A Coarse-to-Fine Framework for Cloud Removal in Remote Sensing Image Sequence

Yongjun Zhang, Fei Wen, Zhi Gao, and Xiao Ling

***Abstract*— Clouds and accompanying shadows, which exist in optical remote sensing images with high possibility, can degrade or even completely occlude certain ground-cover information in images, limiting their applicabilities for Earth observation, change detection, or land-cover classification. In this paper, we aim to deal with cloud contamination problems with the objective of generating cloud-removed remote sensing images. Inspired by low-rank representation together with sparsity constraints, we propose a coarse-to-fine framework for cloud removal in the remote sensing image sequence. Leveraging on group-sparsity constraint, we first decompose the observed cloud image sequence of the same area into the low-rank component, group-sparse outliers, and sparse noise, corresponding to cloud-free land-covers, clouds (and accompanying shadows), and noise respectively. Subsequently, a discriminative robust principal component analysis (RPCA) algorithm is utilized to assign aggressive penalizing weights to the initially detected cloud pixels to facilitate cloud removal and scene restoration. Moreover, we incorporate geometrical transformation into a low-rank model to address the misalignment of the image sequence. Significantly superior to conventional cloud-removal methods, neither cloud-free reference image(s) nor additional operations of cloud and shadow detection are required in our method. Extensive experiments on both simulated data and real data demonstrate that our method works effectively, outperforming many state-of-the-art approaches.**

***Index Terms*— Cloud and shadow removal, group-sparse, low-rank representation, robust principal component analysis (RPCA).**

I. Introduction

REMOTE sensing images have been applied in a variety of applications, including Earth observation, change detection, land-cover classification, and so on. Due to the proliferation of satellites, such trend is continuously intensifying. However, remote sensing images can be contaminated by clouds and accompanying shadows with high possibility. For example, the Enhanced Thematic Mapper Plus (ETM+) land scenes are reported to be about 35% cloud covered globally [1]. Ground-cover information is degraded by thin clouds and shadows or even completely occluded by thick clouds, which remarkably limits further analysis and applications of such images (see Fig. 1 for some examples). Therefore, removing clouds and their shadows is of great importance to facilitate the utilization of such contaminated images. In particular, the effect of clouds varies according to the thickness. Thin clouds allow part of underlying objects being observed, which are often ambiguous and could be fairly subtle to formulate and solve such cloud associated problems. On the other hand, thick clouds allow no ground-cover information being observed, thus solutions are required urgently to overcome such a challenging problem. Despite a substantial amount of efforts in this direction, removing cloud contamination effectively for a batch of temporal images remains an open problem. Therefore, in this paper, we focus on the case of thick clouds and accompanying shadows in the remote sensing image sequence.

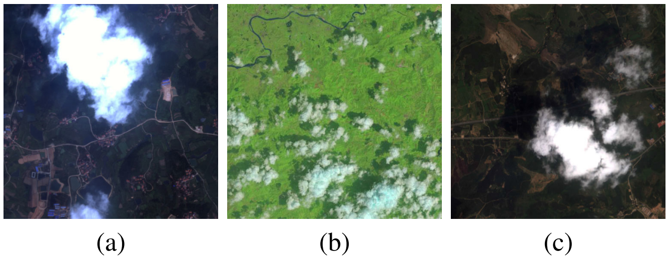


Fig. 1. Cloud-contaminated satellite images. (a) GF-2 true color data. (b) Landsat-8 natural-look data. (c) ZY-3 true color data.

Currently, the available methods of cloud removal can be roughly classified into two categories [2]: individual-based [3]–[5] and multitemporal-based methods [2], [6]–[14]. We will detail these works in Section II. In a nutshell, although these methods can remove cloud contamination and generate visually plausible results, their results are sensitive to the size of the cloud and their efficiencies are quite low. Moreover, extensive cloud-free reference images are usually assumed to be available and accurately preregistered to the cloud contaminated image(s), which is difficult, if not impossible, to fulfill in practice. With the development of satellite technologies and the easier access to their data, it has been feasible to obtain a sequence of satellite images located in the same position. Therefore, methods based on batch processing,

Manuscript received June 30, 2018; revised October 19, 2018 and February 21, 2019; accepted March 4, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0505003 and in part by the National Natural Science Foundation of China under Grant 41571434. *(Corresponding author: Yongjun Zhang.)*

Y. Zhang and F. Wen are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zhangyj@whu.edu.cn; wenfei@whu.edu.cn).

Z. Gao is with the Temasek Laboratories, National University of Singapore, Singapore 117411 (e-mail: gaozhinus@gmail.com).

X. Ling is with the Future Cites Laboratory, Singapore-ETH Center, Singapore (e-mail: ling.xiao@arch.ethz.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2903594

which claim to press the maximal benefits from multitemporal correlations, have been reported with promising results. In our previous work [15], a two-pass robust principal component analysis (TRPCA) method was proposed for cloud removal in a satellite image sequence, which was significantly superior to other methods, neither cloud-free reference images nor specific algorithms of cloud detection [16], [17] were required. Moreover, it has demonstrated to achieve better accuracy and significant efficiency improvement compared to information clone [8] and sparse representation methods [13]. However, the first pass of TRPCA applied a plain RPCA followed by morphological operation without taking into account the cluster property of clouds and shadows. In addition, it required that the image sequence should be accurately prealigned to enforce the low-rank constraint.

On the basis of [15], here, we propose a coarse-to-fine framework leveraging on group-sparsity for cloud removal in the satellite image sequence. Considering the fact that clouds and shadows are typically spatially coherent, group-structured sparsity is formulated to better model sparse outlier clusters. Moreover, adaptive weights are assigned to such groups to facilitate convergence. Benefiting from the group-sparsity constraint, we obtain satisfied initial masks of clouds and shadows without any postprocessing. In addition, we incorporate geometrical transformation into RPCA to refine the alignment of the image sequence. In other words, our method no longer requires that the input images of a sequence are accurately aligned in advance. In summary, our newly proposed method outperforms [15] in terms of both accuracy and efficiency, and with wider applicability as well.

The remainder of this paper is organized as follows. Section II discusses related works. Section III is devoted to the details of this paper. Section IV presents our extensive experiments, and the conclusion is summarized in Section V.

