Superpixel-Oriented Thick Cloud Removal Method for Multitemporal Remote Sensing Images

Qin Jiang, Xi-Le Zhao, Jie Lin, *Graduate Student Member, IEEE*, Jing-Hua Yang, Jiangtao Peng, *Senior Member, IEEE*, and Tai-Xiang Jiang

***Abstract*— Since the information across all bands of the cloud-contaminated region is missing, thick cloud removal for remote sensing images (RSIs) is still a challenging problem. Recently, the availability of rich spatial–spectral–temporal information for multitemporal RSIs provides the possibility for addressing the thick cloud removal problem. However, existing methods explore the holistic redundancy of multitemporal RSIs and neglect the important semantic clue of multitemporal images. In this letter, we propose a superpixel-oriented thick cloud removal (STORM) model for multitemporal images, where the multitemporal superpixel as the generic unit allows us to exploit redundancy with semantic clue in a low-rank optimization problem. To harness the resultant irregular fourth-order tensor (i.e., multitemporal superpixels) in the optimization problem, we cleverly introduce the weighted tensor to transform the irregular tensor into the regular tensor, which naturally leads to a standard low-rank tensor optimization problem. To tackle the tensor optimization problem, we develop a proximal alternating minimization (PAM)-based algorithm. Extensive simulated and real experiments on multitemporal RSIs acquired by Sentinel-2 and Landsat-8 satellites demonstrate the superior performance of the proposed method over the comparison methods.**

***Index Terms*— Proximal alternating minimization (PAM), semantic clue, superpixel, tensor ring (TR) decomposition, thick cloud removal.**

NOMENCLATURE

$x, \mathbf{x}, \mathbf{X}, \mathcal{X}$	Scalar, vector, matrix, and tensor.
$\mathcal{X}(i_1, i_2, i_3, i_4)$	The (i_1, i_2, i_3, i_4)th **element** of a 4-D tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$.
$\mathcal{X}(:, :, i_3, i_4)$	**Slice** of a 4-D tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$, defined by fixing all but two indices.
$\|\mathcal{X}\|_0$	l_0**-norm** of a tensor $\mathcal{X}$, number of nonzero elements of tensor $\mathcal{X}$.
$\|\mathcal{X}\|_F$	**Frobenius** norm of a tensor $\mathcal{X}$ and $\|\mathcal{X}\|_F = (\sum_{i_1,i_2,i_3,i_4} \|\mathcal{X}(i_1, i_2, i_3, i_4)\|^2)^{1/2}$.
$\mathbf{X}_{(i)}$	The first type of **unfolding operator** along the ith dimension, which is represented as $\mathbf{X}_{(i)} \in \mathbb{R}^{I_i \times I_1, \ldots, I_{i-1} I_{i+1} \ldots I_4}$.
$\mathbf{X}_{\langle i \rangle}$	Second type of **unfolding operator** along the ith dimension, which is represented as $\mathbf{X}_{\langle i \rangle} \in \mathbb{R}^{I_i \times I_{i+1}, \ldots, I_4 I_1 \ldots I_{i-1}}$.

Manuscript received 24 November 2023; accepted 12 December 2023. Date of publication 18 December 2023; date of current version 28 December 2023. This work was supported in part by NSFC under Grant 12371456, Grant 62131005, and Grant 12171072; in part by the Sichuan Science and Technology Program under Grant 23ZYZYTS0042; in part by the National Key Research and Development Program of China under Grant 2020YFA0714001; in part by the Natural Science Foundation of Anhui Province under Grant 1908085QA08; and in part by the Natural Science Research Projects of Anhui Province under Grant KJ2020A0535. *(Corresponding author: Xi-Le Zhao.)*

Qin Jiang, Xi-Le Zhao, and Jie Lin are with the School of Mathematical Sciences and the Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (e-mail: xlzhao122003@163.com).

Jing-Hua Yang is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 611756, China.

Jiangtao Peng is with the Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan, Hubei 430062, China.

Tai-Xiang Jiang is with the School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan 610074, China.

Digital Object Identifier 10.1109/LGRS.2023.3344163

I. INTRODUCTION

REMOTE sensing images (RSIs) acquired from spaceborne satellites often suffer from cloud contamination, resulting in missing ground information [1]. Therefore, the image reconstruction of missing information poses a challenging task due to its crucial role in enhancing the quality of images for subsequent applications [2], [3], [4], [5].

