Robust Thick Cloud Removal for Multi-Temporal Remote Sensing Images Using Coupled Tensor Factorization

Jie Lin, Ting-Zhu Huang, Xi-Le Zhao, Yong Chen, Qiang Zhang, and Qiangqiang Yuan

Abstract—The existing non-blind cloud and cloud shadow (cloud/shadow) removal methods for remote sensing (RS) images are based on the assumption that cloud/shadow masks are accurately given. Since the masks are usually manually labeled or detected by cloud detection methods, whose accuracy cannot be well guaranteed, the cloud/shadow removal effect may be affected. In this paper, we suggest a robust thick cloud/shadow removal (RTCR) method that meets the problem with an inaccurate mask. To faithfully reconstruct the multi-temporal information, a coupled tensor factorization is used to explore the relationship between the abundances of the multi-temporal images in the same scene. Moreover, an efficient algorithm is developed to solve the proposed model based on the augmented Lagrange multiplier method. The experimental results under accurate masks and inaccurate masks demonstrate its robustness and superiority for thick cloud/shadow removal.

Index Terms—Robust cloud/shadow removal, coupled tensor factorization, multi-temporal remote sensing image, augmented Lagrange multiplier.

I. INTRODUCTION

WITH the development of remote sensing (RS) technology, lots of temporal images with the high spatial and spectral resolution are captured by satellites. This greatly promotes the applications of RS images in environmental monitoring, urban planning, surface classification, etc [1–5]. However, RS images are inevitably contaminated by thick cloud/shadow and the authentic reflectivity information is lost, which severely limits its regular exploitation, such as unmixing, target detection, and classification [6, 7]. Therefore, designing a thick cloud/shadow removal method to reconstruct the cloud-contaminated region is an essential preprocessing task.

J. Lin, T.-Z. Huang, and X.-L. Zhao are with the School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China.

Y. Chen is with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, Jiangxi 330022, China.

Q. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China.

Q. Yuan is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China.
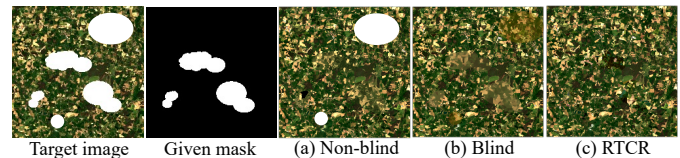

Fig. 1. Comparison of reconstructed results of three kinds of cloud removal methods. (a) Non-blind method [8]; (b) Blind method [9]; (c) The proposed RTCR method.

A. Related Works

In the past decades, plenty of methods have been proposed to address the issue of thick cloud/shadow removal. According to whether the cloud/shadow mask is required, these methods can be divided into two categories: non-blind methods and blind methods.

1) Non-blind methods: The non-blind methods need the given cloud/shadow mask to identify corrupted regions and then use cloud/shadow-free regions to reconstruct the underlying information. With the given mask, the thick cloud/shadow removal task can be transformed into the inpainting or completion task [10–12]. The most classic methods to tackle this issue are spatial-based methods, including interpolation-based methods [13–15], variation-based methods [16–18], and exemplar-based methods [19, 20]. These methods only use the relationship among the local or nonlocal regions in the spatial dimension and perform well in rebuilding small cloud/shadow-contaminated regions [21]. To utilize the high correlations among the spectral bands, spectral-based methods for missing information reconstruction were proposed. Shen *et al.* [22], Wang *et al.* [23], Rakwatin *et al.* [24], and Li *et al.* [25] reconstructed the missing region of Aqua Moderate-resolution Imaging Spectroradiometer band 6 based on the spectral relation with other bands. However, in the real scenario, all spectral bands of RS images are usually contaminated by large-scale thick clouds, which means that only using spatial or spectral information is not enough to reconstruct underlying information [26, 27]. Since the satellites revisit the same region with a regular revisiting cycle and the cloud/shadow-contaminated region of the acquired RS images cannot overlap all the time, the cloud/shadow-free region in the multi-temporal reference images is powerful complementary information to reconstruct cloud/shadow-contaminated the region [28, 29]. To fully exploit all available information, Benabdelkader *et al.* [30] and Melgani [31] reconstructed the contextual information of cloud-contaminated regions in multi-
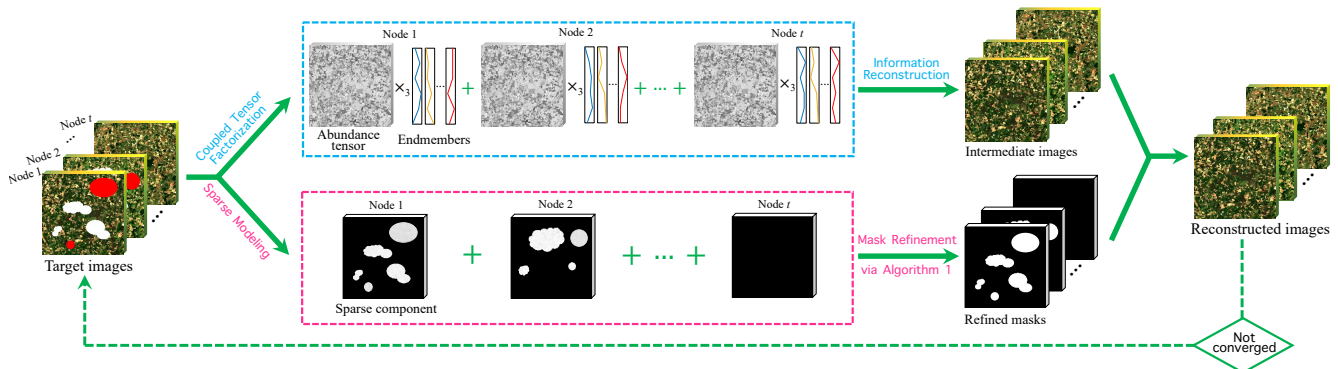
Fig. 2. Flowchart of the proposed robust thick cloud/shadow removal (RTCR) for multi-temporal RS images. (The red color regions in target images denote the undetected clouds.)

temporal images by employing spatial, spectral, and temporal relations. Based on the temporal correlation of multi-temporal images, Cheng *et al.* [32] built a pixel-offset based spatio-temporal Markov random fields global function to locate similar pixels and the missing pixel is filled using the known similar pixels; Wang *et al.* [8] proposed a temporally contiguous robust matrix completion model and developed an augmented Lagrangian method with inexact proximal gradient (ALM-IPG) to solve the model; Lin *et al.* [33] presented a patch-based information reconstruction algorithm; Shuai *et al.* [34] presented a spectral angle distance weighted reconstruction method to reconstruct the MODIS land surface temperature product. Using multi-temporal images as reference images, Zeng *et al.* [35] proposed a weighted linear regression (WLR) model to fill the missing pixels. To jointly exploit the spatial, spectral, and temporal information, Ng *et al.* [36] proposed an adaptive weighted tensor completion method to recover the missing data; Ji *et al.* [37] suggested a nonlocal low-rank tensor completion method, which makes full use of the correlations of all the three domains. The temporal-based methods show a more effective performance for large and thick cloud removal. Recently, deep learning (DL) has achieved great success in the field of remote sensing image processing and DL-based methods for cloud removal have been proposed. Zhang *et al.* [38] proposed a unified spatial-temporal-spectral framework based on a deep convolutional neural network (CNN) that employs a unified deep CNN combined with spatial-temporal-spectral supplementary information. Later, Zhang *et al.* [39] proposed a framework that combines the global-local spatio-temporal information in remote sensing imagery with the non-linear learning capability of deep neural network. All the above methods exhibit high effectiveness for cloud/shadow removal. However, these methods rely heavily on the mask. When the given cloud mask is not accurate, it is difficult to obtain the ideal cloud/shadow removal result.

2) Blind methods: The blind methods remove the cloud/shadow without the given cloud/shadow mask. To overcome the dependence on the mask, many blind methods with promising performance have been proposed. Without the given mask, Tseng *et al.* [40] proposed an automatic multi-temporal method, which contains multi-spectral image enhancement, cloud/shadow determination, and image mosaicking, to generate cloud-free mosaic images from multi-temporal SPOT images; Lin *et al.* [41] used the temporal correlation of multi-temporal images and suggested a method that clones infor-

mation from cloud-free patches to the corresponding detected cloud-contaminated patches; Meraner *et al.* [42] designed a deep residual neural network architecture for the cloud removal of multi-spectral Sentinel-2 imagery and utilized SAR-optical data fusion to exploit the synergistic properties of the two imaging systems to guide the image reconstruction. Li *et al.* [43] proposed a nonnegative matrix factorization and error correction method, which does not require cloud detection and the cloud-free information of the cloud-contaminated image can be maximally retained. Recently, Chen *et al.* [9] proposed a blind method named spatial-spectral total-variation regularized low-rank sparsity decomposition (TVLRSD) framework, which embodies an efficient threshold method to detect cloud. The method gets rid of the dependence on the mask and shows superior performance for the cloud removal of multi-temporal RS images. However, the threshold in the threshold method needs to be set manually and the effect of cloud removal is very sensitive to the threshold. An inappropriate threshold can lead to incomplete cloud removal or spectral distortion.

B. Motivation

First, the calibrated RS imagery products usually come with manually labeled cloud/shadow masks, while sometimes the given masks are inaccurate. Non-blind methods reconstruct the underlying information completely according to the given mask, which cannot reconstruct all information when the given mask with missed detection of clouds, as shown in Fig. 1 (a); the blind method reconstructs the underlying information without the mask, which will lead to color distortion since there is no true information as guidance, as shown in Fig. 1 (b). Thus, ***how to make a balance between the non-blind methods and the blind method to achieve the reasonable use of the masks that comes with RS imagery products?***

To tackle this issue, we introduce a sparse term into our method to characterize the cloud/shadow component and develop an adaptive threshold algorithm to refine the given inaccurate mask. The method can achieve a robust cloud/shadow removal with an inaccurate mask.

Second, the exiting temporal-based methods usually use the direct relationship between original multi-temporal images. ***Is there any latent relationships between the multi-temporal RS images, which can be exploited to finely reconstruct the multi-temporal information?***

As the distribution of surface material is constant over a short period and the same material shows different spectral
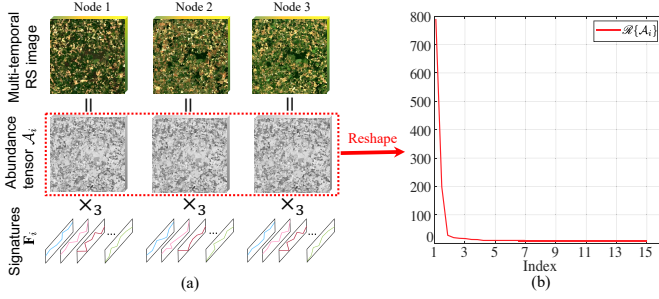
Fig. 3. Redundancy between abundance tensors $\mathcal{A}_i \in \mathbb{R}^{m \times n \times r} (i = 1, 2, 3)$. ($\mathscr{R}\{\mathcal{A}_i\} \in \mathbb{R}^{mn \times 3r}$ is the matrixing of all $\mathcal{A}_i$s.)

signatures at different time nodes, we observe that the multi-temporal images in the same scene share the same abundances over a short period. This observation motivates us to find that there is redundancy between the abundances of multi-temporal images, which can be described by rank. Inspired by subspace representation [44], we use a coupled tensor factorization to decompose the RS image at each time node into orthogonal basis and coefficient, which are more compact spectral signature set and corresponding abundances from the unmixing perspective [45, 46]. Fig. 3 shows the relationship between abundance tensors, i.e., the reshaped matrix of abundance tensors is low-rank.

C. Contribution

Based on the above-mentioned strategy and observation, a **R**obust **T**hick **C**loud/shadow **R**emoval (RTCR) method using coupled tensor factorization is suggested to address the cloud/shadow removal with an inaccurate mask (see Fig. 2 for the flowchart). The contributions are three-fold:

- We suggest a robust thick cloud/shadow removal method. Compared with non-blind methods, a sparse term is introduced to characterize the cloud/shadow component and an adaptive threshold algorithm is designed to refine the given mask, which makes the method robust to the given inaccurate mask.
- We propose a multi-temporal information reconstruction model. The model uses a coupled tensor factorization to explore a deeper relationship that the redundancy between the abundances of multi-temporal images. This provides a new perspective for cloud/shadow removal.
- We develop an efficient augmented Lagrange multiplier (ALM) method-based algorithm for solving the proposed model. The simulated and real experiments on three datasets acquired by the Sentinel-2 and Landsat-8 are implemented, which demonstrates its superiority and robustness and for thick cloud/shadow removal under accurate masks and inaccurate masks.

D. Organization

The remainder of this paper is organized as follows: Section II gives the proposed RTCR method. Section III reports the experimental results. Section IV concludes this paper.