II. Related Works

A. Individual-Based Methods

Assuming that the remaining cloud-free regions have similar texture features as cloud contaminated regions, individual-based methods typically deal with cloud contamination in a single image without additional auxiliary information. Such reconstruction is also called inpainting that synthesizes cloud-contaminated regions via propagating from local or nonlocal cloud-free pixels. For example, Chen *et al.* [3] improved a fragment-based image completion algorithm [18] to remove clouds and shadows in high-resolution remote sensing images. It iteratively chose small image fragments of fixed size in the clear parts guided by their confidence map and duplicated it into the contaminated regions following a coarse-to-fine multiscale strategy. Leveraging on the geometric flow curves estimated by Bandelet transformation, Maalouf *et al.* [4] propagated the geometrical information into cloud-contaminated areas for reconstruction. In [5], three different strategies were utilized to facilitate patch search for more accurate propagation. These individual-based methods can yield visually plausible reconstruction results in some cases. However, they are sensitive to the land-cover types underneath clouds and the size of the contaminated area due to the inherited limitation of inpainting. Furthermore, as uncertainty and error accumulate along with propagation, the individual-based methods can hardly deal with thick cloud of large size.

B. Multitemporal-Based Methods

More relevant to our work, multitemporal-based methods that take advantage of other temporal remote-sensed images have been more popularly investigated. As discussed in [15], multitemporal information can be utilized either explicitly or implicitly, depending on whether the correlation between the contaminated regions and supplementary temporal information is explicitly formulated or implicitly learned. In an explicit manner, assuming that the differences between the cloud image and reference images are small, Tseng *et al.* [6] directly replaced cloud-contaminated pixels with the data of the same location from other cloud-free images. By applying color matching and multiscale wavelet fusion, those seam effects around cloud region boundaries can be properly eliminated. Inspired by Poisson image editing [7], cloud-free patches were cloned to their corresponding cloud regions by solving a group of constrained Poisson equations in [8]. Thanks to the boundary constraint and gradient propagation, such patch clone method can generate plausible results. To further exploit the correlations between cloud regions in the target image and cloud-free regions in both target and supplementary cloud-free images, Cheng *et al.* [9] utilized Markov Random Field (MRF) to locate local or nonlocal similar pixels in the remaining cloud-free region of the target image. It is similar to inpainting-based methods [5], but the strategy of similar pixel estimation is more reliable with the guide of multitemporal images. Moreover, Zhu *et al.* [10] improved a neighborhood similar pixel interpolator (NSPI) approach in [19] to predict cloud-contaminated pixels considering both spectral–spatial and spectral–temporal information. Similarly, Chen *et al.* [11] formulated a linear least-square regression model to search candidate pixels spatially and temporally and applied a weighted regression for the final reconstruction of cloud areas. With the aid of using similar pixel information from the same contaminated image when reconstructing cloud pixels, the above three similar-pixel-based methods can handle radiometric differences and seasonal changes of multitemporal images to some extent.

Recently, a number of learning-based methods have been introduced to remove clouds in remote sensing images. Lorenzi *et al.* [12] assumed that pixels in the cloud-contaminated region can be expressed as a linear combination of sampled pixels in the remaining cloud-free region, thus the problem was formulated and solved using sparse representation. Analogously, in [2], dictionary learning was performed on target cloud image and reference image separately in the spectral domain. Then, cloud removal was conducted by combining dictionary learned from the target image and coefficients learned from the reference image. However, such methods were sensitive to land cover type and cloud size. Utilizing the local temporal correlations and nonlocal spatial correlations, Li *et al.* [13] introduced a patch-matching-based multitemporal

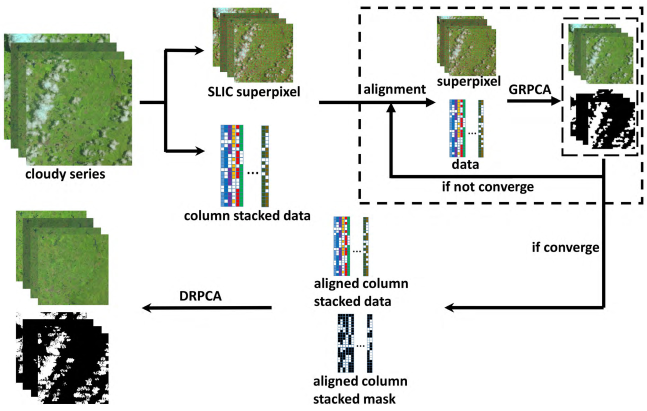


Fig. 2. Flowchart of the proposed coarse-to-fine framework for cloud removal in remote sensing image sequence.

group sparse representation (PM-MTGSR) method to recover cloud regions. Although sparse representation methods can obtain consistent reconstruction results, they show a slight blur effect because of their representation error. Inspired by deep learning technique, Zhang *et al.* [14] proposed a unified spatial–temporal–spectral framework based on deep convolutional neural network (CNN) to reconstruct missing information caused by sensor failure and remove thick cloud in remote sensing images as well. However, it required large amount of training data set and was sensitive to the properties of data.

In summary, all available methods essentially recover only one target cloud image at each time, no matter how the relationship between contaminated pixels and cloud-free pixels is exploited. Though visually plausible recovery results can be generated by these methods, they are sensitive to cloud size and inefficient to process image sequence. Hence, we propose a batch-processing approach based on RPCA framework to remove cloud from image sequence with high efficiency and accuracy. In addition, our method is able to deal with the registration error challenge faced by all available methods. We introduce a 2-D affine transformation model to enable our method to handle misaligned images of a sequence.

III. Methodology

Fig. 2 shows an overview of our coarse-to-fine framework. The input image sequence of the same area obtained at different times can be misaligned. First, simple linear iterative clustering (SLIC) superpixel segmentation and arranging each image to a column of a matrix are conducted as preprocessing. Then, group-sparsity constrained RPCA (GRPCA) combined with geometrical transformation is applied to detect cloud and shadow regions initially and also generate a well-aligned image sequence. The dotted box denotes our extension based on group sparsity to align the misaligned image sequence. Finally, discriminative RPCA (DRPCA) is conducted to remove clouds and shadows to obtain a sequence of cloud removed images.