In the past, due to technical limitations, the observed RSIs are usually single temporal. For small-scale regions missing, inpainting methods [6], [7] utilize neighboring pixels to interpolate and fill in missing regions. For large-scale regions missing, partial differential equations [8] are introduced to reconstruct the missing data regions. Besides, exemplar-based texture synthesis methods [9], [10] generate expansive image regions from sample textures. Nevertheless, the above methods are often significantly constrained by similar spatial contextual structures due to the complete loss of information in regions affected by cloud contamination.

Since satellites periodically capture RSIs, more multitemporal images can be obtained with the development of technology. Multitemporal RSIs contain complement information that contributes to missing region reconstruction [11]. Multitemporal-based image reconstruction methods mainly include pixelwise methods and holistic-based methods. Pixelwise methods mainly search pixels with similar characteristics to reconstruct the cloud-contaminated regions. Zeng et al. [12] considered multitemporal images as referable information to obtain the locally similar pixels and built a regression model. Furthermore, Chen et al. [13] employed similar pixels exhibiting both local and nonlocal similarity to predict cloud-contaminated target pixels by applying spatially and temporally weighted regression. Benabdelkader and Melgani [14] introduced a pixelwise approach that effectively captures the spatial and spectral correlations within the image to enhance the contextual reconstruction process. Due to the independent selection of reference pixels for the missing pixels without considering the neighboring pixels, the pixelwise reconstruction strategy causes visual edge errors [15], [16]. Holistic-based methods utilize the low rankness of multitemporal RSIs to exploit holistic redundancy to reconstruct all missing pixels simultaneously, which alleviates the visual

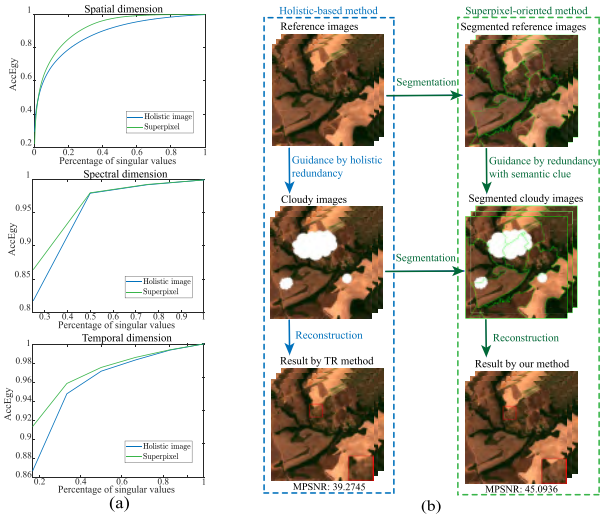

Fig. 1. Comparison between the proposed superpixel-oriented method and holistic-based method. (a) AccEgy with the corresponding percentage of the average of singular values of all superpixels and the percentage of singular values of holistic image along the spatial, spectral, and temporal dimensions. (b) Illustration of the proposed superpixel-oriented method and holistic-based method for cloud removal.

edge errors to a certain extent. Chen et al. [17] reshaped the multitemporal images to a low-rank matrix and conducted subsequent low-rank reconstruction. Considering the smooth property of the RSI, He et al. [18] proposed a low-rank tensor completion method combining tensor ring (TR) decomposition [19] with total variation (TV). Furthermore, Wang et al. [20] flexibly exploited different low rankness of TR factors by designing proper weights and combined TV from three directions to recover the missing information. Typical deep learning-based methods [21], [22] remove thick clouds using neural networks with powerful representation capability and have achieved satisfactory results. Zhang et al. [23] combined the handcrafted prior with the deep prior, which utilizes the low rankness of multitemporal images and leverages deep spatiotemporal feature expression ability by the 3-D convolutional neural network. However, the holistic-based methods only consider the holistic redundancy, which ignores the semantic clue of multitemporal images.

Motivation: To capture the semantic clue of multitemporal RSIs, we utilize the multitemporal superpixels as the generic unit, which consists of pixels with high semantic similarity and possesses stronger low rankness compared to the holistic image (i.e., superpixel achieves the same accumulation energy ratio (AccEgy) with fewer singular values, see Fig. 1). Since the irregular shape of superpixels limits the subsequent modeling, we introduce weighted tensors to represent the superpixels regularly, which allows us to incorporate the multiscale semantic clue coarse-to-fine to further boost the reconstruction performance. Moreover, we leverage TR decomposition to describe the spatial–spectral–temporal information within the regular multitemporal superpixels and further construct a low-rank optimization model for thick cloud removal.