II. THE PROPOSED METHOD

A. Preliminary

In this paper, we denote tensor by capitalized calligraphic letters, e.g., $\mathcal{A}$; matrices are represented as capitalized letters, e.g., $\mathbf{A}$; scalars are expressed with lowercase letters, e.g., x. An mth-order tensor is represented as $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots n_m}$ and its elements are denoted as $a_{i_1, i_2, \cdots, i_m}$. The unfolding matrix $\mathbf{A}_{(k)} = \text{unfold}_k(\mathcal{A}) \in \mathbb{R}^{n_k \times \prod_{i \neq k} n_i}$ is composed by taking the k-mode vectors of $\mathcal{A}$ as its columns. Conversely, the unfolding matrices along the k-mode can be transformed back to the tensor by $\mathcal{A} = \text{fold}_k(\mathbf{A}_{(k)}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots n_m}$. The nuclear norm of matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$, where $\sigma_i(\mathbf{A})$ is the ith singular value of $\mathbf{A}$. The l_1-norm and Frobenius norm of tensor $\mathcal{A}$ are defined as $\|\mathcal{A}\|_1 = \sum_{i_1, i_2, \cdots, i_m} |a_{i_1, i_2, \cdots, i_m}|$ and $\|\mathcal{A}\|_F = (\sum_{i_1, i_2, \cdots, i_m} |a_{i_1, i_2, \cdots, i_m}|^2)^{1/2}$, respectively. For $\{\mathcal{A}_j\}_{j=1}^{n_4}$ where $\mathcal{A}_j \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, operators $\mathscr{R}$ and $\mathscr{C}$ are defined as $\mathbf{A} = \mathscr{R}\{\mathcal{A}_j\} \in \mathbb{R}^{n_1 n_2 \times n_3 n_4}$ and $\mathcal{A} = \mathscr{C}\{\mathcal{A}_j\} \in \mathbb{R}^{n_1 \times n_2 \times n_3 n_4}$, respectively.

Definition (i-mode product [47]) For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots n_m}$ and a matrix $\mathbf{F} \in \mathbb{R}^{k \times n_i}$, the i-mode product of them is denoted as $\mathcal{X} = \mathcal{A} \times_i \mathbf{F} \in \mathbb{R}^{n_1 \times \cdots \times n_{i-1} \times k \times n_{i+1} \times \cdots \times n_m}$ and

$$x_{n_1, \cdots, n_{i-1}, k, n_{i+1}, \cdots, n_m} = \sum_{j_i=1}^{n_i} a_{i_1, i_2, \cdots, i_N} f_{k, j_i}.$$
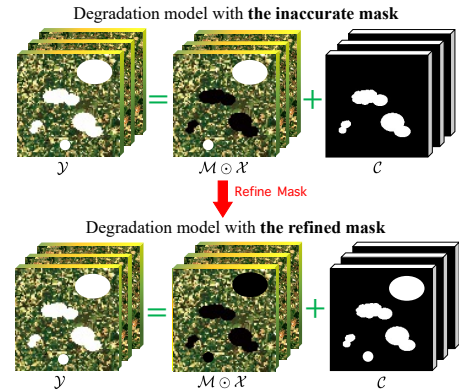
B. Problem Formulation

Fig. 4. Illustration of the degradation model with different masks.

For the thick cloud/shadow contaminated observations with t time nodes $\mathcal{Y} \in \mathbb{R}^{m \times n \times bt}$ (m and n are spatial resolutions, and b is the number of spectral bands), the degradation model can be expressed as

$$\mathcal{Y} = \mathcal{M} \odot \mathcal{X} + \mathcal{C},$$

where $\odot$ denotes the point-wise product, $\mathcal{M} \in \mathbb{R}^{m \times n \times bt}$ is a binary tensor with the zeros representing missing pixels, $\mathcal{X} \in \mathbb{R}^{m \times n \times bt}$ is the cloud-free images, and $\mathcal{C} \in \mathbb{R}^{m \times n \times bt}$ is cloud/shadow component. Fig. 4 illustrates the degradation model with the inaccurate mask and the refined mask.

C. Mask Refinement

To refine the given cloud/shadow mask $\mathcal{M}^0$, we use a threshold algorithm (Algorithm 1) to detect the undetected cloud/shadow from the error component $\mathcal{E} = \mathcal{Y} - \mathcal{X}$.

Algorithm 1 Adaptive Threshold Algorithm for Mask Refinement

Input: Error component $\mathcal{E}$, given mask $\mathcal{M}^0$, and corresponding cloud/shadow index set Ω^0.
1: Initialize: $\Omega = \Omega^0$, $\mathcal{M} = \mathcal{M}^0$.
2: **for** $i = 1 : t$ **do**
3:      **for** $p_1 = 1 : m$ **do**
4:          **for** $p_2 = 1 : n$ **do**
5:              Compute $a = \text{mean}\,[\mathcal{E}(p_1, p_2, (i\text{-}1)b\text{+}1{:}ib)]$;
6:              Compute
             $\tau = \min\{|\text{mean}[(\mathcal{E})_{\Omega^0}(p_1, p_2, (i\text{-}1)b\text{+}1{:}ib)]|\}$;
7:              $\Omega = \Omega \cup (p_1, p_2, (i\text{-}1)b\text{+}1{:}ib)$, if $|a| > \tau$;
8:          **end for**
9:      **end for**
10: **end for**
11: Let $(\mathcal{M})_\Omega = 0$;
Output: Refined mask $\mathcal{M}$.

Under the guidance of the given mask, the threshold value τ in the proposed algorithm can be automatically set according to the pixel values of the error component $\mathcal{E}$ corresponding to the given cloud/shadow region. The strategy may result in the over-detection of the cloud/shadow but the over-detected region can be faithfully reconstructed (see Section III-D1). Moreover, we embed the threshold algorithm in each iteration of Algorithm 2, which will help introduce true complementary information for subsequent information compensation, which circumvent the problem of the blind method.

D. Information Reconstruction

For the ith time node RS image $\mathcal{X}_i \in \mathbb{R}^{m \times n \times b}$, we decompose it as

$$\mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i,$$

where $\mathcal{A}_i \in \mathbb{R}^{m \times n \times r}$ is abundance tensor, $\mathbf{F}_i \in \mathbb{R}^{b \times r}$ is semi-orthogonal basis and contains r spectral signatures, and $\times_3$ is 3-mode product. There are two necessities for the orthogonality of basis. First, it allows us to use fewer distinguished spectral signatures to represent the image, leading to less computational complexity. Second, it guarantees the closed-from solution for updating $\mathbf{F}_i$ [48].

Based on our observation and the sparse prior of the cloud/shadow [9], the low-rankness and the sparsity of the rearrangement of abundance tensors $\mathbf{A} = \mathscr{R}\{\mathcal{A}_i\} \in \mathbb{R}^{mn \times rt}$ and the cloud/shadow component $\mathcal{C}$ can be characterized by rank function and l_0-norm, respectively. For observed multi-temporal images $\mathcal{Y} \in \mathbb{R}^{m \times n \times bt}$, the information reconstruction model is formulated as

$$\min_{\mathcal{X}, \mathcal{C}, \mathcal{A}_i, \mathbf{F}_i} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_0 + \alpha \text{Rank}(\mathbf{A}),$$
$$s.t. \quad \mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i, \ \mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}. \tag{1}$$

where i denotes the ith time node, $\mathcal{X} \in \mathbb{R}^{m \times n \times bt}$ is cloud-free image, $\mathcal{Y} \in \mathbb{R}^{m \times n \times bt}$ is target image, and $\mathbf{A} \in \mathbb{R}^{mn \times rt}$ is the rearrangement of all $\mathcal{A}_i$s, which is obtained by vectorizing each band of all $\mathcal{A}_i$s into a column and then rearranging them into a matrix. α and β are regularization parameters.

For a tractable optimization, we substitute the rank function and l_0-norm with nuclear norm and l_1-norm, respectively. Thus, we formulated (1) as

$$\min_{\mathcal{X}, \mathcal{C}, \mathcal{A}_i, \mathbf{F}_i} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_1 + \alpha\|\mathbf{A}\|_*,$$
$$s.t. \quad \mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i, \ \mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}. \tag{2}$$

Note that for both the images with short intervals or long intervals, the nuclear norm can flexibly capture the redundancy of their abundances (see Section III-D2). By introducing auxiliary variable $\mathbf{W} = \mathbf{A}$, the problem can be rewritten as

$$\min_{\mathcal{X}, \mathcal{C}, \mathcal{A}_i, \mathbf{F}_i, \mathbf{W}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_1 + \alpha\|\mathbf{W}\|_*,$$
$$s.t. \ \mathbf{W} = \mathbf{A}, \mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i, \ \mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}. \tag{3}$$

Based on the ALM method, we deduce the following augmented Lagrangian function for problem (2):

$$L(\mathbf{F}_i, \mathcal{A}_i, \mathbf{W}, \mathcal{C}, \mathcal{X}; \mathcal{P}_i, \mathbf{Q}) = \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_1$$
$$+ \sum_{i=1}^{t} \left\{ \langle \mathcal{P}_i, \mathcal{X}_i - \mathcal{A}_i \times_3 \mathbf{F}_i \rangle + \frac{\rho}{2}\|\mathcal{X}_i - \mathcal{A}_i \times_3 \mathbf{F}_i\|_F^2 \right\}$$
$$+ \alpha\|\mathbf{W}\|_* + \langle \mathbf{Q}, \mathbf{W} - \mathbf{A} \rangle + \frac{\gamma}{2}\|\mathbf{W} - \mathbf{A}\|_F^2. \tag{4}$$

where $\mathcal{P}_i$ and $\mathbf{Q}$ are Lagrange multipliers. Then, each variable can be updated as follows:

1) Update $\{\mathbf{F}_i\}$. Each $\mathbf{F}_i$-subproblem is as follows:

$$\mathbf{F}_i^{k+1} = \arg\min_{\mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}} \left\|(\mathcal{X}_i^k + \mathcal{P}_i^k/\rho) - \mathcal{A}_i^k \times_3 \mathbf{F}_i\right\|_F^2.$$

This problem can be solved in closed-form solution [48]

$$\mathbf{F}_i^{k+1} = \mathbf{V}\mathbf{U}^T, \tag{5}$$

where $\mathbf{U}\mathbf{S}\mathbf{V}^T = (\mathbf{A}_i^k)_{(3)} \left(\mathbf{X}_i^k + \mathbf{P}_i^k/\rho\right)_{(3)}^T$ is the singular value decomposition (SVD).

2) Update $\{\mathcal{A}_i\}$. The $\{\mathcal{A}_i\}$-subproblem is as follows:

$$\{\mathcal{A}_i^{k+1}\} = \arg\min_{\{\mathcal{A}_i\}} \frac{\gamma}{2}\left\|\left(\mathbf{W}^k + \mathbf{Q}^k/\gamma\right) - \mathbf{A}\right\|_F^2$$
$$+ \sum_{i=1}^{t} \frac{\rho}{2}\left\|(\mathcal{X}_i^k + \mathcal{P}_i^k/\rho) - \mathcal{A}_i \times_3 \mathbf{F}_i^{k+1}\right\|_F^2.$$

To solve this problem, we introduce the following Theorem.

Theorem 1: If $\mathbf{F}$ is a semi-orthogonal matrix, i.e., $\mathbf{F}^T\mathbf{F} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, then

$$\arg\min_{\mathcal{A}}\|\mathcal{X} - \mathcal{A} \times_3 \mathbf{F}\|_F^2 = \arg\min_{\mathcal{A}}\|\mathcal{X} \times_3 \mathbf{F}^T - \mathcal{A}\|_F^2.$$

The proof of the theorem will be provided in Appendix A.