A. Overview of RPCA

RPCA [20] has obtained stunning performance in a variety of applications including target detection, anomaly detection, and so on. Leveraged on the intrinsic low dimensionality of massive multidimensional data such as video and images, RPCA assumes such data are composed of a low-rank component and a sparse component. Mathematically, given a sequence of images or frames of a video and arranging them as the columns of a large matrix $M \in R^{m\times n}$, then M can be decomposed into a low-rank matrix L and a sparse matrix S, which is estimated by minimizing the following constrained optimization problem:

$$\begin{aligned}\min_{L,S}\quad & |L\|_* + \lambda\|S\|_1\\ \text{s.t.}\quad & M = L + S\end{aligned}\tag{1}$$

as a surrogate for the original problem

$$\begin{aligned}\min_{L,S}\quad & \operatorname{rank}(L) + \lambda\|S\|_0\\ \text{s.t.}\quad & M = L + S\end{aligned}\tag{2}$$

where $\|L\|_*$ denotes the nuclear norm of matrix L, i.e., the sum of its singular values, $\|S\|_0$ denotes the number of nonzero elements in the matrix, and $\|S\|_1$ denotes the sum of the absolute value of each element of matrix S, and λ is a positive balance value.

As an example in background and foreground separation [21], low-rank L corresponds to the background and sparse S contains foreground moving objects. Similarly, when dealing with satellite sequence of cloud images, the desirable clear ground-cover can be treated as background, whereas the clouds and accompanying shadows are treated as sparse foreground. Modeling the background by low-rank approximation is known to be able to absorb the global illumination changes. However, the sparsity prior with l_1-norm regularization treats each pixel independently, it ignores the possible structure or relations between pixels [22]. In practice, the foreground objects are usually spatially coherent clusters and such spatial correlations should be incorporated to facilitate detection.

B. Superpixel Group-Structured Sparsity

Inspired by recent studies of structured sparsity in computer vision [23], [24], we introduce a nonoverlapping group-sparsity norm that can incorporate prior structures on spatial coherent outliers. Given an observed matrix $M \in R^{m\times n}$ and $M = [vec(I_1)|vec(I_2)|\ldots|vec(I_n)]$, where $I_i, i = 1, 2, \ldots, n$, are input images and vec denotes stacking an image as a column, m is the number of pixels in each image, and n is the total number of images. We define the group-sparsity norm as follows:

$$\psi(S) = \sum_{j=1}^{n}\sum_{i=1}^{K} w_j^i \left\|S_{g_j^i}\right\|_\infty \tag{3}$$

where K is the number of groups in each image, $S_{g_j^i}$ denotes every single group in sparse component S, and w_j^i are the groupwise weights which will be detailed in Section III-C.

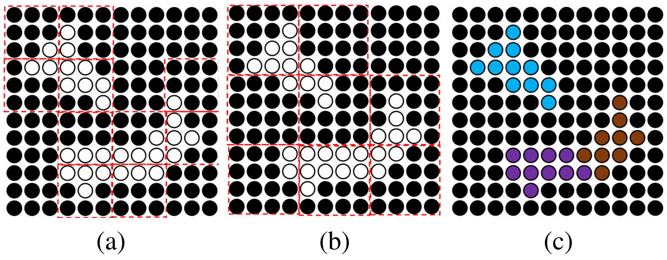


Fig. 3. Comparison of nonoverlapping groups of block-structured and superpixel structured. (a) and (b) Blocks with size of 3 and 4, respectively. (c) Superpixel groups.

$\|\cdot\|_\infty$ denotes the l_∞ norm of the vector, which is the maximum value of pixels in a group. l_∞ norm encourages the rest of variables within the same group to take arbitrary values, and that is exactly what we desire as the coherent pixels of the same object have similar magnitude.

Considering the spatial coherence of foreground pixels, a meaningful group of pixels should consider the shape and structure of objects in the image. The straightforward idea to segment pixels into groups is to cluster them into blocks as in [25]. However, the block is too restrictive to deal with a random shape in natural scenes. We show an example in Fig. 3 that, for nonoverlapping groups, blocks are either too small to encode spatial coherence prior or too large to bring in too much background pixels, which will lead to unsatisfactory detection results. To overcome such limitation, we introduce a new group structure that adapts well to objects in remote sensing images. As shown in Fig. 3, each image can be segmented into superpixels. Superpixel technique clusters pixels into perceptually meaningful regions according to their feature similarity, such as color, texture, location, and so on, which is flexible to cover random-shaped natural objects. Due to their proper approximation to the boundaries of objects, no further postprocessing is required to generate group-sparse outlier regions.

As the preprocessing step in our framework, the superpixel segmentation should be efficient and easy to use in practice. Thus, we adopt the SLIC method that has been demonstrated as the state-of-the-art superpixel method [26]. SLIC applies k-means clustering to generate superpixels with higher speed and better segmentation performance. Only two parameters of the SLIC method are needed to be set, i.e., the number of superpixels we want to obtain for an image and a compactness factor that controls adherence of each superpixel to object boundaries.

C. GRPCA With Geometrical Transformation

1) Problem Formulation: To overcome the deficiency of plain RPCA which does not consider the spatial coherence of natural objects, we introduce the group-structured sparsity into RPCA framework. Apart from group-structured sparse outliers, there are many noiselike sparse entries with an arbitrary large value in the observed matrix that are neither structured nor belong to the low-rank model. Therefore, we modify the decomposition into three parts, namely, a low-rank part and a group sparse part as usual, and an additional part of noiselike sparse outliers modeled by l_1-norm. We formulate the new GRPCA equation as

$$\begin{aligned}\min_{L,S}\quad & \|L\|_* + \lambda\psi(S) + \gamma\|N\|_1\\ \text{s.t.}\quad & D = L + S + N\end{aligned} \tag{4}$$

where $\psi(S)$ is the superpixel group-structured sparsity norm defined in (3), denoting group sparse outliers. $\|N\|_1$ denotes noiselike sparse outliers. λ and γ are the positive values controlling the sparsity of group-structured and noiselike outliers respectively.

