{\mu}\ (i = 1, \dots, N)\}. \tag{20}$$

The first constraint serves as data-fidelity with the $\mathbf{v}$-centered ℓ_2-ball of the radius $\varepsilon > 0$. The second constraint characterizes sparse noise with the zero-centered ℓ_1-ball of the radius $\alpha > 0$. Since we impose the above constraints instead of adding terms to the objective function, it becomes easier to adjust the parameters. These advantages are also addressed, for example, in [27]–[31]. The third constraint is a box constraint which represents the dynamic range of $\mathbf{u}$.

D. Optimization

We develop an efficient solver for Prob. (17) based on P-PDS [22]. Using the indicator functions, we rewrite Prob. (17) into an equivalent form:

$$\begin{aligned} \min_{\mathbf{u},\mathbf{s}} &\|\mathbf{D}_{\text{vh}} \mathbf{D}_\lambda \mathbf{u}\|_1 + \omega_1 \|\mathbf{D}_{\mathcal{G}_{\text{sp}}} \mathbf{u}\|_1 + \omega_2 \|\mathbf{D}_{\mathcal{G}_\lambda} \mathbf{u}\|_1 \\ &+ \iota_{B_{2,\varepsilon}^{\mathbf{v}}}(\mathbf{u} + \mathbf{s}) + \iota_{B_{1,\alpha}}(\mathbf{s}) + \iota_{R_{\underline{\mu},\bar{\mu}}}(\mathbf{u}). \end{aligned} \tag{21}$$

Prob. (21) can be solved by P-PDS [22] given by Alg. 1. The proximity operator of $\|\cdot\|_1$ is calculated by

$$[\text{prox}_{\gamma\|\cdot\|_1}(\mathbf{x})]_i = \text{sgn}(x_i) \max\{0, |x_i| - \gamma\}, \tag{22}$$

which is equivalent to soft thresholding. The projections onto $B_{2,\varepsilon}^{\mathbf{v}}$ and $R_{\underline{\mu},\bar{\mu}}$ are calculated by

$$P_{B_{2,\varepsilon}^{\mathbf{v}}}(\mathbf{x}) = \begin{cases} \mathbf{x}, & \text{if } \mathbf{x} \in B_{2,\varepsilon}^{\mathbf{v}}, \\ \mathbf{v} + \frac{\varepsilon(\mathbf{x}-\mathbf{v})}{\|\mathbf{x}-\mathbf{v}\|_2}, & \text{otherwise}, \end{cases} \tag{23}$$

$$[P_{R_{\underline{\mu},\bar{\mu}}}(\mathbf{x})]_i = \begin{cases} \underline{\mu}, & \text{if } x_i < \underline{\mu}, \\ \bar{\mu}, & \text{if } x_i > \bar{\mu}, \\ x_i, & \text{otherwise}. \end{cases} \tag{24}$$

The projection onto $B_{1,\alpha}$ can be efficiently computed by a fast ℓ_1-ball projection algorithm [32]. According to [23], we set the stepsize parameters as $\gamma_{\mathbf{u}} = \frac{1}{53}$, $\gamma_{\mathbf{s}} = 1$, $\gamma_{\mathbf{y}_1} = \gamma_{\mathbf{y}_2} = \gamma_{\mathbf{y}_3} = \gamma_{\mathbf{y}_4} = \gamma_{\mathbf{y}_5} = \frac{1}{2}$.

REFERENCES

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral remote sensing: principles and applications*. CRC press, 2007.

[2] H. Grahn and P. Geladi, *Techniques and applications of hyperspectral image analysis*. John Wiley & Sons, 2007.

[3] P. S. Thenkabail, J. G. Lyon, and A. Huete, *Hyperspectral remote sensing of vegetation*. CRC press, 2016.

[4] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, p. 2659, 2020.

[5] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, 2014.

[6] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, 2022.

[7] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, 2017.

[8] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[9] W. Ma, J. M. Bioucas-Dias, T. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, 2014.

[10] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 61–85, 2015.

[11] 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, 2018.