According to Theorem 1, we have

$$\{\mathcal{A}_i^{k+1}\} = \arg\min_{\{\mathcal{A}_i\}} \sum_{i=1}^{t} \left\{ \frac{\gamma}{2}\left\|\mathscr{R}_i^{-1}\left(\mathbf{W}^k + \mathbf{Q}^k/\gamma\right) - \mathscr{R}_i^{-1}\mathbf{A}\right\|_F^2 \right.$$
$$\left. + \frac{\rho}{2}\left\|(\mathcal{X}_i^k + \mathcal{P}_i^k/\rho) \times_3 (\mathbf{F}_i^{k+1})^T - \mathcal{A}_i\right\|_F^2 \right\}.$$

Then, we get each $\mathcal{A}_i$-subproblem,

$$
\mathcal{A}_i^{k+1} = \arg \min_{\mathcal{A}_i} \frac{\gamma}{2} \left\| \mathscr{R}_i^{-1} \left(\mathbf{W}^k + \mathbf{Q}^k/\gamma \right) - \mathcal{A}_i \right\|_F^2
$$
$$
+ \frac{\rho}{2} \left\| \left(\mathcal{X}_i^k + \mathcal{P}_i^k/\rho \right) \times_3 \left(\mathbf{F}_i^{k+1} \right)^T - \mathcal{A}_i \right\|_F^2.
$$

Taking the derivative with respect to $\mathcal{A}_i$, we have the exact solution of $\mathcal{A}_i$:

$$
\mathcal{A}_i^{k+1} = \frac{(\rho \mathcal{X}_i^k + \mathcal{P}_i^k) \times_3 (\mathbf{F}_i^{k+1})^T + \mathscr{R}_i^{-1}(\gamma \mathbf{W}^k + \mathbf{Q}^k)}{\rho + \gamma}. \quad (6)
$$

3) Update W. The **W**-subproblem is as follows:

$$
\mathbf{W}^{k+1} = \arg \min_{\mathbf{W}} \alpha \|\mathbf{W}\|_* + \frac{\gamma}{2} \left\| \left(\mathbf{A}^{k+1} - \mathbf{Q}^k/\gamma \right) - \mathbf{W} \right\|_F^2.
$$

This problem can be solved by singular value thresholding [49] and its closed-form solution:

$$
\mathbf{W}^{k+1} = \mathbf{U}\text{Diag}\{(\boldsymbol{\Sigma} - \alpha/\gamma)_+\}\mathbf{V}^T, \quad (7)
$$

where $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is from the SVD of $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{A}^{k+1} - \mathbf{Q}^k/\gamma$ and $(\boldsymbol{\Sigma}_{jj} - \alpha/\gamma)_+ = \max\{\boldsymbol{\Sigma}_{jj} - \alpha/\gamma, 0\}$.

4) Update $\mathcal{C}$. The $\mathcal{C}$-subproblem is as follows:

$$
\mathcal{C}^{k+1} = \arg \min_{\mathcal{C}_i} \frac{1}{2} \|(\mathcal{Y} - \mathcal{M} \odot \mathcal{X}^k) - \mathcal{C}\|_F^2 + \beta \|\mathcal{C}\|_1.
$$

The solution to this problem is given by

$$
\mathcal{C}^{k+1} = \text{soft}_\beta(\mathcal{Y} - \mathcal{M} \odot \mathcal{X}^k), \quad (8)
$$

where $\text{soft}_\beta(\mathcal{A}) = \text{sgn}(a_{ijk})\max(|a_{ijk}| - \beta, 0)$.

5) Update $\{\mathcal{X}\}$. $\mathcal{X}$-subproblem is as follows:

$$
\mathcal{X}^{k+1} = \arg \min_{\mathcal{X}} \frac{1}{2} \left\| (\mathcal{Y} - \mathcal{C}^{k+1}) - \mathcal{M} \odot \mathcal{X} \right\|_F^2
$$
$$
+ \sum_{i=1}^t \frac{\rho}{2} \left\| (\mathcal{A}_i^{k+1} \times_3 \mathbf{F}_i^{k+1} - \mathcal{P}_i^k/\rho) - \mathcal{X}_i \right\|_F^2. \quad (9)
$$

Let $\mathcal{X} = \mathscr{C}\{\mathcal{X}_i\} \in \mathcal{R}^{m \times n \times bt}$, (9) can be rewritten as

$$
\mathcal{X}^{k+1} = \arg \min_{\mathcal{X}} \frac{1}{2} \left\| (\mathcal{Y} - \mathcal{C}^{k+1}) - \mathcal{M} \odot \mathcal{X} \right\|_F^2
$$
$$
+ \frac{\rho}{2} \left\| \mathscr{C} \{ \mathcal{A}_i^{k+1} \times_3 \mathbf{F}_i^{k+1} - \mathcal{P}_i^k/\rho \} - \mathcal{X} \right\|_F^2. \quad (10)
$$

The closed-form solution of $\mathcal{X}$ is

$$
\mathcal{X}^{k+1} = \left[\mathcal{M} \odot \left(\mathcal{Y} - \mathcal{C}^{k+1} \right) + \mathscr{C} \{ \rho \mathcal{A}_i^{k+1} \times_3 \mathbf{F}_i^{k+1} - \mathcal{P}_i^k \} \right]
$$
$$
\oslash (\mathcal{M} + \rho \mathcal{J}), \quad (11)
$$

where $\oslash$ denotes the point-wise division and $\mathcal{J}$ is all-ones tensor.

6) Update $\{\mathcal{P}_i\}$ and Q. Each $\mathcal{P}_i$ and $\mathbf{Q}$ are updated as

$$
\begin{cases} \mathcal{P}_i^{k+1} = \mathcal{P}_i^k + \rho(\mathcal{X}_i^{k+1} - \mathcal{A}_i^{k+1} \times_3 \mathbf{F}_i^{k+1}), \\ \mathbf{Q}^{k+1} = \mathbf{Q}^k + \gamma(\mathbf{W}^{k+1} - \mathbf{A}^{k+1}). \end{cases} \quad (12)
$$

Finally, the developed algorithm for solving the proposed model is summarized in Algorithm 2. In the algorithm, we initialize $\mathbf{F}_i^0 = \mathbf{U}(:, 1 : r)$, where $\mathbf{U}$ is from the SVD of $\mathbf{U}\mathbf{S}\mathbf{V}^T = (\mathbf{X}_i^k + \mathbf{P}_i^k/\rho)_{(3)}$ and r is estimated by HySime [50].

Algorithm 2 ALM Algorithm for Cloud/Shadow Removal

Input: Target RS images $\mathcal{Y}$, regularization parameters α and β, and penalty parameters γ and ρ.
1: Initialize: $\mathcal{X} = \mathcal{Y}$, $\mathcal{C} = \mathcal{P}_i = \mathcal{O}$, and $\mathbf{W} = \mathbf{Q} = \mathbf{0}$.
2: **while** not converged **do**
3:      Update $\{\mathbf{F}_i^{k+1}\}$ by (5);
4:      Update $\{\mathcal{A}_i^{k+1}\}$ by (6);
5:      Update $\mathbf{W}^{k+1}$ by (7);
6:      Update $\mathcal{C}^{k+1}$ by (8);
7:      Update $\mathcal{X}^{k+1}$ by (11);
8:      Update $\{\mathcal{P}_i^{k+1}\}$ and $\mathbf{Q}^{k+1}$ by (12);
9:      Refine mask $\mathcal{M}$ by Algorithm 1;
10:     Check the convergence condition:
         $\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F^2/\|\mathcal{X}^k\|_F^2 \leq 10^{-4}$.
11: **end while**
Output: Reconstructed RS images $\mathcal{X}$.

III. EXPERIMENTS AND RESULTS

To demonstrate the effectiveness of the proposed method for thick cloud/shadow removal, we select 4 state-of-the-art methods as comparison methods, including non-blind completion method HaLRTC [51], non-blind cloud removal methods ALM-IPG [8] and WLR [35], and blind cloud removal method TVLRSDC [9]. We set the parameters of the comparison methods according to the source code settings or the authors' suggestions. The parameter setting of the proposed RTCR is reported in Section III-D6. We use $\mathcal{J} - \mathcal{M}$ for the visual display of the cloud mask in this Section.

A. Simulated Experiment

Since it is difficult to quantitively assess the cloud/shadow removal results in real experiments, we perform simulated experiments to quantitively and qualitatively verify the effectiveness of the proposed RTCR. The detailed information about the test datasets and the evaluation indexes are as follows:
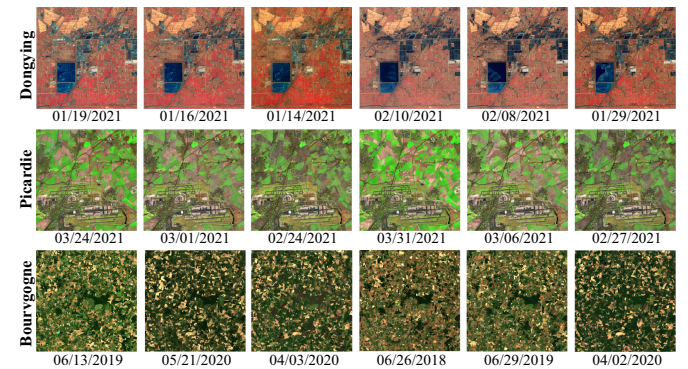

Fig. 5. Pseudo-color images (R: B4, G: B3, B: B2) of three simulated datasets. ("MM/DD/YYYY" means the date of taking the image.)

Datasets: To test the robustness of RTCR on different type datasets, two Sentinel-2 L2A/B datasets, and a Landsat-8 OLI dataset are selected in simulated experiments. The pseudo-color images of the referred datasets are shown in Fig. 5.

- Dongying[1]. This dataset is taken over Dongying, China, by Sentinel-2, and each time node contains 4 spectral bands (B2, B3, B4, and B8) with 10m spatial resolution. The sub-images of size $500 \times 500 \times 4$ of 6 time nodes are used in experiments.
- Picardie[1]. This dataset is taken over Picardie, France, by Sentinel-2, and each time node contains 6 spectral bands (B5, B6, B7, B8A, B11, and B12) with 20m spatial resolution. The sub-images of size $1000 \times 1000 \times 6$ of 6 time nodes are used in experiments.
- Bourgogne[2]. This dataset is taken over Bourgogne, France, by Landsat-8, and each time node contains 7 spectral bands (B1, B2, B3, B4, B5, B6, and B7) with 30m spatial resolution. The sub-images of size $400 \times 400 \times 7$ of 6 time nodes are used in experiments.

Evaluation indexes: We select three picture quality indexes to measure the results quantitatively, including the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) [52], and the correlation coefficient (CC) [53], which are defined as follows:

$$\text{PSNR}_i = 10 \times \log \frac{pq}{\|\mathbf{X}_i - \hat{\mathbf{X}}_i\|_F^2},$$

$$\text{SSIM}_i = \frac{(2\mu_{\mathbf{X}_i}\mu_{\hat{\mathbf{X}}_i} + c_1)(2\sigma_{\mathbf{X}_i\hat{\mathbf{X}}_i})}{(\mu_{\mathbf{X}_i}^2 + \mu_{\hat{\mathbf{X}}_i}^2 + c_1)(\sigma_{\mathbf{X}_i}^2 + \sigma_{\hat{\mathbf{X}}_i}^2 + c_2)},$$

$$\text{CC} = \frac{\sum_{j=1}^k (\mathcal{X}_j - \mu_{\mathcal{X}})(\hat{\mathcal{X}}_j - \mu_{\hat{\mathcal{X}}})}{\sqrt{\sum_{j=1}^k (\mathcal{X}_j - \mu_{\mathcal{X}})^2 \sum_{j=1}^k (\hat{\mathcal{X}}_j - \mu_{\hat{\mathcal{X}}})^2}}$$

where p and q denote the size of each band, $\mathbf{X}_i$ and $\hat{\mathbf{X}}_i$ denote the ith band of original image and reconstructed image; $\mu_{\mathbf{X}_i}$ and $\mu_{\hat{\mathbf{X}}_i}$ denote the average values of $\mathbf{X}_i$ and $\hat{\mathbf{X}}_i$, $\sigma_{\mathbf{X}_i}^2$ and $\sigma_{\hat{\mathbf{X}}_i}^2$ stand for the variances, and $\sigma_{\mathbf{X}_i\hat{\mathbf{X}}_i}$ is the covariance between $\mathbf{X}_i$ and $\hat{\mathbf{X}}_i$, c_1 and c_2 are default constants; $\mathcal{X}_j$ and $\hat{\mathcal{X}}_j$ are the original and the reconstructed values of the jth contaminated pixels, respectively, k denotes the number of contaminated pixels, $\mu_{\mathcal{X}}$ and $\mu_{\hat{\mathcal{X}}}$ are the average values.

To verify the robustness of RTCR to the given masks with different accuracy, the experimental procedure adopted is 1) to give the accurate mask and assess the information reconstruction performance and 2) to provide the inaccurate mask (with undetected clouds) and evaluate both the mask refinement and information reconstruction performance.

1) Accurate mask: In the three datasets, we select three temporal images to add simulated cloud/shadow as cloud-contaminated images and employ the rest three as the reference images (see Table I and Fig. 5 for details).

For the three simulated data in each dataset, we add different simulated cloud/shadow to them, which corresponds to Case 1 to 3. For the Dongying dataset and Bourgogne dataset, we mainly test the effectiveness of the proposed RTCR for the multiple small, middle, and large clouds removal; For the Picardie dataset, we mainly test the effectiveness of the proposed RTCR for the large cloud/shadow and huge cloud removal. The detailed setting to the three datasets in simulated experiments is presented in Table I.

[1]https://earthexplorer.usgs.gov
[2]https://theia.cnes.fr/atdistrib/rocket/#/home

TABLE I
SELECTION OF SIMULATED DATA AND SETTING OF SIMULATION CLOUD IN DIFFERENT CASES OF THREE DATASETS.