However, the low-rank assumption of background may no longer hold if the images are not well aligned. Due to the complicate acquisition processes, satellite images acquired at different times are always misaligned to some extent. Inspired by the work in [27], we can model the alignment between satellite images as 2-D affine transformation. Suppose that $I_1, I_2, ..., I_n$ denote n input images of the same area but not well aligned, and there exist n transformations $\tau_1, \tau_2, ..., \tau_n$ such that the transformed images $I_1 \circ \tau_1, I_2 \circ \tau_2, ..., I_n \circ \tau_n$ are pixel-level aligned, where $\tau_i \in R^p, i = 1, 2, ..., n$ and $p = 6$ indicates affine transformation. Then, the matrix $D \circ \tau = [I_1 \circ \tau_1, I_2 \circ \tau_2, ..., I_n \circ \tau_n]$ has low rank, where $D = [vec(I_1)|vec(I_2)|...|vec(I_n)]$, and τ represents the set of n transformations. Therefore, the final constrained optimization problem is formulated as

$$\begin{aligned}\min_{L,S}\quad & \|L\|_* + \lambda\psi(S) + \gamma\|N\|_1\\ \text{s.t.}\quad & D\circ\tau = L + S + N.\end{aligned} \tag{5}$$

2) Iterative Optimization: We adopt Augmented Lagrange Multiplier (ALM) [28] method to solve the optimization problem of (5) since it has been widely used to solve RPCA-based problems [29], [30]. The augmented Lagrange function is defined as

$$\begin{aligned}f(L,S,N,\tau,Y,\mu) = {}& \|L\|_* + \lambda\psi(S) + \gamma\|N\|_1\\ & + \langle Y, D\circ\tau - L - S - N\rangle\\ & + \frac{\mu}{2}\|D\circ\tau - L - S - N\|_F^2\end{aligned} \tag{6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, Y is the Lagrange multiplier, and μ is a positive scalar. ALM optimizes variables L, S, N, and τ alternatively and updates Y, μ iteratively. Especially, L, S, N, and τ are estimated by solving subproblems as follows:

$$\begin{aligned}L_{k+1} &= \min_{L} f(L, S_k, N_k, \tau, Y_k, \mu_k)\\ S_{k+1} &= \min_{S} f(L_k, S, N_k, \tau, Y_k, \mu_k)\\ N_{K+1} &= \min_{N} f(L_k, S_k, N, \tau, Y_k, \mu_k)\\ \tau &= \tau + \triangle\tau\end{aligned} \tag{7}$$

and the Lagrange multiplier Y and balance value μ are updated as

$$\begin{aligned}Y_{k+1} &= Y_k + \mu_k(D\circ\tau - L_k - S_k - N_k)\\ \mu_{k+1} &= \rho\mu_k.\end{aligned} \tag{8}$$

Given the support of other variables at the kth iteration, the estimation of L_{k+1} for the first subproblem of (7) is formulated as

$$\arg\min_{L} \|L\|_* + \frac{\mu}{2}\left\|\left(D \circ \tau - S_k - N_k + \frac{1}{\mu}Y\right) - L\right\|_F^2. \quad (9)$$

According to the lemma in [31], given a matrix Z, the solution to the optimization problem

$$\min_{X} \frac{1}{2}\|Z - X\|_F^2 + \alpha\|X\|_* \quad (10)$$

is given by $X = \Theta_\alpha(Z)$, where Θ_α means the singular value thresholding

$$\Theta_\alpha(Z) = U\Sigma_\alpha V^T. \quad (11)$$

Here, $\Sigma_\alpha = diag[(d_1-\alpha)_+, \ldots, (d_r-\alpha)]$, $U\Sigma V$ is the singular value decomposition (SVD) of Z, $\Sigma = diag[d_1, \ldots, d_r]$, and $t_+ = max(t, 0)$. Therefore, the solution of (9) is $U\Sigma_{1/\mu}(D \circ \tau - S_k - N_k + 1/\mu Y)V^T$.

For the second subproblem of (7), the optimization of S can be written as

$$\min_{S} \lambda\psi(S) + \frac{\mu}{2}\left\|\left(D \circ \tau - L_k - N_k + \frac{1}{\mu}Y\right) - S\right\|_F^2. \quad (12)$$

Since the superpixel-structured groups are nonoverlapping, we can minimize (12) with respect to each group S_{g^i} separately (we remove j to help understanding). For brevity, we denote λw_j^i as ω, $S_{g_j^i}$ as s, and $(D \circ \tau - L_k - N_k + \frac{1}{\mu}Y)_j^i$ as h. Then, (12) is as the same as minimizing the problem as follows:

$$\min_{s} \frac{\mu}{2}\|s - h\|_2^2 + \omega\|s\|_\infty. \quad (13)$$

Here, we use the proximal method [32] to solve (13).

To estimate N, we solve the optimization problem as

$$\min_{N} \gamma\|N\|_1 + \frac{\mu}{2}\left\|\left(D \circ \tau - L_k - S_k + \frac{1}{\mu}Y - N\right)\right\|_F^2. \quad (14)$$

As in [28], the solution for (14) is $\mathcal{S}_{\gamma/\mu}(D \circ \tau - L_k - S_k + \frac{1}{\mu}Y)$, where $\mathcal{S}$ is a soft-thresholding operator that is defined as

$$\mathcal{S}_\omega(x) = \max(x - u, 0) + \min(x + u, 0). \quad (15)$$

The groupwise weight value w_j^i is fairly important in our nonoverlapping GRPCA method, especially for the cloud removal task. Before discussing the details, we show an intuitive comparison between the cloud pixels and cloud-free pixels. Based on the RPCA, cloud pixels are decomposed as sparse outliers. By applying a plain RPCA decomposition (the input data are normalized to $0 \sim 1$), the observed matrix is decomposed into a low-rank component and a sparse one as shown in Fig. 4. Quantitatively, we can roughly see the magnitude of different kinds of pixels in the sparse component as in Fig. 5. It is obvious that clouds and shadows pixels depart from the low-rank background (zero line shown in Fig. 5) with nearly the same magnitude and they are significantly larger than cloud-free pixels. According to the optimization problem (13), the max absolute value of s is bounded by h. Thus, the closer h is to the zero vector, such as cloud-free candidate groups, the higher possibility $\|s\|_\infty$ is to be zero with fixed ω, which acts as a shrinkage operator to force s to zeros. In addition, though cloud-free candidate groups are more likely to be set as zeros, one global balance value for all groups in the whole sequence is insufficient to segment cloud groups and cloud-free groups accurately as it always favors the most prominent features (i.e., the groups in images with large illumination variations). Therefore, we introduce a groupwise weight to punish such bias and facilitate the shrinkage to more accurately detect cloud candidate groups.