This letter makes two main contributions.

1) We propose a superpixel-oriented thick cloud removal (STORM) model for multitemporal RSIs, which can exploit redundancy with semantic clue within the multitemporal superpixels by introducing weighted tensors to represent the irregular 4-D superpixels regularly, leading to finer local details preservation compared to holistic-based methods.
2) We develop a proximal alternating minimization (PAM)-based algorithm to solve the proposed thick cloud removal model. Extensive experiments on multitemporal RSIs acquired by Sentinel-2 and Landsat-8 satellites demonstrate that the proposed method achieves promising results, outperforming the comparison methods.

II. Notations

We summarize the notations used throughout this letter in the Nomenclature.

III. Method

A. Problem Formulation

Assume that the observed image $\mathcal{Y} \in \mathbb{R}^{m\times n\times b\times t}$ consists of the cloud-free image $\mathcal{X} \in \mathbb{R}^{m\times n\times b\times t}$ and sparse cloud component $\mathcal{S} \in \mathbb{R}^{m\times n\times b\times t}$, and the model can be written as

$$\mathcal{Y} = \mathcal{X} + \mathcal{S} \tag{1}$$

where $m \times n$, b, and t represent the size of spatial dimensions, spectral dimension, and temporal dimension, respectively.

B. Multitemporal Superpixels

We suggest using multitemporal superpixels as the generic unit, which is defined as regions in the image at different times that contains pixels with similar color, texture, and semantic information. First, we tackle the average image

$$Y_{\text{ave}} = \frac{1}{bt_r}\sum_{i_3=1}^{b}\sum_{i_4=1}^{t_r}\bar{\mathcal{Y}}(:,:,i_3,i_4)$$

from the reference images $\bar{\mathcal{Y}} \in \mathbb{R}^{m\times n\times b\times t_r}$ as the segmentation target, and the reference images are cloud-free images taken from t_r time nodes in the same scene, which contain the complete structure and local details similar to the cloud-contaminated images. Then, we employ the entropy rate superpixel segmentation method [24] to segment the image $Y_{\text{ave}} \in \mathbb{R}^{m\times n}$ for D times and obtain label maps containing K_i $(i = 1, \ldots, D)$ superpixels. According to the label maps, we generate multiscale superpixels on the observed image coarse-to-fine to obtain multiscale semantic clue. Since superpixels are irregular and hard to utilize directly, we convert them into regular tensors by introducing weighted tensor $\mathcal{W}_{ij}$ and denote the ijth 4-D multitemporal superpixels as $\mathcal{W}_{ij} \odot R_{ij}\mathcal{X}$, where $\odot$ denotes the pointwise product. In particular, R_{ij} is an operator that finds out the envelope cube of the multitemporal superpixels and $\mathcal{W}_{ij}$ is a binary tensor that denotes the superpixel region pixels by 1 and others by 0, respectively. In this way, we can flexibly integrate the semantic clue of multitemporal superpixels, and how to further characterize the relationship among the multitemporal superpixels for multitemporal RSI reconstruction remains a challenge.

C. TR Decomposition

To fully utilize the redundancy of multitemporal superpixels, we leverage TR decomposition, which can describe the spatial–spectral–temporal information within the multitemporal superpixels. For multitemporal superpixels $\mathcal{W}_{ij} \odot R_{ij}\mathcal{X} \in \mathbb{R}^{m_{ij}\times n_{ij}\times b\times t}$, we decompose it into the following 3-D factor tensors, i.e., $\mathcal{G}_{ij}^{(1)} \in \mathbb{R}^{r_1\times m_{ij}\times r_2}$, $\mathcal{G}_{ij}^{(2)} \in \mathbb{R}^{r_2\times n_{ij}\times r_3}$,