Dataset	Case	Simulated data	Cloud/Shadow region			
			Small	Middle	Large	Huge
	Case 1	01/19/2021	2	–	1	–
Dongying	Case 2	01/16/2021	–	1	–	–
	Case 3	01/14/2021	1	2	–	–
	Case 1	03/24/2021	2	–	1	–
Picardie	Case 2	03/01/2021	–	–	1	–
	Case 3	02/24/2021	–	–	–	1
	Case 1	06/13/2019	2	–	1	–
Bourgogne	Case 2	05/21/2020	–	1	–	–
	Case 3	04/03/2020	1	2	–	–

TABLE II
QUANTITATIVE EVALUATION OF CLOUD/SHADOW REMOVAL RESULTS BY ALL METHODS ON THREE DATASETS.

		Dongying					
Case	Index	Target	HaLRTC	ALM-IPG	WLR	TVLRSDC	RTCR
	PSNR	12.802	36.913	41.761	39.097	43.131	**46.536**
Case 1	SSIM	0.8686	0.9702	0.9931	0.9837	0.9927	**0.9963**
	CC	0.3315	0.9776	0.9962	0.9842	0.9953	**0.9974**
	PSNR	15.138	38.761	39.226	38.710	42.237	**45.788**
Case 2	SSIM	0.9334	0.9856	0.9934	0.9829	0.9935	**0.9964**
	CC	0.1198	0.9741	0.9876	0.9821	0.9920	**0.9975**
	PSNR	14.579	40.754	31.726	37.197	38.221	**45.656**
Case 3	SSIM	0.8937	0.9861	0.9738	0.9818	0.9926	**0.9963**
	CC	0.3445	0.9898	0.9426	0.9636	0.9865	**0.9961**
Time (min)		—	**3.104**	6.876	4.962	8.210	3.763
		Picardie					
Case	Index	Target	HaLRTC	ALM-IPG	WLR	TVLRSDC	RTCR
	PSNR	9.632	44.295	44.010	44.717	49.614	**52.206**
Case 1	SSIM	0.8344	0.9873	0.9876	0.9888	0.9961	**0.9978**
	CC	0.1270	0.9353	0.9214	0.9686	0.9773	**0.9932**
	PSNR	7.5131	40.850	45.286	47.261	49.530	**50.515**
Case 2	SSIM	0.7632	0.9682	0.9896	0.9938	0.9959	**0.9975**
	CC	0.1134	0.8736	0.9281	0.9745	0.9797	**0.9873**
	PSNR	4.166	39.032	38.630	44.514	45.513	**47.048**
Case 3	SSIM	0.5188	0.9567	0.9620	0.9955	0.9904	**0.9941**
	CC	0.1097	0.7238	0.6346	0.8057	0.9184	**0.9258**
Time (min)		—	1.065	1.705	4.882	9.617	**0.988**
		Bourgogne					
Case	Index	Target	HaLRTC	ALM-IPG	WLR	TVLRSDC	RTCR
	PSNR	11.081	36.357	33.056	32.757	35.675	**39.117**
Case 1	SSIM	0.8378	0.9412	0.9345	0.9279	0.9522	**0.9700**
	CC	0.2373	0.9488	0.9011	0.8295	0.9406	**0.9743**
	PSNR	14.108	36.185	35.996	30.001	36.223	**38.414**
Case 2	SSIM	0.9194	0.9651	0.9744	0.9056	0.9706	**0.9807**
	CC	0.4975	0.9785	0.9752	0.8653	0.9787	**0.9869**
	PSNR	11.807	35.950	36.765	37.202	38.091	**40.596**
Case 3	SSIM	0.8543	0.9538	0.9713	0.9778	0.9742	**0.9854**
	CC	0.3216	0.9598	0.9689	0.9786	0.9780	**0.9877**
Time (min)		—	1.964	1.316	3.373	3.159	**0.865**

We present the quantitative comparison in Table II. The highest PSNR, SSIM, and CC are highlighted in bold. Table II shows that all comparison methods obtain satisfactory results under different cases on three datasets. For multiple small contaminated regions, i.e., on Dongying and Bourgogne datasets, the performance of HaLRTC is similar to that of ALM-IPG and WLR. When it becomes large or even huge contaminated regions, i.e., on Picardie dataset, the cloud removal methods WLR, TVLRSDC, and RTCR have better cloud removal results than the completion method HaLRTC. Considering three evaluation indexes comprehensively, cloud removal methods
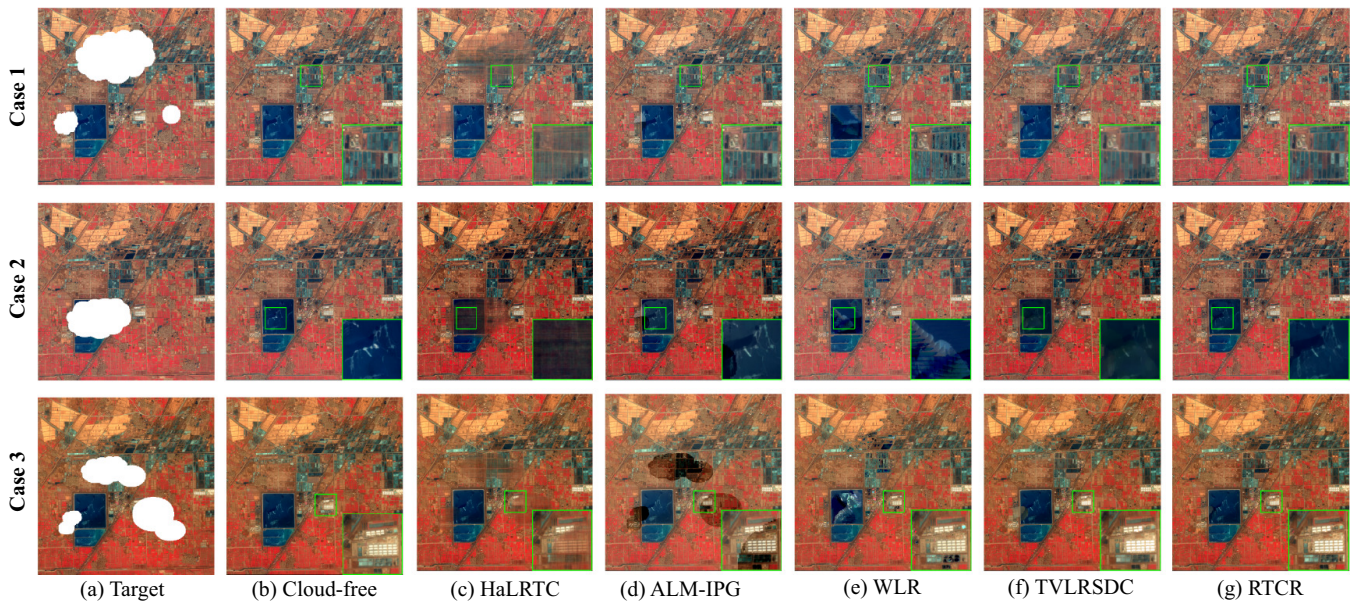
Fig. 6. (a) Pseudo-color images (R:B8, G:B4, B:B3) of simulated target images in Cases 1-3 of Dongying dataset; (b) Cloud-free images; (c)-(g) Cloud removal results by all methods.
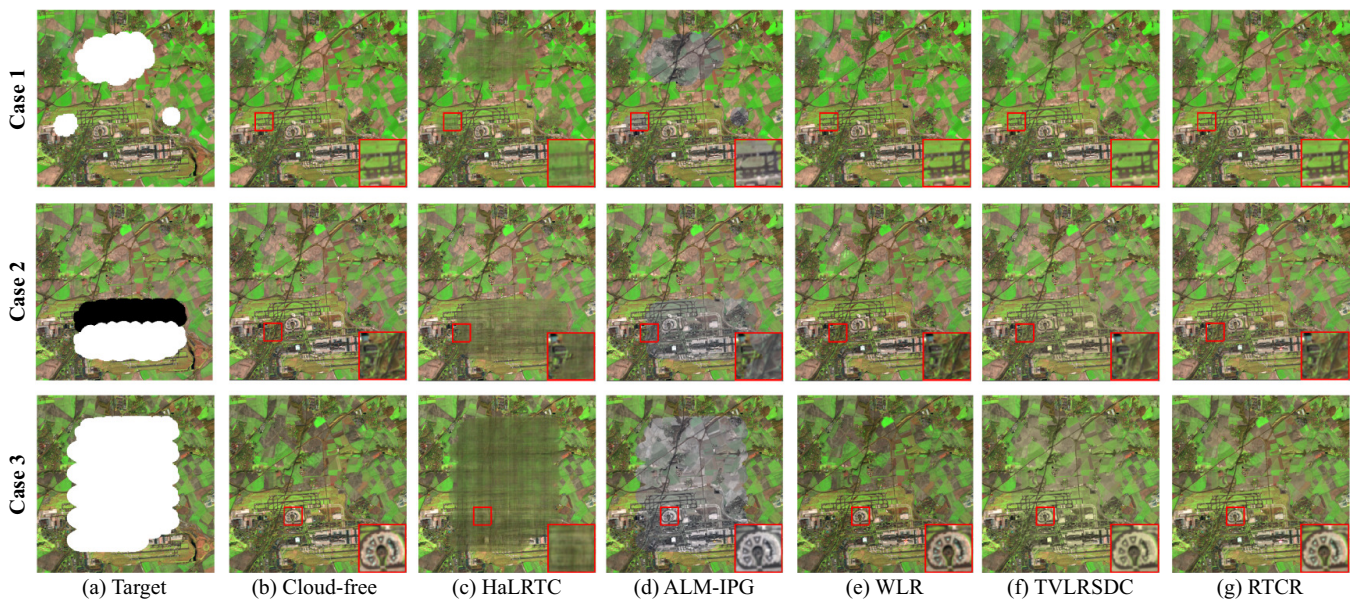

Fig. 7. (a) Pseudo-color images (R:B11, G:B7, B:B5) of simulated target images in Cases 1-3 of Picardie dataset; (b) Cloud-free images; (c)-(g) Cloud removal results by all methods.

obtain better results than the completion method HaLRTC, especially for the removal of large region clouds. For the cloud removal methods, RTCR achieves superior performance than others. Regarding PSNR, the results of the proposed RTCR achieve around 31-44dB improvement over the target images on Dongying and Picardie datasets and exceed around 24-29dB than the target images on the Bourgogne dataset. On the whole, RTCR significantly outperforms all comparison methods in terms of all evaluation indexes.

In terms of visual quality, the target images, cloud-free images, and reconstructed results on three datasets under different cases are shown in Figs. 6-8, respectively, where the same area of the subfigures in each figure is enlarged for a

clear comparison. Fig. 6 displays the pseudo-color images of the reconstructed results in Cases 1-3 of the Dongying dataset. From Fig. 6, we observe that HaLRTC and ALM-IPG cannot finely reconstruct the original texture details under the clouds. From Fig. 6 (f), TVLRSDC fails to preserve some details in Case 2. Overall, WLR and RTCR achieve a more satisfactory cloud removal effect. Fig. 7 illustrates the pseudo-color images of the reconstructed results in Cases 1-3 of the Picardie dataset. From Fig. 7 (c)-(g), it can be observed that, for large and huge cloud removal, all methods achieve outstanding cloud removal effect and detail preservation except for HaLRTC and ALM-IPG. Fig. 8 displays the pseudo-color images of the reconstructed results in Cases 1-3 of the Bourgogne dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140800, IEEE Transactions on Geoscience and Remote Sensing
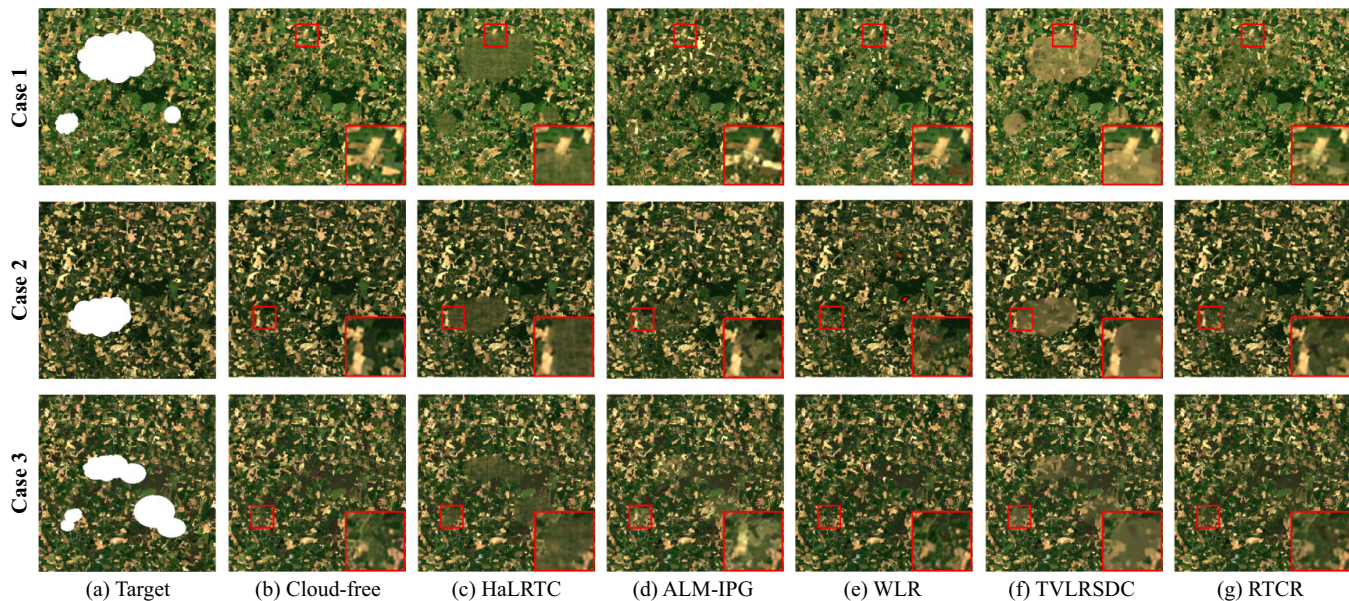
8