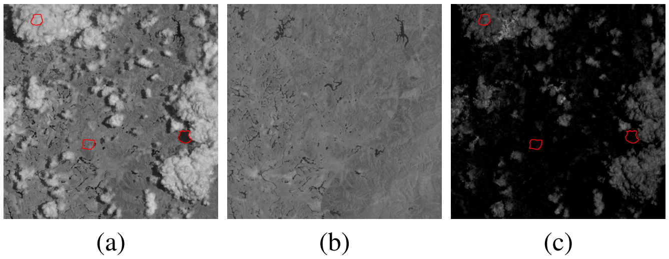


Fig. 4. Plain RPCA decomposition of cloud image. (a) Original cloud image. (b) Low-rank component. (c) Spare component. Red circular curves: sampling region of cloud, cloud-free, and shadow.

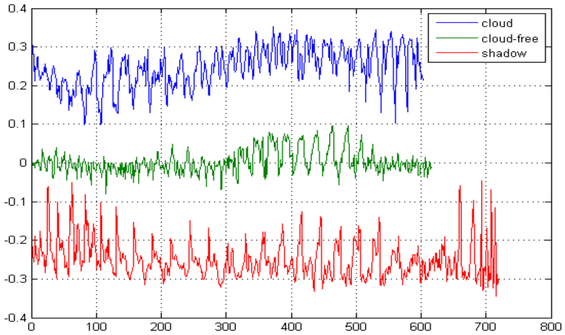


Fig. 5. Value distribution of sampled pixels in sparse component S as marked in Fig. 4.

As for minimizing (13), a larger ω diminishes $\|s\|_\infty$ with fixed h, which can also shrink the group s to zeros. λ is a balance value to a tradeoff between low-rank and sparse components, so aggressive values are set for w_j^i. We design groupwise weight w_j^i as $1/\|S_{g_j^i}\|_{\tilde{\infty}}/\|S_j\|_\infty$, and we define $\|S_{g_j^i}\|_{\tilde{\infty}}$ as a pseudo $\|\cdot\|_\infty$ norm, which is the max value of a group of variables after clipping the extreme values with a ratio number. On the one hand, cloud candidate groups are weighted with lower values because of higher $\|S_{g_j^i}\|_{\tilde{\infty}}$ and cloud-free candidate groups are weighted with higher values because of lower $\|S_{g_j^i}\|_{\tilde{\infty}}$, which will promote their separation by shrinkage. On the other hand, we normalize each group weight according to the image it belongs to by $\|S_j\|_\infty$. This, within image normalization, is crucial since the illumination of every image is different, and it will always favor the prominent groups among the whole sequence without such normalization. In addition, using a pseudo norm to generate the groupwise weight makes it more robust to noise entries with fairly large magnitude.

Now, we investigate how to update τ with respect to $\triangle\tau$. The constraint $D \circ \tau = L + S + N$ is nonlinear.

Following [27] and [33], we approximate this constraint by linearizing around the current value of τ. At each iteration, we have $D \circ \tau = D \circ \widehat{\tau} + J_{\widehat{\tau}}(\triangle\tau)$, where J_τ is the Jacobian matrix $\partial D/\partial\tau|_{\tau=\widehat{\tau}}$. Thus, τ can be updated as follows:

$$\widehat{\tau} = \widehat{\tau} + \min_{\triangle\tau} \left\| D \circ \widehat{\tau} + J_{\widehat{\tau}}(\triangle\tau) - \widehat{L} - \widehat{S} - \widehat{N} + \frac{1}{\mu} Y \right\|_F^2 . \quad (16)$$

The minimization of $\triangle\tau$ in (16) is a weighted least squares problem that has a closed-form solution. To accelerate the convergence, we initialize τ by roughly aligning each image D_j to the middle one by the multiresolution method in [34].

Finally, we summarize the whole nonoverlapping GRPCA in Algorithm 1. We initialize L, S, N to zeros and set $\mu_0 = 2/\|D\|_2$, $\rho = 1.5$ and $J(D) = max(\|D\|_2, \gamma^{-1}\|D\|_\infty)$.

Algorithm 1: Nonoverlapping Group Sparse RPCA

```
  Input: Matrix $D \in R^{m\times n}$, SLIC superpixel labels, positive
         scalar $\lambda, \gamma$;
  Output: low-rank $L$, sparse $S$ and $N$, transformation $\tau$
1 Initialize: $\widehat{\tau} = \tau_0$ (prealigned), $L_0, S_0, N_0, Y_0 = D/J(D)$, $\mu_0 > 0$; $\rho > 1$; $k = 0$.
  while not converged do
2     // Lines 4-5 solve $L_{k+1} = \arg\min_L f(L, S_k, N_k, \widehat{\tau}, Y_k, \mu_k)$ï¼›
      $(U, \Lambda, V) = svd(D \circ \widehat{\tau} - S_k - N_k + \mu_k^{-1} Y_k)$;
      $L_{k+1} = U S_{\mu_k^{-1}}[\Lambda] V^T$;
      // Line 7 solves $S_{k+1} = \arg\min_S f(L_{k+1}, S, N_k, \widehat{\tau}, Y_k, \mu_k)$;
      $S_{k+1} = prox(S_g)$;
      // Line 9 solves $N_{k+1} = \arg\min_N f(L_{k+1}, S_{k+1}, N, \widehat{\tau}, Y_k, \mu_k)$;
      $N_{k=1} = S_{\gamma/\mu_k}(D \circ \widehat{\tau} - L_{k+1} - S_{k+1} + \mu_k^{-1} Y_k)$ // Line 10 and 11 updates $\widehat{\tau}$;
      $\triangle\tau = \min_{\triangle\tau} \|D \circ \widehat{\tau} + J_{\widehat{\tau}}(\triangle\tau) - \widehat{L} - \widehat{S} - \widehat{N} + \mu_k^{-1} Y_k\|_F^2$;
      $\widehat{\tau} = \widehat{\tau} + \triangle\tau$;
      $Y_{k+1} = Y_k + \mu_k(D \circ \widehat{\tau} - L_{k+1} - S_{k+1} - N_{k+1})$;
      $\mu_{k+1} = \rho\mu_k$;
      $k \leftarrow k + 1$;
3 end
4 Output: $L, S, N, \widehat{\tau}$
```

D. Discriminative RPCA for Cloud Removal

The low-rank component generated in the first GRPCA step is far from the cloud-free reconstructed result that we expect. It is either too smooth or ghostly with a single balance value λ as discussed in [15]. For the purpose of reconstructing cloud contaminated images, we hope to recover pixels in cloud and shadow region while maintaining original cloud-free pixels at the same time. Therefore, as proposed in [15], we assign different balance values for cloud and shadow pixels and cloud-free pixels according to the initial region obtained in the first GRPCA step, which we name it as DRPCA. Within the cloud-covered region, a lower balance value ensures that all cloud- and shadow-polluted pixels will be thoroughly decomposed into sparse outlier matrix without leaving any ghostly presence in the background, yet not incurring a large false positive rate. On the other hand, for the cloud-free region, we set a relatively large value to guarantee original cloud-free information maintenance.