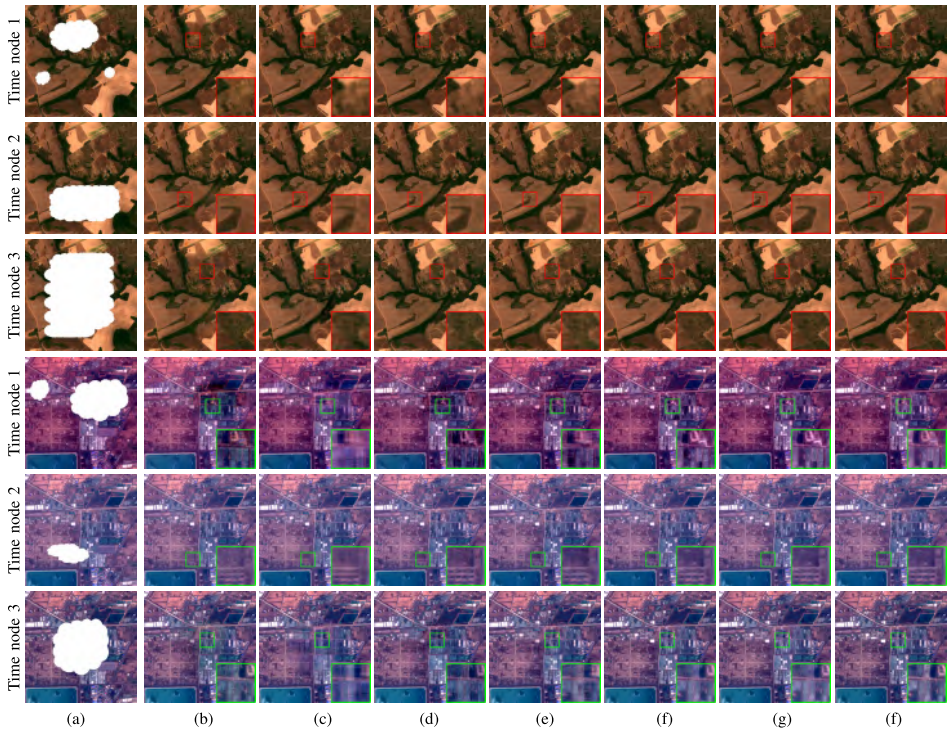

Fig. 2. Comparison of cloud removal results by all methods at different time nodes on the Brazil dataset and Dongying dataset. (a) Degraded image. (b) TNN. (c) TR. (d) TVTR. (e) TRLRF. (f) STORM. (g) STORM+. (f) Ground truth.

$\mathcal{G}_{ij}^{(3)} \in \mathbb{R}^{r_3 \times b \times r_4}$, and $\mathcal{G}_{ij}^{(4)} \in \mathbb{R}^{r_4 \times t \times r_1}$. Here, $\mathbf{r} = (r_1, r_2, r_3, r_4)$ denotes the TR rank. We represent the elementwise form as

$$[\mathcal{W}_{ij} \odot R_{ij}\mathcal{X}](i_1, i_2, i_3, i_4) = \sum_{j_1,\ldots,j_4=1}^{r_1,\ldots,r_4} \prod_{d=1}^{4} \mathbf{G}_{ij}^{(d)}(j_d, i_d, j_{d+1}).$$

In particular, $j_5 = j_1$. The TR decomposition of the superpixels is simply rewritten as $\mathcal{W}_{ij} \odot R_{ij}\mathcal{X} = \Phi(\mathcal{G}_{ij})$, where Φ is the operator to obtain approximated tensor through transforming TR factors and $\mathcal{G}_{ij} = \{\mathcal{G}_{ij}^{(1)}, \mathcal{G}_{ij}^{(2)}, \mathcal{G}_{ij}^{(3)}, \mathcal{G}_{ij}^{(4)}\}$.

D. Proposed Model

Equipping with the above preliminary, we use l_0-norm to regularize the sparsity of cloud components and propose an STORM model formulated as

$$\begin{aligned} \min_{\mathcal{X},\mathcal{S},\mathcal{G}_{ij}} \ & \|\mathcal{S}\|_0 \\ \text{s.t. } & \mathcal{W}_{ij} \odot R_{ij}\mathcal{X} = \Phi(\mathcal{G}_{ij}), \quad i = 1, \ldots, D \\ & \qquad\qquad\qquad\qquad\quad j = 1, \ldots, K_i \\ & \mathcal{Y} = \mathcal{X} + \mathcal{S}, \quad \mathcal{Y}_\Omega = \mathcal{X}_\Omega \end{aligned} \tag{2}$$

where D denotes the total number of the segmentation, K_i denotes the total number of the superpixels in the ith segmentation, and Ω denotes the index set of cloud-free regions.