Fig. 8. (a) Pseudo-color images (R:B4, G:B3, B:B2) of simulated target images in Cases 1-3 of Bourgogne dataset; (b) Cloud-free images; (c)-(g) Cloud removal results by all methods.
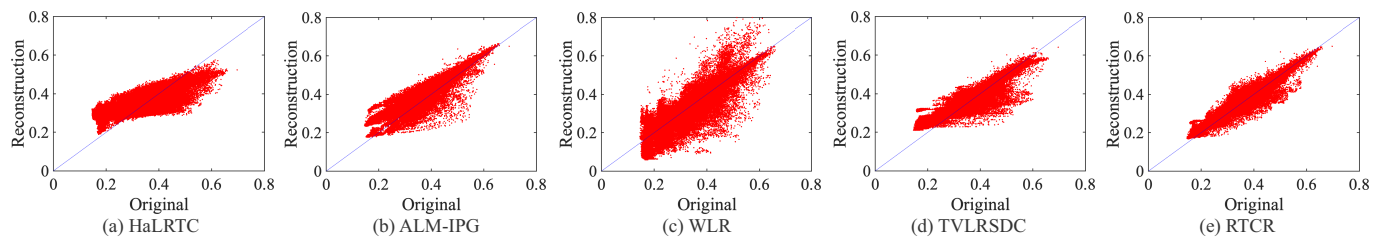

Fig. 9. Scatter diagrams between the original and reconstructed pixels of the cloud-contaminated regions in Case 2 of Dongying dataset.

Fig. 10. Scatter diagrams between the original and reconstructed pixels of the cloud-contaminated regions in Case 3 of Picardie dataset.

From Fig. 8 (c), HaLRTC fails to recover the details of the original information. Although WLR seems to have more ideal results on the whole, from the enlarged area, there are still have distortions compared with the cloud-free images. From the results of ALM-IPG and TVLRSDC, there is a color difference between the reconstructed region and the cloud-free region. Compared with all methods, the proposed RTCR achieves the closest results to the cloud-free images.

To further measure the reconstruction quality of the cloud-contaminated region, we display the scatter diagrams between the original pixels and the reconstructed pixels by different methods on three datasets. Fig. 9 shows the scatter diagrams in Case 2 of the Dongying dataset. From Fig. 9 (c), there are some points that deviate from the diagonal line. The points of ALM-IPG, TVLRSDC, and RTCR are mostly distributed surrounding the diagonal line. Fig. 10 shows the scatter diagrams in Case 3 of the Picardie dataset. From Fig. 10 (b),

some points deviate from the blue diagonal, which illustrates the color aberration in the reconstructed result of ALM-IPG. Although all methods except HaLRTC and ALM-IPG achieve satisfactory visual performance in Fig. 7, the points of RTCR in Fig. 10, which are mostly distributed surrounding the diagonal line, demonstrates the advantage of the proposed RTCR. Fig. 11 shows the scatter diagrams in Case 1 of the Bourgogne dataset. It is clear that the points of RTCR are mostly distributed surrounding the diagonal line. From Figs. 9-11, RTCR achieves the best reconstruction performance for the cloud-contaminated regions.

From the experiments with the accurate mask, we can observe that RTCR shows a clear superiority on multi-temporal information reconstruction. This demonstrates the advantage of using coupled tensor decomposition to reproduce the multi-temporal relationship of target images.
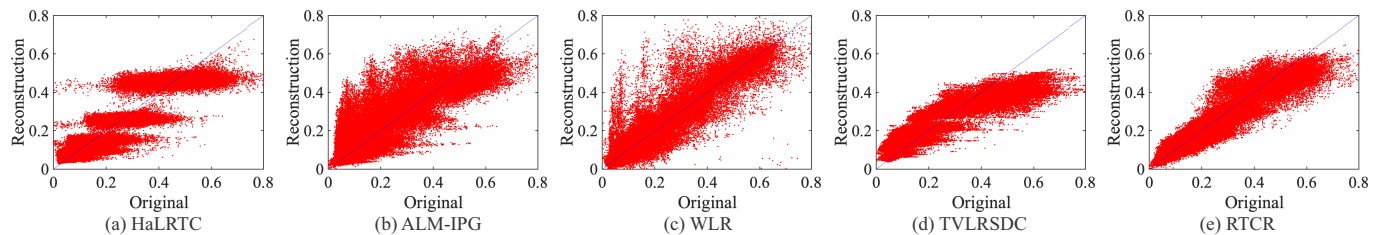
Fig. 11. Scatter diagrams between the original and reconstructed pixels of the cloud-contaminated regions in Case 1 of Bourgogne dataset.
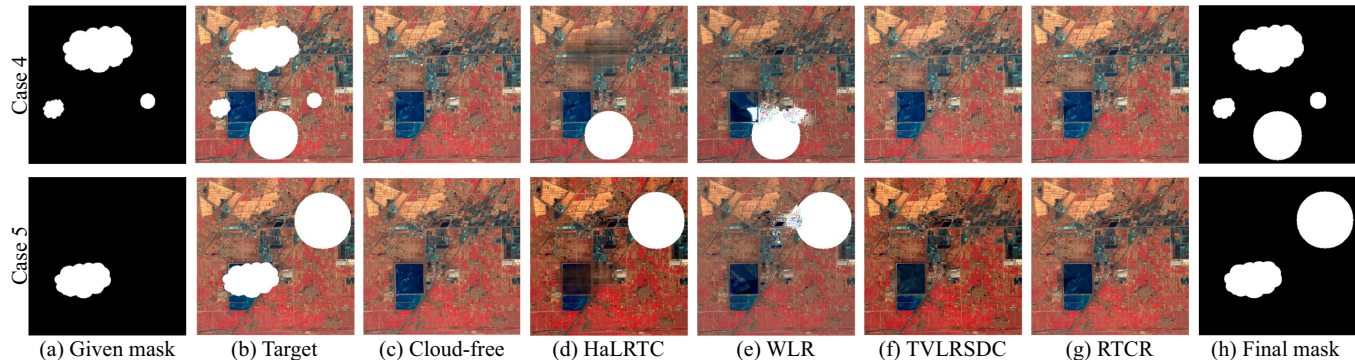

Fig. 12. (a) Given masks; (b) Pseudo-color images (R:B8, G:B4, B:B3) of the simulated target images in Cases 4-5 of Dongying dataset; (c) Cloud-free images; (d)-(g) Cloud removal results by four methods; (h) Final mask by RTCR.

2) Inaccurate mask: There are two types of inaccuracy in cloud masks: the mask contains misdetected clouds (i.e., over-detection) and the mask has undetected clouds (i.e., under-detection). The proposed RTCR can easily handle the former (see Section III-D1), and we focus on the latter in the following experiments.

To verify the robustness and effectiveness of the proposed RTCR for cloud/shadow removal under inaccurate masks, we add simulated clouds/shadows to the target images in Cases 1-2 of the Dongying dataset and Picardie dataset and denote them as Cases 4-5. We use the cloud mask in Cases 1-2 as the mask input of the proposed RTCR and the newly added clouds represent the undetected clouds. We select non-blind methods HaLRTC and WLR, and the blind method TVLRSDC as our comparison method.

TABLE III
QUANTITATIVE EVALUATION OF CLOUD/SHADOW REMOVAL RESULTS BY FOUR METHODS ON TWO DATASETS.

Dataset	Case	Index	Target	HaLRTC	WLR	TVLRSDC	RTCR
Dongying	Case 4	PSNR	11.005	15.066	14.583	41.420	**44.691**
		SSIM	0.8100	0.8955	0.8772	0.9894	**0.9955**
		CC	0.3584	0.5854	0.5735	0.9936	**0.9970**
	Case 5	PSNR	11.450	13.645	13.412	39.336	**44.062**
		SSIM	0.8452	0.8921	0.8479	0.9873	**0.9952**
		CC	0.0809	0.2626	0.2877	0.9854	**0.9975**
	Time (min)		—	**3.054**	4.879	9.738	3.610
Picardie	Case 4	PSNR	6.819	11.879	13.673	45.757	**47.896**
		SSIM	0.7055	0.8530	0.7757	0.9912	**0.9952**
		CC	0.0988	0.1398	0.1607	0.9559	**0.9838**
	Case 5	PSNR	5.903	10.573	10.822	46.398	**50.052**
		SSIM	0.6530	0.8413	0.7405	0.9912	**0.9971**
		CC	0.0649	0.0068	0.0205	0.9597	**0.9847**
	Time (min)		—	**1.047**	4.971	4.542	1.508

To quantitatively and qualitatively evaluate the effect of cloud removal, Table III reports the quantitative comparison of four methods in Cases 4-5 of the Dongying dataset and Picardie dataset. Since the non-blind methods cannot

reconstruct the undetected cloud/shadow regions, they achieve unsatisfactory results. Compare with all methods, it can be seen that the RTCR achieves better performance in all three evaluation measures in all cases of two datasets. Figs. 12-13 (d)-(g) display the visual comparison of reconstructed results by four methods in Cases 4-5 of Dongying dataset and Picardie dataset. We observe that the non-blind methods fail to remove the undetected cloud/shadow due to the inaccurate mask. Both TVLRSDC and RTCR attain visually satisfactory results compared with cloud-free images, which demonstrates the two methods get rid of the dependence on the given mask. For a deeper comparison, Figs. 14-15 shows the scatter diagrams between the original and reconstructed pixels on Dongying dataset and Picardie dataset, where the red points represent the pixels under undetected cloud/shadow, and the green represents the pixels under the cloud/shadow in the given mask. We observe that HaLRTC completely fails to remove the undetected cloud/shadow and WLR removes part of the undetected cloud/shadow. From Fig. 14 (d) and Fig. 15 (d), it can be seen that both the red and green points of RTCR are more concentrated on the diagonal line than all methods, which indicates the robustness and effectiveness of RTCR for cloud/shadow removal under the inaccurate given mask. To verify the effect of the mask refinement of RTCR, Figs. 12-13 (a) and (h) show the given masks and the final masks by RTCR. It can be observed that the proposed method captures undetected clouds and shadows successfully.

From the experiments under the inaccurate mask, the proposed RTCR shows promising performance in mask refinement and information reconstruction, which circumvents the shortages of both non-blind methods and the blind method TVLRSDC.

Regarding the running time in Tables II and III, HaLRTC is faster than RTCR, and RTCR is faster than ALM-IPG, WLR, and TVLRSDC. Considering the cloud removal effects, the proposed RTCR attains the best effect/efficiency trade-off.