[12] H. Shen, M. Jiang, J. Li, C. Zhou, Q. Yuan, and L. Zhang, "Coupling model- and data-driven methods for remote sensing image restoration and fusion: Improving physical interpretability," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 231–249, 2022.

[13] K. Naganuma and S. Ono, "A general destriping framework for remote sensing images using flatness constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[14] 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.

[15] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, 2018.

[16] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, 2018.

[17] T. Ince, "Hyperspectral image denoising using group low-rank and spatial-spectral total variation," *IEEE Access*, vol. 7, pp. 52 095–52 109, 2019.

[18] S. Takeyama, S. Ono, and I. Kumazawa, "A constrained convex optimization approach to hyperspectral image restoration with hybrid spatio-spectral regularization," *Remote Sens.*, vol. 12, no. 21, 2020.

[19] M. Wang, Q. Wang, J. Chanussot, and D. Hong, "l_0-l_1 hybrid total variation regularization and its applications on hyperspectral image mixed noise removal and compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7695–7710, 2021.

[20] S. Takemoto, K. Naganuma, and S. Ono, "Graph spatio-spectral total variation model for hyperspectral image denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[21] S. Takemoto and S. Ono, "Enhancing spatio-spectral regularization by structure tensor modeling for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2023, pp. 1–5.

[22] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2011, pp. 1762–1769.

[23] K. Naganuma and S. Ono, "Variable-wise diagonal preconditioning for primal-dual splitting: Design and applications," *IEEE Trans. Signal Process.*, vol. 71, pp. 3281–3295, 2023.

[24] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[25] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[26] P. L. Combettes and N. N. Reyes, "Moreau's decomposition in banach spaces," *Math. Program.*, vol. 139, no. 1, pp. 103–114, 2013.

[27] M. V. Afonso, J. M. Bioucas-Dias, and M. 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.

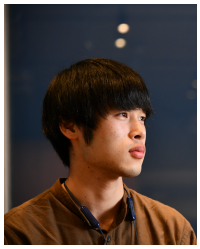
[28] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal Image Video Process.*, vol. 9, no. 8, pp. 1737–1749, 2015.

[29] 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.

[30] S. Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1108–1112, 2017.

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

[32] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Math. Program.*, vol. 158, no. 1, pp. 575–585, 2016.



Shingo Takemoto (S'22) received a B.E. degree and M.E. degree in Information and Computer Sciences in 2021 from Sophia University and in 2023 from Institute of Science Tokyo (formerly Tokyo Institute of Technology), respectively. He is currently pursuing a Ph.D. degree at the Department of Computer Science in the Institute of Science Tokyo. Since April 2024, he has been a Research Fellow (DC2) of Japan Society for the Promotion of Science (JSPS). His current research interests are in signal and image processing and optimization theory. Mr. Takemoto received the Student Award from IEEE SPS Tokyo Joint Chapter in 2022.

Shunsuke Ono (S'11–M'15–SM'23) received a B.E. degree in Computer Science in 2010 and M.E. and Ph.D. degrees in Communications and Computer Engineering in 2012 and 2014 from the Tokyo Institute of Technology, respectively. From 2012 to 2014, he was a Research Fellow (DC1) of the Japan Society for the Promotion of Science (JSPS). He was an Assistant, then an Associate Professor with Tokyo Institute of Technology (TokyoTech), Tokyo, Japan, from 2014 to 2024. From 2016 to 2020, he was a Researcher of Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency (JST), Tokyo, Japan. Currently, he is an Associate Professor with Institute of Science Tokyo (Science Tokyo), Tokyo, Japan. His research interests include signal processing, image analysis, optimization, remote sensing, and measurement informatics. He has served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS (2019–2024). Dr. Ono was a recipient of the Young Researchers' Award and the Excellent Paper Award from the IEICE in 2013 and 2014, respectively, the Outstanding Student Journal Paper Award and the Young Author Best Paper Award from the IEEE SPS Japan Chapter in 2014 and 2020, respectively, and the Best Paper Award in APSIPA ASC 2024. He also received the Funai Research Award in 2017, the Ando Incentive Prize in 2021, the MEXT Young Scientists' Award in 2022, and the IEEE SPS Outstanding Editorial Board Member Award in 2023.