We present a concise review of DRPCA algorithm here as a matter of convenience. In the reconstruction step, the new formulation is defined as

$$\begin{aligned} \min_{L,S} \quad & \|L\|_* + \alpha\|P_\Omega(S)\|_1 + \beta\|P_{\Omega^-}(S)\|_1 \\ \text{s.t.} \quad & \widehat{D} = L + S \end{aligned} \quad (17)$$

where $\widehat{D}$ denotes observed matrix after transformation and $\widehat{D} = \{vec(I_1 \circ \tau_1), vec(I_2 \circ \tau_2), ..., vec(I_n \circ \tau_n)\}$, Ω is the cloud and shadow mask obtained in the first step, Ω^- denotes cloud-free region, and α and β are the two discriminative balance values. Similar to (4), (17) remains a constrained convex optimization problem, and we apply inexact ALM to solve it. For more details, readers can refer to our previous work [15].

IV. Experiments and Discussion

A. Experiment Settings

Our experiments were conducted on simulated and real cloud images. The simulation experiments were comprised of two parts, aiming to demonstrate our method in handling misaligned image sequence and compare reconstruction accuracy with other methods, respectively. The real image experiments aimed to specify the improvement related to group sparsity in initial cloud region detection. To demonstrate the performance of the proposed coarse-to-fine framework, we chose Landsat-8 Operational Land Imager (OLI) and Sentinel-2 scenes to form our data sets due to their convenient access. All the Landsat-8 OLI images used in our experiments are Landsat-8 natural-look products that are compressed and stretched to create an optimization for image selection and visual interpretation. The Landsat-8 natural-look color image is composed of three bands (bands 4, 5, and 6)[1] and the reflectance values are scaled to 1 ~ 255 range using a gamma stretch with a gamma of 2. The stretch is designed to emphasize vegetation without clipping the extreme values. The Sentinel-2 images are top of atmosphere (TOA) products downloaded from Google Earth Engine, and we use false color Sentinel-2 images (composed of bands 3, 4, and 8)[2] in our experiments.

In all our simulation experiments, only Landsat-8 images were used. The Landsat-8 OLI image products in Tier-1 level are well registered to subpixel level. To simulate a misaligned image sequence, we randomized six affine transformation parameters and warped each image to a new projected frame except for the middle image in the sequence. The images in each sequence were cropped from original Landsat-8 scenes with the size of 518 × 518 pixels. In the experiments of reconstruction comparison, for the reason that real clouds and

[1] The spatial resolution of Landsat-8 bands 4, 5, and 6 is 30 m and their wavelengths are 0.636–0.673, 0.851–0.879, and 1.566–1.651μm, respectively. For more details, refer to the website (https://Landsat.usgs.gov/).

[2] The spatial resolution of Sentinel-2 bands 3, 4, and 8 is 10 m and their wavelengths are 560, 665, and 842 nm.

shadows are fairly hard to simulate, we alternatively drew clouds in each image with pure white shapes and ignored the effect of cloud shadows. As for real image tests, we cropped block from real cloud satellite images to form cloud image sequence, and the cloud contamination rate of each block image ranged from 1% to 50%. In addition, in order to test the robustness of our method, real image experiments were conducted on three Landsat-8 sites with different land-cover types and two Sentinel-2 sites with respect to urban and mountain area. The sequence simulated for misalignment experiment was composed of 28 frames. The numbers of images in three simulated sequences in reconstruction accuracy comparison were 28, 22, and 23, respectively. In real image experiments, the three Landsat-8 sequences were composed of 30, 32, and 28 images, and the numbers of two Sentinel-2 sequences were 32 and 17, respectively. The time interval between images in the same sequence ranges from one cycle[3] to several cycles.

As for parameters setting of our method, we first set two constant parameters of the SLIC method at the preprocessing step. The desired number of superpixels for an image is set to the minimum value of its rows and columns, and the compactness factor is set to 30. At coarse detection step, the balance values, λ and γ , were set to 0.2 and $1/\sqrt{\max(m,n)}$, respectively. The stop criterion of Algorithm 1 is $\|D \circ \tau - L - S - N\|_F/\|D\|_F < 10^{-5}$. At the final reconstruction step, all parameters were set as the same as those in [15]. More specifically, in order to make the proposed method robust to noiselike large magnitude outliers, we set a ratio value to 0.1 to select pseudo max value in each group, i.e., $\|\cdot\|_{\tilde{\infty}}$ was the ratio indexed largest value in a group when computing groupwise weights.

B. Test on Simulated Sequences

1) Misaligned Sequence: To verify the ability to handle the misaligned image sequence of the proposed framework, we first experimented on a sequence of simulated misaligned images. As shown in Fig. 6(a), the misalignment between images is large. We simulated the transformation for images in the sequence except for the middle frame. Since all other cloud removal methods are based on the assumption that the target cloud image and the reference images are well aligned (i.e., subpixel-level aligned), we did not compare the reconstruction accuracy at this stage and left it to the next part. When estimating transformation parameters, the solution of weighted least square problems may converge to a local minimum value. Therefore, we initially prealigned images to the middle frame of the sequence to facilitate convergence and avoid being trapped in a local minimum as well. As shown in the last row of Fig. 6, we applied no transformation on the middle frame to better visualize the differences between images before and after alignment. As we can see, the proposed method transformed all the images very close to the middle frame. Cloud and shadow regions were properly detected without any postprocessing, which demonstrated the feasibility of our GRPCA method for processing misaligned sequences.