(a) Given mask   (b) Target   (c) Cloud-free   (d) HaLRTC   (e) WLR   (f) TVLRSDC   (g) RTCR   (h) Final mask
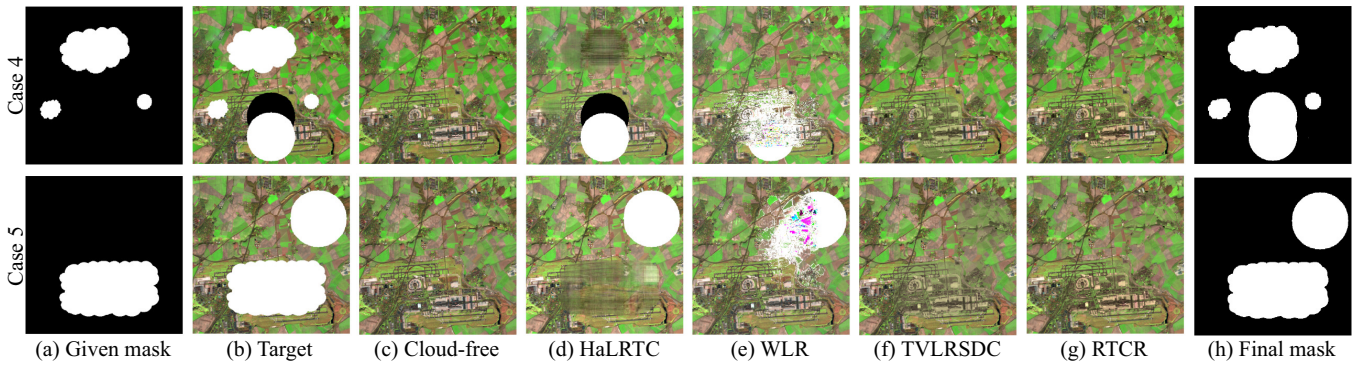
Fig. 13. (a) Given mask; (b) Pseudo-color image (R:B11, G:B7, B:B5) of the simulated target images in Cases 4-5 of Picardie dataset; (c) Cloud/Shadow-free images; (d)-(g) Cloud/Shadow removal results by four methods; (h) Final mask by RTCR.
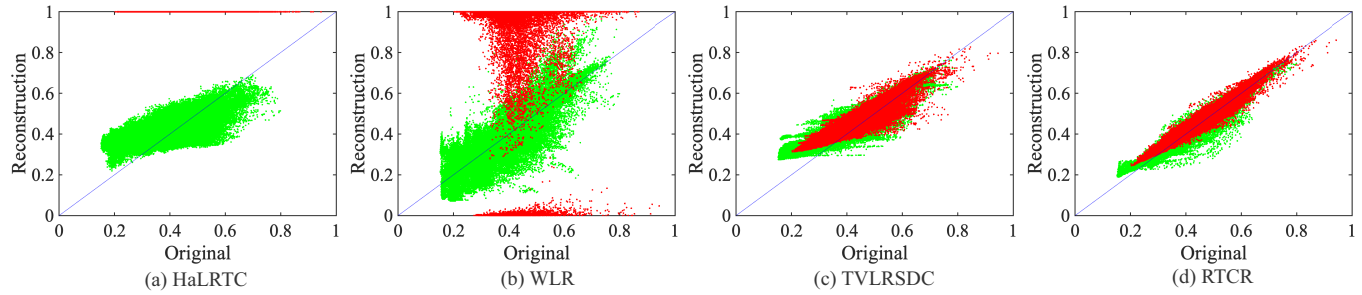

(a) HaLRTC   (b) WLR   (c) TVLRSDC   (d) RTCR

Fig. 14. Scatter diagrams between the original and reconstructed pixels of the cloud-contaminated regions in Case 4 of Dongying dataset. The red points represent the pixels in undetected cloud/shadow-contaminated regions, and the green represents the pixels in the cloud/shadow region of given masks.

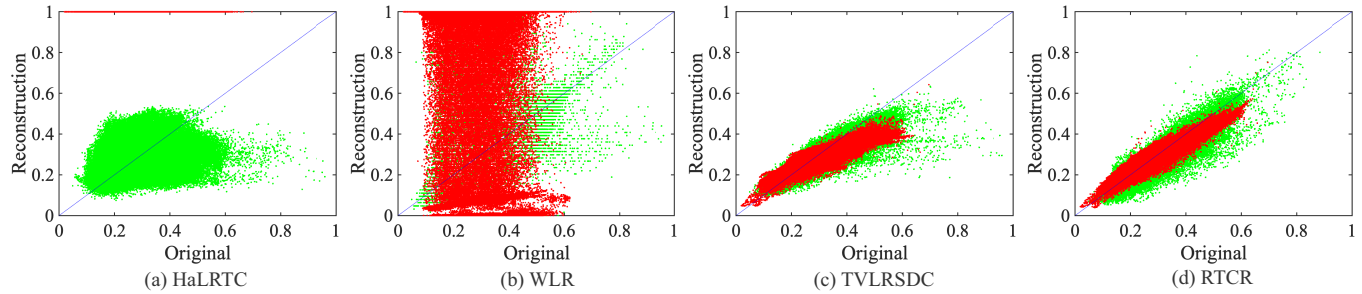(a) HaLRTC   (b) WLR   (c) TVLRSDC   (d) RTCR

Fig. 15. Scatter diagrams between the original and reconstructed pixels of the cloud-contaminated regions in Case 4 of Picardie dataset. The red points represent the pixels in undetected cloud/shadow-contaminated regions, and the green represents the pixels in the cloud/shadow region of given masks.
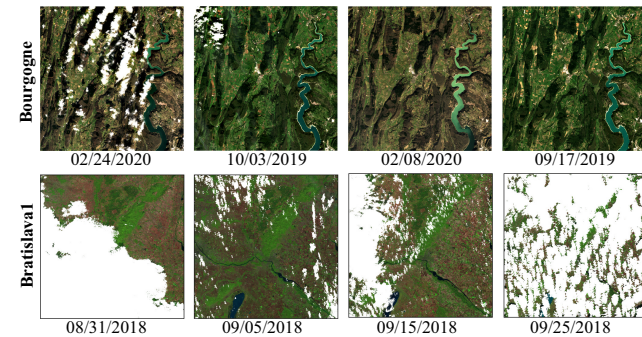
B. Real Experiment

Fig. 16. Pseudo-color images of Bourgogne (R:B4, G:B3, B:B2) and Bratislava1 (R:B11, G:B7, B:B5) datasets. ("MM/DD/YYYY" means the date of taking the image.)

Two datasets are selected to test the performance of the proposed RTCR under real scenarios. The pseudo-color images of the referred datasets are shown in Fig. 16.

- Bourgogne[2]. This dataset is taken by Landsat-8 and each time node contains 7 spectral bands with 30m spatial resolution. The sub-images of size $600 \times 600 \times 7$ of 4 time nodes are used in experiments.
- Bratislava[1]. This dataset is taken over Bratislava, Slovakia, by Sentinel-2 and each time node contains 6 spectral bands with 20m spatial resolution. The full-images of size $5490 \times 5490 \times 6$ of 4 time nodes are used in experiments.

We select Bourgogne dataset to test the effectiveness of the RTCR under inaccurate masks. For cloud-contaminated images, the corresponding cloud/shadow masks are also acquired from the website[2]. Fig. 17 shows the pseudo-color images of the target and reconstructed images on Bourgogne dataset, where the same area of subfigures is enlarged for a more detailed comparison. For the top row of Fig. 17, from given masks and target images, there is an undetected cloud in the upper right corner of the image. From Fig. 17 (c)-(g), all non-blind methods cannot remove the undetected cloud due

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140800, IEEE Transactions on Geoscience and Remote Sensing
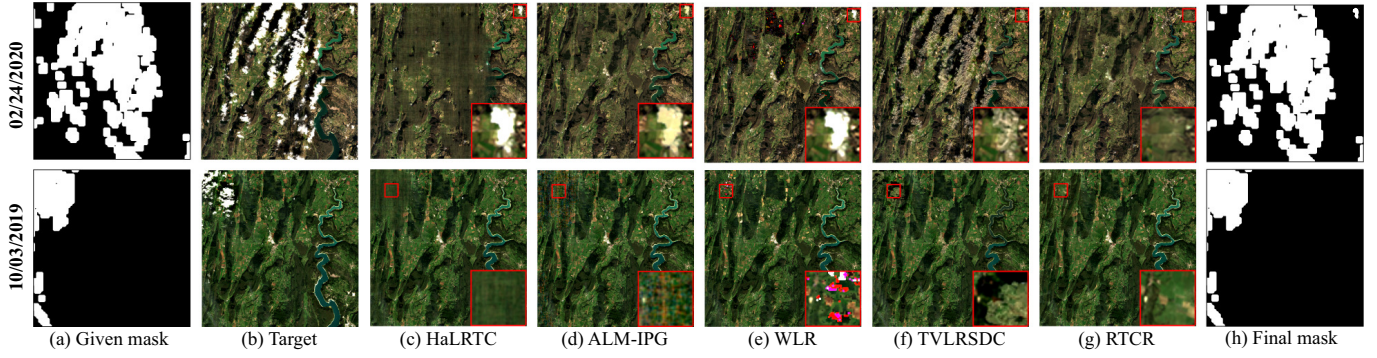
11

Fig. 17. (a) Given masks; (b) Pseudo-color images (R:B4, G:B3, B:B2) of the real target images on Bourgogne dataset; (c)-(g) Cloud/Shadow removal results by all methods; (h) Final mask by RTCR.

to the inaccurate given mask. The blind method TVLRSDC partly removes the undetected cloud and the proposed RTCR completely removes it, which implies the superiority of an adaptive threshold value in RTCR than the fixed in TVLRSDC. For the bottom row of Fig. 17, from the enlarged area of (c)-(g), it can be observed the proposed RTCR gains the best detail reconstruction. Fig. 17 (h) shows the final masks by RTCR, we can observe that the undetected cloud is captured. The experiments on Bourgogne dataset verify the robustness of the proposed RTCR for cloud removal under inaccurate masks. Moreover, we select Bratislava dataset to test the effectiveness of the RTCR on large scene images. Fig. 18 shows the pseudo-color images of target images and reconstructed images by RTCR on Bratislava dataset. The corresponding masks are generated by MSCFF [54]. It can be observed that for both the large and small region contamination, RTCR can finely reconstruct the underlying information. All experimental results verify the effectiveness of the proposed RTCR on multi-temporal images with scenes of different sizes.

C. Complexity Analysis

For an input image $\mathcal{Y} \in \mathbb{R}^{m \times n \times bt}$ whose spatial size is much larger than its spectral-temporal size, the main per-iteration cost lies in the update of $\mathbf{F}_i (i = 1, 2, \cdots, t)$ and $\mathbf{W}$. The major cost for updating $\mathbf{F}_i$ is computing an SVD on $mn \times b$ matrix and the cost is $\mathcal{O}(mnb^2)$. The major cost for updating $\mathbf{W}$ is computing an SVD on $mn \times bt$ matrix and the cost is $\mathcal{O}(mnr^2t^2)$, where $r < b$. So the cost of Algorithm 2 is $\mathcal{O}(mnr^2t^2)$. Compared with iterative optimization methods HaLRTC, ALM-IPG, and TVLRSDC, whose computational complexities are $\mathcal{O}(mnb^2t^2)$, the proposed algorithm enjoys a lower computational complexity although it occupies more running memory.

D. Discussion

1) Rationality behind the Threshold Algorithm: We explain the rationality behind the proposed threshold algorithm. In the algorithm, we use the minimum mean value of the spectral pixels within the given cloud/shadow region as the threshold value, which is adaptively and automatically calculated at each iteration instead of predefined. Fig. 19 and 20 show the results and corresponding PSNR values on Dongying dataset and Picardie dataset. We can observe from Fig. 19
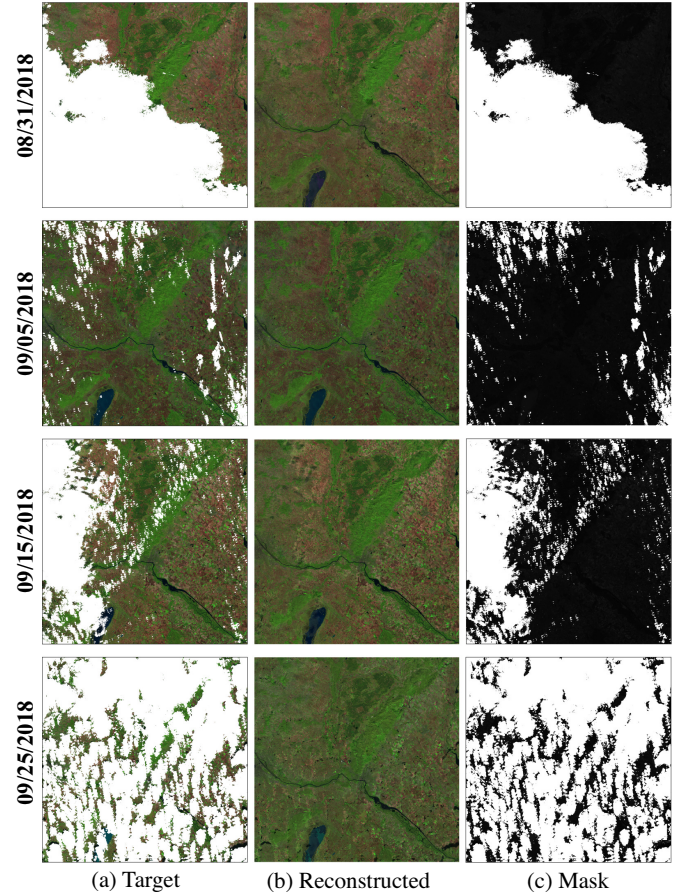

Fig. 18. Pseudo-color images (R:B11, G:B7, B:B5) of the real target images on Bratislava1 dataset, pseudo-color images of the cloud/shadow removal results by RTCR, and corresponding masks.