We reformulate problem (2) as

$$\min_{\mathcal{X},\mathcal{S},\mathcal{G}_{ij}} \|\mathcal{S}\|_0 + \frac{\rho}{2}\|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + \iota_{\mathbb{Q}}(\mathcal{X}) + \sum_{i=1}^{D}\sum_{j=1}^{K_i} \frac{\beta}{2}\left\|\mathcal{W}_{ij} \odot R_{ij}\mathcal{X} - \Phi(\mathcal{G}_{ij})\right\|_F^2 \tag{3}$$

where

$$\iota_{\mathbb{Q}}(\mathcal{X}) := \begin{cases} 0, & \text{if } \mathcal{X} \in \mathbb{Q} \\ \infty, & \text{otherwise} \end{cases}$$

with $\mathbb{Q} := \{\mathcal{X} : \mathcal{X}_\Omega = \mathcal{Y}_\Omega\}$ and ρ and β being regularization parameters.

As model (3) is hard to directly optimize, the PAM algorithm is employed to optimize it by alternately updating

$$\begin{cases} \mathcal{G}_{ij}^{(d),l+1} = \underset{\mathcal{G}_{ij}^{(d)}}{\operatorname{argmin}} f\left(\mathcal{G}_{ij}^{(d)}, \mathcal{X}^l, \mathcal{S}^l\right) + \frac{\alpha}{2}\left\|\mathcal{G}_{ij}^{(d)} - \mathcal{G}_{ij}^{(d),l}\right\|_F^2 \\ \mathcal{X}^{l+1} = \underset{\mathcal{X}}{\operatorname{argmin}} f\left(\mathcal{G}^{l+1}, \mathcal{X}, \mathcal{S}^l\right) + \frac{\alpha}{2}\left\|\mathcal{X} - \mathcal{X}^l\right\|_F^2 \\ \mathcal{S}^{l+1} = \underset{\mathcal{S}}{\operatorname{argmin}} f\left(\mathcal{G}^{l+1}, \mathcal{X}^{l+1}, \mathcal{S}\right) + \frac{\alpha}{2}\left\|\mathcal{S} - \mathcal{S}^l\right\|_F^2 \end{cases}$$

where $f(\mathcal{G}, \mathcal{X}, \mathcal{S})$ is the objective function in (3), $\mathcal{G}_{ij}^{(d)}$ denotes the dth TR factor of the ijth multitemporal superpixel $\mathcal{W}_{ij} \odot R_{ij}\mathcal{X} \in \mathbb{R}^{m_{ij} \times n_{ij} \times b \times t}$, l denotes the iteration index, and $\alpha > 0$ is a proximal parameter.

1) Updating $\mathcal{G}_{ij}^{(d)}$: The $\mathcal{G}_{ij}^{(d)}$ subproblem is

$$\begin{aligned} \mathcal{G}_{ij}^{(d),l+1} &= \arg\min_{\mathcal{G}_{ij}^{(d)}} \frac{\alpha}{2}\left\|\mathcal{G}_{ij}^{(d)} - \mathcal{G}_{ij}^{(d),l}\right\|_F^2 \\ &\quad + \frac{\beta}{2}\left\|\mathcal{W}_{ij} \odot R_{ij}\mathcal{X}^l - \Phi\left(\mathcal{G}_{ij}^{(1:d-1),l+1}, \mathcal{G}_{ij}^{(d)}, \mathcal{G}_{ij}^{(d+1:4),l}\right)\right\|_F^2. \end{aligned} \tag{4}$$

The solution to problem (4) is given by

$$\mathcal{G}_{ij}^{(d),l+1} = \text{fold}_2(\mathbf{H}) \tag{5}$$

$$\mathbf{H} = \left(\beta\left(\mathcal{W}_{ij} \odot R_{ij}\mathcal{X}^l\right)_{\langle n \rangle}\mathbf{G}_{ij\langle 2\rangle}^{(\neq d)} + \alpha\mathbf{G}_{ij(2)}^{(d)}\right) \times \left(\beta\mathbf{G}_{ij\langle 2\rangle}^{(\neq d),T}\mathbf{G}_{ij\langle 2\rangle}^{(\neq d)} + \alpha\mathbf{I}\right)^{-1} \tag{6}$$

where $\mathbf{I}$ is an identity matrix and fold_2 [18] denotes the inverse operator of the first mode-2 unfolding.