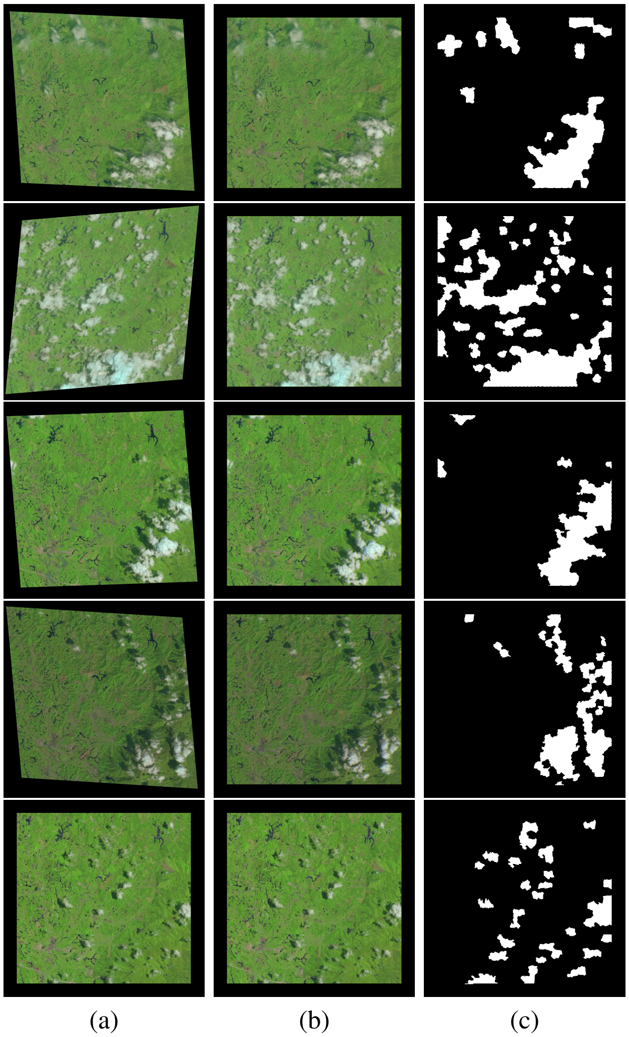
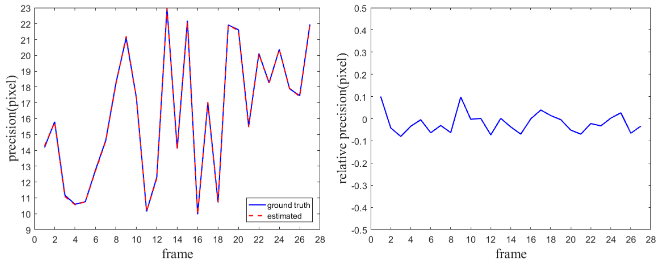


Fig. 6. Experiments on simulated misaligned sequence. The last row is the middle frame with no simulated affine transformation. (a) Simulated misaligned images with random affine transformation. (b) Well-aligned images after GRPCA combined with geometrical transformation. (c) Detected initial cloud and shadow masks.

Moreover, we evaluated the precision of the estimated transformation parameters. The evaluation was conducted by calculating the root-mean-square error (RMSE) of the affine transformation error of pixels in each image. Specifically, given $(x^{,}, y^{,}) = f(x, y)$ denoting 2-D affine transformation, we set $dS = (x^{,} - x)^2 + (y^{,} - y)^{2^{(1/2)}}$ as affine transformation error of each pixel. As shown in Fig. 7(a), the RMSE of affine transformation error of original images in the sequence ranged from 10 to 23 pixels, and the estimated transformation error by our method highly coincided with the simulated ground truth. Fig. 7(b) shows the relative accuracy of the estimated and the simulated transformation error, which is less than 0.1 pixel for all images in the sequence, demonstrating the superior performance of our method. A sequence of well-aligned images and their cloud masks were generated at this stage, which would be used in the DRPCA step for the final cloud removal and reconstruction.

[3] The Landsat-8 collects images of the Earth with a 16-day repeat cycle. The Sentinel-2 revisits the same place in every 5 days.



(a) (b)

Fig. 7. Evaluation of the precision of estimated affine transformation parameters. (a) Comparison of RMSE of affine transformation error. (b) Relative error between estimated and simulated transformations.

2) Reconstruction Accuracy Comparison: After the coarse detection of clouds and their shadows in the first step, we applied DRPCA method to finely reconstruct the missing information contaminated by clouds and shadows. To evaluate the accuracy of our reconstruction approach, and compared to our previous work [15], another representative method, i.e., similar-pixel-based spatial and temporal weighted regression (STWR) was included for comparison. For the Poisson information clone (PIC) and PM-MTGSR methods, due to their dependence on cloud-free images, we manually chose the best suitable reference images for them. The same reference images were also used to generate direct pixel replacement (PR) results to visualize the radiometric difference. As shown in Fig. 8, all four approaches achieve visually plausible reconstruction results and exhibit no obvious discontinuity over cloud and shadow boundaries. More importantly, compared to the original cloud-free images, our reconstruction results show slightly better consistency with the remaining cloud-free region than the other three methods most of the time. This indicates that information reconstruction modeled by low-rank is more powerful to maintain consistency over the entire image. Furthermore, to quantitatively assess the reconstruction accuracy of the four methods, RMSE, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used to evaluate their reconstruction accuracy. As shown in Table I, compared to the other three methods, the proposed DRPCA has an overall win except for the Image 3 in reconstructing cloud-contaminated sequence images. In our reconstruction experiments, Image 3 is the first image in its sequence, i.e., the earliest acquired image, it differs the most compared to other images. Based on the low-rank assumption, our DRPCA method recovers the contaminated areas through iterative SVD, which will tend to favor those principal components of the remaining cloud-free counterparts in the sequence. Therefore, the reconstruction result of Image 3 by the DRPCA may be slightly worse than other methods with selected reference images.

In addition, we presented an efficiency evaluation of the four methods. Three images with different cloud-cover rate were selected in each sequence to calculate their reconstruction time. Since DRPCA recovers a sequence of images one time so that the time of reconstructing a single image is divided by the total number of images. The STWR method iteratively reconstructs each cloud image until all cloud-contaminated images are recovered; therefore, its time is calculated by dividing the whole time with the number of cloud target images. As shown in Table II, our method spent the least amount of time and is independent of the cloud-cover rate compared to the three other methods. This is because the DRPCA iteratively recovers the whole cloud sequence by one SVD and one soft-thresholding step each time, and it converges within less than 15 iterations when processing sequences of about 30 images as we find in our experiments. On the other hand, other methods essentially reconstruct one target image at each time. PIC reconstructs each image by solving a group of Poisson equations with boundary constraints, which is time-consuming when cloud size and the number of boundaries increase. STWR iteratively recovers cloud images patch by patch so that its reconstruction time increases along with cloud-cover rate and number of cloud patches and images. The PM-MTGSR method spends too much time for reconstruction due to dictionary learning and sparse coding. In summary, DRPCA achieves a tremendous improvement of reconstruction efficiency in dealing with cloud remote sensing sequence, and guarantees a high-quality reconstruction accuracy as well.