and 20, the large threshold value (τ=0.80) leads to under-detection of cloud/shadow, and the small threshold values (Dongying dataset: τ=0.10; Picardie dataset: τ=0.20) lead to over-detection while producing satisfactory results. This observation motivates us to set the minimum mean value as the threshold value. In addition, the result by the proposed adaptive threshold algorithm is better than the best result of fixed threshold values (Dongying dataset: τ=0.20; Picardie dataset: τ=0.40), which demonstrates that the threshold can adaptively adjust the threshold value and lead to a promising performance.
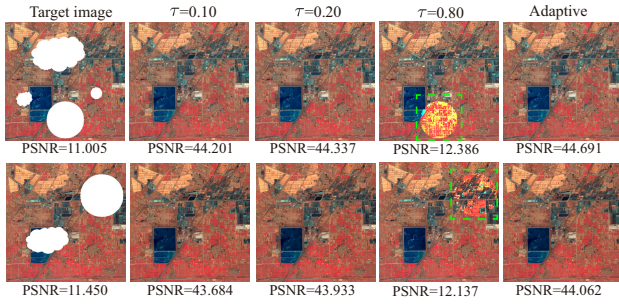
Fig. 19. Cloud/Shadow removal effect with respect to different threshold values on Dongying dataset.
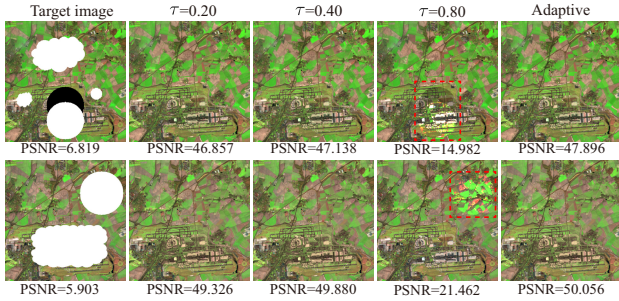

Fig. 20. Cloud/Shadow removal effect with respect to different threshold values on Picardie dataset.

2) Flexibility of the Proposed Model: We analyze the optimal rank of the reshaped matrix with respect to different temporal intervals. The multi-temporal images with short intervals (acquired times: 02/24, 02/27, 03/01, 03/06, 03/24, and 03/31/2021) and long intervals (acquired times: 05/20/2020, 05/30/2020, 09/17/2020, 09/22/2020, 02/24/2021, and 02/27/2021) are selected from Picardie dataset. We use the same mask as Cases 1-3 of Picardie dataset to generate clouds/shadows. The r (dimension of the orthogonal basis) is set as 5. Fig. 21 shows the changes of PSNR values with respect to different ranks under different temporal intervals. For the images with short temporal intervals, from Fig. 21 (a), we can observe that the ideal rank is r, and the PSNR values are stable when the rank is between r and $3r$. For the images with long temporal intervals, from Fig. 21 (b), we can observe that the ideal rank is $2r$, and the PSNR values are stable when the rank is between $2r$ and $3r$. For the different ideal ranks corresponding to different temporal intervals, our model can flexibly capture the redundancy of abundances with ground cover changes by minimizing the nuclear norm in varying degrees.
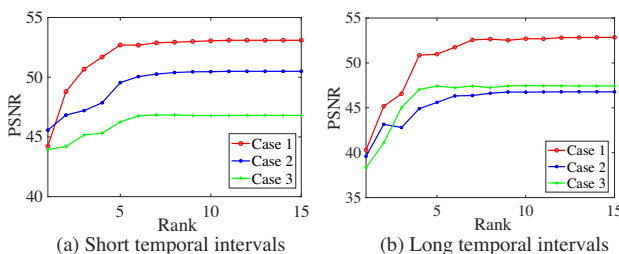

Fig. 21. Reconstruction performance with respect to the rank of the reshaped matrix of abundance tensors under different temporal intervals.
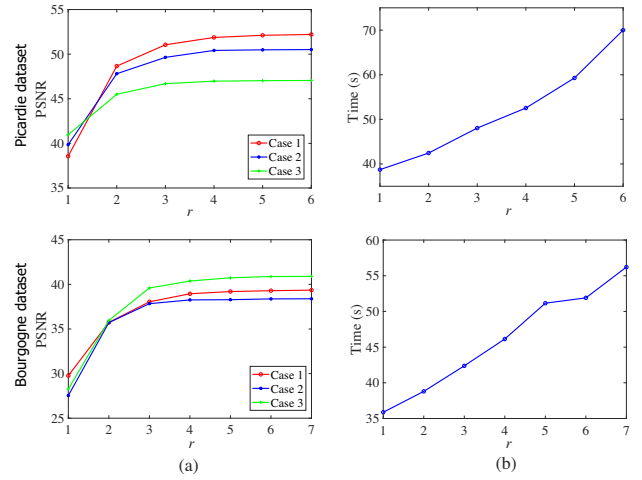

Fig. 22. Influence of the dimension of the orthogonal basis on the performance of cloud removal on different datasets. (a) PSNR versus the dimension r; (b) Running time versus the dimension r.

3) Influence of the Dimension of the Orthogonal Basis: The size of orthogonal basis $\mathbf{F}_i$ is $b \times r$ in the proposed method. We analyze the influence of the dimension r of the orthogonal basis on the performance of cloud removal. Picardie dataset and Bourgogne dataset with complex scenarios (the ground truth distributions including urban scenes, lakes, rivers, farmlands, airports, etc.) are selected. Fig. 22 shows the changes in the PSNR values and time (in seconds) with different dimensions. From Fig. 22 (a) and (b), we can observe that the PSNR values remain at a high level when the dimensions are larger than 4 and the running time is increasing with the increase of the dimension. Thus, we set the size of the basis to be $b \times r$ $(r < b)$ instead of $b \times b$ for balancing efficiency and effectiveness.

4) Influence of the Temporal Number: We test the influence of the temporal number on the reconstruction performance of the proposed method. For an objective evaluation, we keep the three target images in Cases 1-3 of Picardie dataset fixed and sequentially introduce six reference images (acquired times: 03/31/2021, 03/06/2021, 02/27/2021, 03/29/2021, 05/20/2020, and 09/17/2020) to change the temporal number. Table IV displays the PSNR values of reconstructed results of target images with different temporal numbers of reference images. It can be observed that as the temporal number increases, the PSNR values of all target images increase.

TABLE IV
PSNR (DB) VERSUS DIFFERENT TEMPORAL NUMBER OF REFERENCE IMAGES ON PICARDIE DATASET.

Target image	Temporal number of reference images					
	1	2	3	4	5	6
03/24/2021	51.79	52.294	52.206	52.881	52.913	52.952
03/01/2021	47.253	50.332	50.515	50.647	50.703	50.734
02/24/2021	45.237	46.080	47.047	47.243	47.326	47.486

5) Influence of the Cloud Cover Ratio: We test the influence of the cloud cover ratio on the reconstruction performance of the proposed method. Dongying dataset and Picardie dataset are considered for test and the simulated cloud/shadow of Case 1-2 in simulated experiments are added to the first two images.
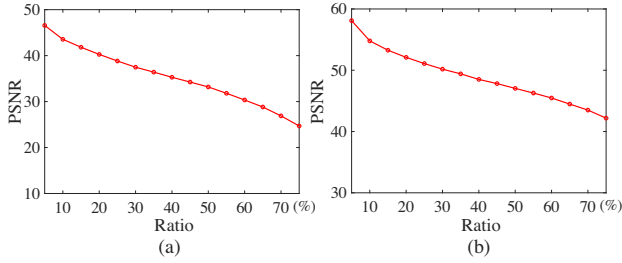
Fig. 23. Reconstruction performance with respect to different cloud cover ratio on different datasets. (a) Dongying dataset; (b) Picardie dataset.

For the third image, we change the cloud/shadow cover ratio from 5% to 75% with the increment 5%. Fig. 23 shows the changes of PSNR values with respect to different cloud cover ratios of the third image. It can be observed that the PSNR values decrease with the increase of the cloud cover ratio.

6) Parameter Analysis: There are four parameters in the proposed method, including the regularization parameters α and β and penalty parameters γ and ρ. We use the PSNR index as an evaluation indicator to analyze each parameter on the Picardie dataset and the Bourgogne dataset.

Regularization parameters α and β: The parameters α and β are used to balance the low-rank term, the sparse term, and the fidelity term. Fig. 24 (a) shows that the PSNR value changes when α is selected from the candidate set $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50\}$ and the PSNR value reach the highest value when α is 0.5. Fig. 24 (b) shows that the proposed method achieves a satisfactory result when β varies from 0.2 to 1. In the proposed method, we empirically set $\alpha = 0.5$ and $\beta = 0.5$ in all experiments.

Penalty parameters γ and ρ: The parameters γ and ρ are utilized to balance the penalty term and other terms. Fig. 24 (c) and (d) shows the PSNR value changes when ρ and β are selected from the candidate set $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50\}$, respectively. We observe that the PSNR remains at a high level when γ is selected from $\{0.005, 0.01, 0.05, 0.1, 0.5, 1, 5\}$ and the PSNR reaches the highest value when $\rho = 0.1$. Thus, we suggest the γ is selected from $\{0.005, 0.01, 0.05, 0.1, 0.5, 1, 5\}$ and set $\rho = 50$ in all experiments.

7) Convergence behavior: We numerically analyze the convergence behavior of the developed ALM algorithm. Defining the relative change (RelCha) in kth iteration as

$$\text{RelCha} = \frac{\|\mathcal{X}^k - \mathcal{X}^{k-1}\|_F}{\|\mathcal{X}^{k-1}\|_F},$$

Fig. 25 displays the curves of the RelCha value by Algorithm 2 on the Picardie dataset and Bourgogne dataset. It is clear that for different datasets, the RelCha values gradually tend to zero. This illustrates the convergence of the developed algorithm.

IV. CONCLUSION

In this paper, we have proposed a robust thick cloud/shadow removal method for multi-temporal RS images. First, a sparse term is introduced into the method to describe the cloud/shadow component and an adaptive threshold algorithm is proposed, which makes a robust cloud/shadow removal

under inaccurate given masks. Besides, a multi-temporal information reconstruction model is suggested. Based on our key observation, the model uses a coupled tensor decomposition to reproduce the latent multi-temporal relationship, achieving faithful reconstruction of underlying information. It is worth noting that the mask refinement and information reconstruction are complementary and collaborative for cloud removal since the refined mask will help to update signatures, and the better-reconstructed results will benefit the cloud/shadow components modeling. Finally, an efficient ALM-based algorithm is developed to solve the proposed model, and both simulated and real experiments on Sentinel-2 L2A dataset and Landsat-8 OLI dataset under different cloud/shadow contamination scenarios with accurate masks and inaccurate masks are conducted. The experimental results demonstrate the robustness and effectiveness of the proposed method for thick cloud/shadow removal.

APPENDIX A
PROOF OF THEOREM 1

***Theorem** 1.* If $\mathbf{F}$ is a semi-orthogonal matrix, i.e., $\mathbf{F}^T\mathbf{F} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, then

$$\arg\min_{\mathcal{A}}\|\mathcal{X} - \mathcal{A} \times_3 \mathbf{F}\|_F^2 = \arg\min_{\mathcal{A}}\|\mathcal{X} \times_3 \mathbf{F}^T - \mathcal{A}\|_F^2. \quad (13)$$

Proof. According to [47], we have $\mathcal{A} \times_3 \mathbf{F} = \text{fold}_3(\mathbf{F}\mathbf{A}_{(3)})$. Thus, formula (13) is equivalent to

$$\arg\min_{\mathbf{A}_{(3)}}\|\mathbf{X}_{(3)} - \mathbf{F}\mathbf{A}_{(3)}\|_F^2 = \arg\min_{\mathbf{A}_{(3)}}\|\mathbf{F}^T\mathbf{X}_{(3)} - \mathbf{A}_{(3)}\|_F^2. \quad (14)$$

From the semi-orthogonality of $\mathbf{F}$, i.e., $\mathbf{F}^T\mathbf{F} = \mathbf{I}$, we get

$$\|\mathbf{F}\mathbf{A}_{(3)}\|_F^2 = \text{Trace}\left[(\mathbf{F}\mathbf{A}_{(3)})^T(\mathbf{F}\mathbf{A}_{(3)})\right] = (\mathbf{A}_{(3)}^T\mathbf{A}_{(3)}) = \|\mathbf{A}_{(3)}\|_F^2.$$

Then, the derivation of (14) is as follows:

$$\arg\min_{\mathbf{A}_{(3)}}\|\mathbf{X}_{(3)} - \mathbf{F}\mathbf{A}_{(3)}\|_F^2$$
$$= \arg\min_{\mathbf{A}_{(3)}}\|\mathbf{X}_{(3)}\|_F^2 - 2\left\langle\mathbf{X}_{(3)}, \mathbf{F}\mathbf{A}_{(3)}\right\rangle + \|\mathbf{F}\mathbf{A}_{(3)}\|_F^2$$
$$= \arg\min_{\mathbf{A}_{(3)}}\|\mathbf{F}^T\mathbf{X}_{(3)}\|_F^2 - 2\left\langle\mathbf{F}^T\mathbf{X}_{(3)}, \mathbf{A}_{(3)}\right\rangle + \|\mathbf{A}_{(3)}\|_F^2$$
$$= \arg\min_{\mathbf{A}_{(3)}}\|\mathbf{F}^T\mathbf{X}_{(3)} - \mathbf{A}_{(3)}\|_F^2.$$

The proof is completed.