2) Updating $\mathcal{X}$: The $\mathcal{X}$ subproblem is

$$\mathcal{X}^{l+1} = \arg\min_{\mathcal{X}} \frac{\rho}{2}\|\mathcal{Y} - \mathcal{X} - \mathcal{S}^l\|_F^2 + \frac{\alpha}{2}\left\|\mathcal{X} - \mathcal{X}^l\right\|_F^2$$

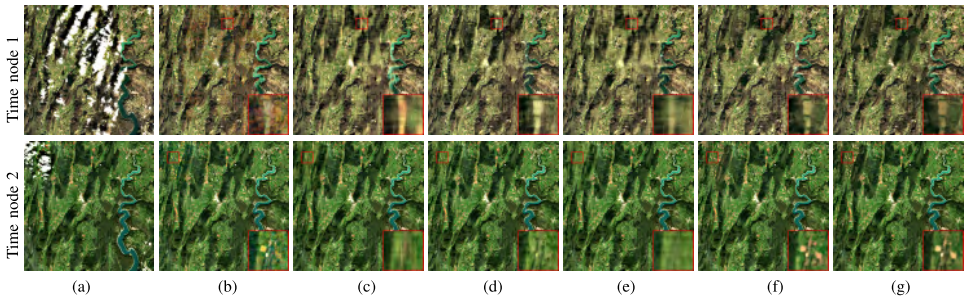

Fig. 3. Comparison of cloud removal results by all methods at different time nodes on the Bourgogne dataset. (a) Degraded image. (b) TNN. (c) TR. (d) TVTR. (e) TRLRF. (f) STORM. (g) STORM+.

TABLE I
QUANTITATIVE RESULTS BY ALL METHODS ON THE BRAZIL DATASET AND DONGYING DATASET

Brazil							
Time node	Index	TNN	TR	TVTR	TRLRF	STORM	STORM+
Time node 1	MPSNR	32.745	42.130	45.070	43.211	50.118	**51.211**
	MSSIM	0.9847	0.9795	0.9862	0.9831	0.9957	**0.9971**
	SAM	0.0040	0.0048	0.0048	0.0042	0.0023	**0.0021**
Time node 2	MPSNR	33.865	40.491	42.487	42.069	45.476	**46.898**
	MSSIM	0.9808	0.9704	0.9791	0.9760	0.9892	**0.9931**
	SAM	0.0086	0.0072	0.0073	0.0064	0.0041	**0.0038**
Time node 3	MPSNR	30.383	35.204	36.529	36.738	39.687	**42.110**
	MSSIM	0.9551	0.9182	0.9382	0.9357	0.9732	**0.9831**
	SAM	0.0198	0.0204	0.0210	0.0187	0.0133	**0.0102**
Dongying							
Time node	Index	TNN	TR	TVTR	TRLRF	STORM	STORM+
Time node 1	MPSNR	33.543	36.530	35.054	37.451	39.441	**40.511**
	MSSIM	0.9694	0.9574	0.9519	0.9643	0.9760	**0.9806**
	SAM	0.0075	0.0049	0.0054	0.0042	0.0033	**0.0030**
Time node 2	MPSNR	43.280	40.963	42.991	42.578	47.689	**48.294**
	MSSIM	0.9910	0.9816	0.9884	0.9867	0.9956	**0.9962**
	SAM	0.0028	0.0039	0.0032	0.0030	**0.0018**	0.0019
Time node 3	MPSNR	33.170	31.217	31.681	32.673	35.213	**35.576**
	MSSIM	0.9625	0.9030	0.9342	0.9291	0.9576	**0.9641**
	SAM	**0.0147**	0.0172	0.0188	0.0225	0.0240	0.0184

$$+\sum_{i=1}^{D}\sum_{j=1}^{K_i}\frac{\beta}{2}\left\|\mathcal{W}_{ij}\odot R_{ij}\mathcal{X}-\Phi\left(\mathcal{G}_{ij}^{l+1}\right)\right\|_F^2+\iota_{\mathbb{Q}}(\mathcal{X}). \tag{7}$$

The problem can be solved by

$$\mathcal{X}^{l+1}=\begin{cases}\left(\widetilde{\mathcal{X}}^{l+1}\right)_{\bar{\Omega}}, & \mathcal{X}\in\bar{\Omega}\\ \mathcal{Y}_{\Omega}, & \mathcal{X}\in\Omega\end{cases} \tag{8}$$

where $\widetilde{\mathcal{X}}^{l+1}=[\rho(\mathcal{Y}-\mathcal{S}^l)+\beta\sum_{i=1}^{D}\sum_{j=1}^{K_i}\Phi(\mathcal{G}_{ij}^{l+1})+\alpha\mathcal{X}^l]\oslash[\rho\mathcal{J}+\beta\sum_{i=1}^{D}\sum_{j=1}^{K_i}\mathcal{W}_{ij}\odot R_{ij}^T R_{ij}+\alpha\mathcal{J}]$, $\oslash$ denotes the pointwise division, $\bar{\Omega}$ denotes the index set indicating cloud region positions, and $\mathcal{J}$ is an all-ones tensor whose elements are all 1.

3) Updating $\mathcal{S}$: The $\mathcal{S}$ subproblem is

$$\mathcal{S}^{l+1}=\arg\min_{\mathcal{S}}\|\mathcal{S}\|_0+\frac{\rho}{2}\|\mathcal{Y}-\mathcal{X}^{l+1}-\mathcal{S}\|_F^2+\tfrac{\alpha}{2}\left\|\mathcal{S}-\mathcal{S}^l\right\|_F^2. \tag{9}$$

The closed-form solution of problem (9) is

$$\mathcal{S}^{l+1}=\text{hard}_{\frac{1}{\rho+\alpha}}\left(\frac{\rho(\mathcal{Y}-\mathcal{X}^{l+1})+\alpha\mathcal{S}^l}{\rho+\alpha}\right) \tag{10}$$

where hard is hard-thresholding operator

$$\text{hard}_{\epsilon}(z)=\begin{cases}z-\epsilon, & \text{if } z>\epsilon\\ 0, & \text{if } z\le\epsilon.\end{cases} \tag{11}$$

The developed algorithm to solve the proposed STROM model is described in Algorithm 1.

Algorithm 1 PAM Algorithm for STORM

Input: The observation $\mathcal{Y}$ and corresponding index set Ω;
Initialization: $\mathcal{X}^0=\mathcal{S}^0=\mathcal{O}$;
while not converged **do**
  for each segmentation $i=1,2,\ldots,D$ **do**
    for multitemporal superpixels $j=1,2,\ldots,K_i$ **do**
      for each TR factor $d=1,2,3,4$ **do**
        Update $\mathcal{G}_{ij}^{(d),l+1}$ via Eq. (5);
      end for
    end for
  end for
  Update $\mathcal{X}^{l+1}$ via Eq. (8);
  Update $\mathcal{S}^{l+1}$ via Eq. (10);
end while
Output: The reconstructed image $\mathcal{X}$;

IV. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed method STORM, we select four methods for comparison, including TNN [25], TR [19], TVTR [18], and TRLRF [26], on simulated and real cases. We perform single superpixel segmentation ($K_1=10$) denoted by STORM and corresponding multisegmentation ($K_1=10, K_2=70$) denoted by STORM+. We employ the mean peak signal-to-noise ratio (MPSNR), mean structural similarity index (MSSIM), and spectral angle mapper (SAM) to evaluate the resemblance between the reconstructed image and the ground truth [27]. Higher MPSNR and MSSIM values and smaller SAM values indicate a better reconstruction effect.

A. Results on the Simulated Experiment

We select the Brazil[1] dataset and Dongying[2] dataset as the simulated datasets to verify the superiority of the proposed methods (STORM and STORM+). Each time node in the datasets taken by Sentinel-2 comprises four spectral bands (bands 2–4 and 8) with a spatial resolution of 10 m. The sizes of datasets are both $400\times400\times4\times6$. We design different masks to simulate clouds of diverse shapes and sizes at the first three time nodes, and the last three time nodes are used as reference images.

[1]https://www.theia-land.fr/en/data-and-services-for-the-land/
[2]https://earthexplorer.usgs.gov

The quantitative results are presented in Table I, with the highest MPSNR, MSSIM values, and smallest SAM highlighted in bold. As shown in Table I, our method performs almost the best on all metrics on all datasets, whose MPSNR values achieve an improvement of about 2–5 dB compared to the second-best method. Due to the utilization of semantic clue of superpixels, the best performance is obtained from STORM. We show the visual comparison in Fig. 2, revealing that STORM can reconstruct the details and textures of the cloud-contaminated regions better compared to the TR, TVTR, and TRLRF. From the results of TNN on the Dongying dataset, we can observe a distinct color difference when comparing the reconstructed result and the ground truth. In conclusion, the proposed STORM+ achieves the closest results to the ground truth compared to all comparison methods. Compared with the STORM method, the STORM+ method can provide more comprehensive local details indicating the effectiveness of the multiscale strategy in our method.