REFERENCES

[1] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Transactions on Geoscience and Remote Sensing*, 2021, doi: 10.1109/TGRS.2020.3046038.

[2] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang, "Non-local meets global: An integrated paradigm for hyperspectral image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, doi: 10.1109/TPAMI.2020.3027563.

[3] H. Zhang, J. Cai, W. He, H. Shen, and L. Zhang, "Double low-rank matrix decomposition for hyperspectral image denoising and destriping," *IEEE Transactions on Geoscience and Remote Sensing*, 2021, doi: 10.1109/TGRS.2021.3061148.

[4] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205–1218, 2019.
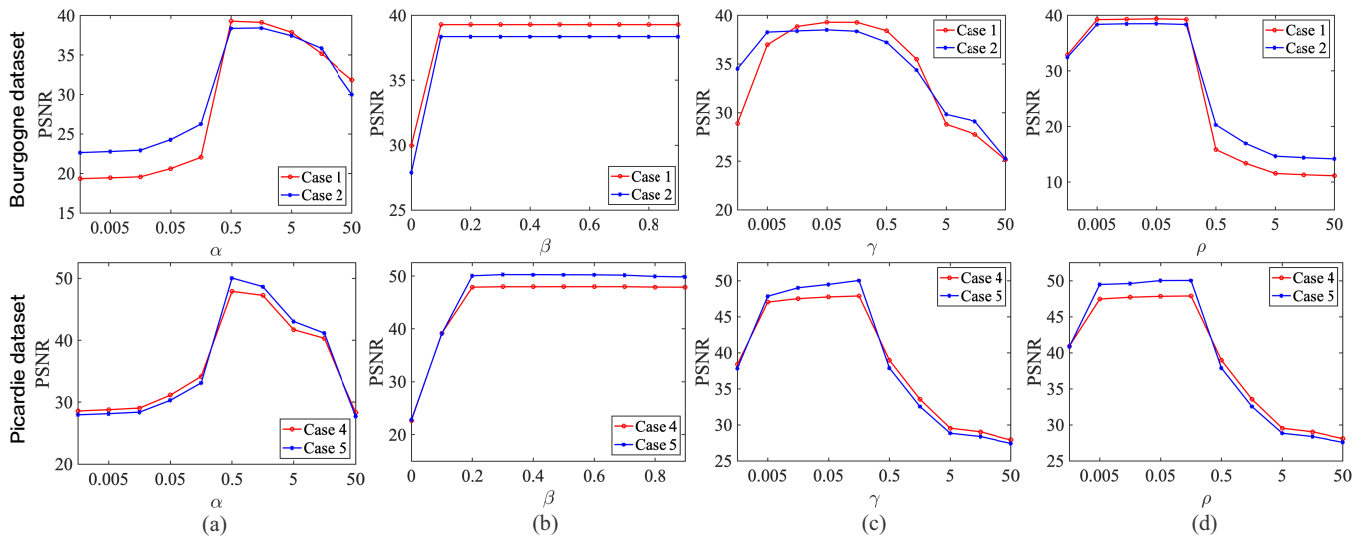
Fig. 24. Sensitivity analysis of the parameters on Bourgogne dataset and Picardie dataset. (a) PSNR versus regularization parameter α; (b) PSNR versus regularization parameter β; (c) PSNR versus penalty parameter γ; (d) PSNR versus penalty parameter ρ.
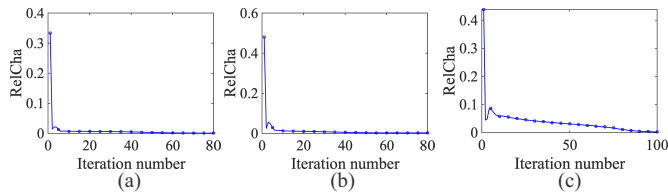

Fig. 25. RelCha versus iteration number in different cases of two datasets. (a) Cases 1-3 of Bourgogne dataset; (b) Cases 4-5 of Picardie dataset; (c) Real experiments of Bourgogne dataset.

[5] Q. Zhang, Q. Yuan, J. Li, Y. Wang, F. Sun, and L. Zhang, "Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019," *Earth System Science Data*, vol. 13, no. 3, pp. 1385–1401, 2021.

[6] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4273–4286, 2021.

[7] C. Kwan, L. Hagen, B. Chou, D. Perez, J. Li, Y. Shen, and K. Koperski, "Simple and effective cloud- and shadow-detection algorithms for landsat and worldview images," *Signal, Image and Video Processing*, vol. 14, no. 1, pp. 125–133, 2020.

[8] J. Wang, P. A. Olsen, A. R. Conn, and A. C. Lozano, "Removing clouds and recovering ground observations in satellite image sequences via temporally contiguous robust matrix completion," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2754–2763.

[9] 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 Journal of Photogrammetry and Remote Sensing*, vol. 157, pp. 93–107, 2019.

[10] X. Li, H. Shen, L. Zhang, and H. Li, "Sparse-based reconstruction of missing information in remote sensing images from spectral/temporal complementary information," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 106, pp. 1–15, 2015.

[11] H. Xu, X. Tang, B. Ai, X. Gao, F. Yang, and Z. Wen, "Missing data reconstruction in vhr images based on progressive structure prediction and texture generation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 266–277, 2021.

[12] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, "Framelet representation of tensor nuclear norm for third-order tensor completion," *IEEE Transactions on Image Processing*, vol. 29, pp. 7233–7244, 2020, doi: 10.1109/TIP.2020.3000349.

[13] A. C. Siravenha, D. Sousa, A. Bispo, and E. Pelaes, "Evaluating inpainting methods to the satellite images clouds and shadows removing," in *Signal Processing, Image Processing and Pattern Recognition*, T.-h. Kim, H. Adeli, C. Ramos, and B.-H. Kang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 56–65.

[14] C. Yu, L. Chen, L. Su, M. Fan, and S. Li, "Kriging interpolation method and its application in retrieval of modis aerosol optical depth," in *International Conference on Geoinformatics*, 2011, pp. 1–6.

[15] X. Zhu, F. Gao, D. Liu, and J. Chen, "A modified neighborhood similar pixel interpolator approach for removing thick clouds in landsat images," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 521–525, 2012.

[16] H. Shen and L. Zhang, "A map-based algorithm for destriping and inpainting of remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1492–1502, 2009.

[17] A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2634–2645, 2010.

[18] Q. Cheng, H. Shen, L. Zhang, and P. Li, "Inpainting for remotely sensed images with a multichannel nonlocal total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 175–187, 2014.

[19] K. He and J. Sun, "Image completion approaches using the statistics of similar patches," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2423–2435, 2014.

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

[21] M. Xu, X. Jia, M. Pickering, and S. Jia, "Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 149, pp. 215–225, 2019.

[22] H. Shen, X. Li, L. Zhang, D. Tao, and C. Zeng, "Compressed sensing-based inpainting of aqua moderate resolution imaging

spectroradiometer band 6 using adaptive spectrum-weighted sparse bayesian dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 894–906, 2014.

[23] L. Wang, J. J. Qu, X. Xiong, H. Hao, Y. Xie, and N. Che, "A new method for retrieving band 6 of aqua modis," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 2, pp. 267–270, 2006.

[24] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Restoration of aqua modis band 6 using histogram matching and local least squares fitting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 2, pp. 613–627, 2009.

[25] X. Li, H. Shen, L. Zhang, H. Zhang, and Q. Yuan, "Dead pixel completion of aqua modis band 6 using a robust m-estimator multiregression," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 768–772, 2014.

[26] 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 Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, 2015.

[27] M. Dao, C. Kwan, B. Ayhan, and T. D. Tran, "Burn scar detection using cloudy modis images via low-rank and sparsity-based models," in *IEEE Global Conference on Signal and Information Processing*, 2016, pp. 177–181.

[28] X. Li, H. Shen, H. Li, and L. Zhang, "Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 8, pp. 3629–3641, 2016.

[29] 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 Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 161–173, 2021.

[30] S. Benabdelkader and F. Melgani, "Contextual spatiospectral postreconstruction of cloud-contaminated images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 204–208, 2008.

[31] F. Melgani, "Contextual reconstruction of cloud-contaminated multitemporal multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 442–455, 2006.

[32] Q. Cheng, H. Shen, L. Zhang, Q. Yuan, and C. Zeng, "Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 92, pp. 54–68, 2014.

[33] C.-H. Lin, K.-H. Lai, Z.-B. Chen, and J.-Y. Chen, "Patch-based information reconstruction of cloud-contaminated multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 163–174, 2014.

[34] T. Shuai, X. Zhang, S. Wang, L. Zhang, K. Shang, X. Chen, and J. Wang, "A spectral angle distance-weighting reconstruction method for filled pixels of the modis land surface temperature product," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 9, pp. 1514–1518, 2014.

[35] 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 Sensing of Environment*, vol. 131, pp. 182–194, 2013.

[36] M. K. Ng, Q. Yuan, L. Yan, and J. Sun, "An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3367–3381, 2017.

[37] T.-Y. Ji, N. Yokoya, X. X. Zhu, and T.-Z. Huang, "Nonlocal tensor completion for multitemporal remotely sensed images' inpainting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3047–3061, 2018.

[38] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56,

[39] Q. Zhang, Q. Yuan, J. Li, Z. Li, H. Shen, and L. Zhang, "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 148–160, 2020.

[40] D.-C. Tseng, H.-T. Tseng, and C.-L. Chien, "Automatic cloud removal from multi-temporal SPOT images," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 584–600, 2008.

[41] C.-H. Lin, P.-H. Tsai, K.-H. Lai, and J.-Y. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 232–241, 2013.

[42] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, 2020.

[43] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 103 – 113, 2019.

[44] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 730–742, 2018.

[45] F. D. Van der Meer and X. Jia, "Collinearity and orthogonality of endmembers in linear spectral unmixing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 491–503, 2012.

[46] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. Zhu, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 2, pp. 52–87, 2021.

[47] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[48] J. Lin, T. Z. Huang, X. L. Zhao, T. X. Jiang, and L. Zhuang, "A tensor subspace representation-based method for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, 2020, doi: 10.1109/TGRS.2020.3032168.

[49] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

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

[51] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

[52] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[53] X. Li, H. Shen, L. Zhang, H. Zhang, Q. Yuan, and G. Yang, "Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7086–7098, 2014.

[54] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 197–212, 2019.

Jie Lin received the B.S. degree in information and computing science from Anhui University of Finance and Economics, Bengbu, China, in 2018. He is currently working toward the Ph.D. degree in the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

His current research interests include low-rank and sparse modeling, tensor decomposition, and high-dimensional image processing.

Qiang Zhang received the B.S. degree in surveying and mapping engineering and M.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree in LIESMARS, Wuhan University, Wuhan, China.

His research interests include remote sensing image quality improvement, multi-source data fusion, deep learning and computer vision. More details about his work can be found at https://qzhang95.github.io/.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China. He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

His current research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Dr. Huang is an Editor of the Scientific World Journal, Advances in Numerical Analysis, the Journal of Applied Mathematics, the Journal of Pure and Applied Mathematics: Advances in Applied Mathematics, and the Journal of Electronic Science and Technology, China.

Xi-Le Zhao received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012. He worked as a post-doc with Prof. M. Ng at Hong Kong Baptist University from 2013 to 2014. He worked as a visiting scholar with Prof. J. Bioucas Dias at University of Lisbon from 2016 to 2017. He is currently a Professor with the School of Mathematical Sciences, UESTC.

His research interests include sparse and low-rank modeling for the low-level inverse problems of multi-dimensional images. More information can be found on his homepage https://zhaoxile.github.io/.

Qiangqiang Yuan (Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is currently a Professor. He published more than 90 research papers. His current research interests include image reconstruction, remote sensing image processing and application, and data fusion.

Yong Chen received the B.S. degree in School of Science from East China University of Technology, Nanchang, China, in 2015, and the Ph.D. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020. He is currently working with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China. From 2018 to 2019, he was a research intern with the Geoinformatics unit, RIKEN Center for Advanced Intelligence Project, Japan.

His research interests include remote sensing image processing and low-rank matrix/tensor representation.