B. Results on the Real Experiment

We select the Bourgogne[3] dataset acquired by Landsat-8 to test the effectiveness of the proposed methods (STORM and STORM+) in real-world scenarios. In this section, the subimage of size $600 \times 600 \times 7 \times 4$ is utilized, and seven spectral bands (bands 1–7) with a spatial resolution of 30 m are encompassed in each time node. The relevant masks of clouds are available online.[3] Fig. 3 presents the visual results of all methods, with the red box magnified four times for more detailed observation. We can observe that STORM and STORM+ reconstruct the details and textures of the cloud-contaminated regions best, which verifies the effectiveness of the proposed STORM and STORM+. This result is attributed to the introduction of semantic clue, which promotes local details preservation in complex scenes.

V. Conclusion

In this letter, we have proposed an STORM model for thick cloud removal. First, multitemporal superpixels are employed as the generic unit, which exploits redundancy with semantic clue. Besides, the weighted tensors are suggested to handle the challenge of superpixels' irregular shape and further incorporate the multiscale semantic clue in regular 4-D tensors. Then, we utilize TR decomposition to describe spatial–spectral–temporal information within multitemporal superpixels and establish a low-rank reconstruction model. Finally, a PAM-based algorithm is developed to solve the proposed model. Extensive experiments on multitemporal RSIs taken by Sentinel-2 and Landsat-8 satellites have demonstrated that the STORM method is superior to the compared methods.

References

[1] D. Hong et al., "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[2] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, and Y. Xiao, "Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–21, 2023, doi: 10.1109/TNNLS.2023.3278866.

[3] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[4] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.

[5] Y. Chen, X. Gui, J. Zeng, X.-L. Zhao, and W. He, "Combining low-rank and deep plug-and-play priors for snapshot compressive imaging," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–13, 2023, doi: 10.1109/TNNLS.2023.3294262.

[6] A. C. Siravenha, D. Sousa, A. Bispo, and E. Pelaes, "The use of high-pass filters and the inpainting method to clouds removal and their impact on satellite images classification," in *Proc. ICIAP*, 2011, pp. 333–342.

[7] J. Chen, X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin, "A simple and effective method for filling gaps in Landsat ETM+ SLC-off images," *Remote Sens. Environ.*, vol. 115, no. 4, pp. 1053–1064, Apr. 2011.

[8] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1200–1211, Aug. 2001.

[9] J. Jia and C.-K. Tang, "Image repairing: Robust image synthesis by adaptive ND tensor voting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, p. 643.

[10] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

[11] C. Long, X. Li, Y. Jing, and H. Shen, "Bishift networks for thick cloud removal with multitemporal remote sensing images," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–17, Feb. 2023.

[12] C. Zeng, H. Shen, and L. Zhang, "Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method," *Remote Sens. Environ.*, vol. 131, pp. 182–194, Apr. 2013.

[13] B. Chen, B. Huang, L. Chen, and B. Xu, "Spatially and temporally weighted regression: A novel method to produce continuous cloud-free Landsat imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 27–37, Jan. 2017.

[14] S. Benabdelkader and F. Melgani, "Contextual spatiospectral postreconstruction of cloud-contaminated images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 204–208, Apr. 2008.

[15] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.

[16] W. Wu, L. Ge, J. Luo, R. Huan, and Y. Yang, "A spectral–temporal patch-based missing area reconstruction for time-series images," *Remote Sens.*, vol. 10, no. 10, p. 1560, Sep. 2018.

[17] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 93–107, Nov. 2019.

[18] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8998–9009, Nov. 2019.

[19] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," 2016, *arXiv:1606.05535*.

[20] M. Wang, Q. Wang, J. Chanussot, and D. Hong, "Total variation regularized weighted tensor ring decomposition for missing data recovery in high-dimensional optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[21] S. Ji, P. Dai, M. Lu, and Y. Zhang, "Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 732–748, Jan. 2021.

[22] Y. Chen, Q. Weng, L. Tang, X. Zhang, M. Bilal, and Q. Li, "Thick clouds removing from multitemporal Landsat images using spatiotemporal neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4400214.

[23] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, "Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 161–173, Jul. 2021.

[24] Y. Tang, L. Zhao, and L. Ren, "Different versions of entropy rate superpixel segmentation for hyperspectral image," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2019, pp. 1050–1054.

[25] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2017.

[26] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, Jul. 2019, pp. 9151–9158.

[27] J. Lin, T.-Z. Huang, X.-L. Zhao, Y. Chen, Q. Zhang, and Q. Yuan, "Robust thick cloud removal for multitemporal remote sensing images using coupled tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5406916.

[3] https://www.theia-land.fr/en/data-and-services-for-the-land/