TABLE I

Quantitative Assessment of Four Different Reconstruction Methods. Images 1–3 Are Related to Simulated Cloud Images in Fig. 8(a). The Best Evaluation Values Between the Three Methods Are Highlighted

Image	Method	RMSE	PSNR	SSIM
Image1	PR	3.4351	37.4119	0.9888
	PIC	2.6683	39.6053	0.9907
	STWR	2.4567	40.3237	0.9919
	PM-MTGSR	2.7165	39.4507	0.9911
	DRPCA	**2.2747**	**40.9925**	**0.9939**
Image2	PR	4.1043	35.8660	0.9956
	PIC	1.9483	42.3293	0.9973
	STWR	1.9157	42.4843	0.9975
	PM-MTGSR	1.5494	44.3274	0.9982
	DRPCA	**1.4853**	**44.6945**	**0.9985**
Image3	PR	3.5869	37.0365	0.9961
	PIC	1.8310	42.8660	0.9975
	STWR	**1.4251**	**45.3671**	**0.9988**
	PM-MTGSR	1.6111	43.9885	0.9971
	DRPCA	1.4728	44.7681	0.9983

TABLE II

Efficiency Comparison Between PIC, PM-MTGSR, and the Proposed DRPCA. The Number of Images in Each Sequence Is 28, 22, and 23, Respectively

Image	Cloud	PIC(s)	STWR(s)	PM-MTGSR(s)	DRPCA(s)
Image1	3.15%	9.0350	60.2920	510.2880	1.12
	5.81%	17.8182	60.2920	530.4540	1.12
	7.96%	24.2246	60.2920	503.6640	1.12
Image2	5.06%	15.0872	55.2710	475.4460	1.10
	6.94%	21.3898	55.2710	459.7200	1.10
	9.38%	30.1769	55.2710	514.5600	1.10
Image3	5.27%	15.2959	35.8587	544.1220	1.10
	5.14%	15.0204	35.8587	528.7680	1.10
	5.34%	15.0625	35.8587	494.8200	1.10

C. Real Images Test

After demonstrating the outperformance of the DRPCA method in reconstructing cloud contaminated region, at this

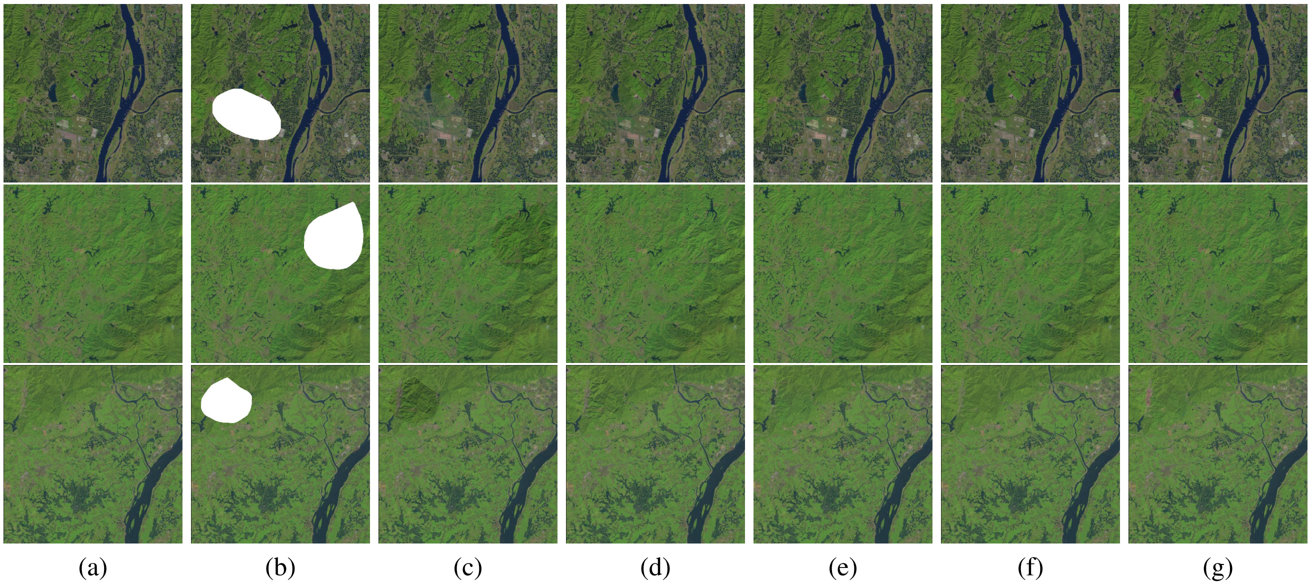


Fig. 8. Three cloud removal examples of the simulated data. (a) Original cloud-free images prepared for simulation. (b) Simulated cloud images. (c) Result of direct PR of the reference image, which is presented to show the radiometric difference between images. (d)–(g) Cloud removal results of PIC, PM-MTGSR, STWR, and our method, respectively.

stage, we focus on specifying the improvement of the proposed GRPCA comparing to the plain RPCA in initial cloud and shadow region detection. In our previous work [15], a plain RPCA was utilized followed by several morphological operations to generate initial cloud and shadow region. However, it has two drawbacks. First, an empirical threshold should be properly set to obtain a binary mask from the sparse component S of plain RPCA. Second, a combination of several morphological operators is required to eliminate noise and to over-cover clouds and shadows, which would absorb in more cloud-free pixels. As shown in Fig. 9, initial cloud and shadow masks generated by plain RPCA over-cover more cloud-free pixels than those of GRPCA because of morphological operations. The proposed GRPCA can obtain the crisper region of clouds and shadows profiting from superpixel good approximation to cloud boundaries. Thanks to group shrinkage, no additional thresholding process is required and the initial mask is generated directly by binarizing the group-sparse component. In addition, benefiting from pseudo $\| \cdot \|_{\widehat{\infty}}$ norm, some noiselike large magnitude sparse outliers that cannot be eliminated by morphological operations are cleaned up by GRPCA as shown in Fig. 9. In short, the proposed GRPCA method can generate crisper cloud and shadow masks and is robust to some noiselike noncloud sparse outliers as well.

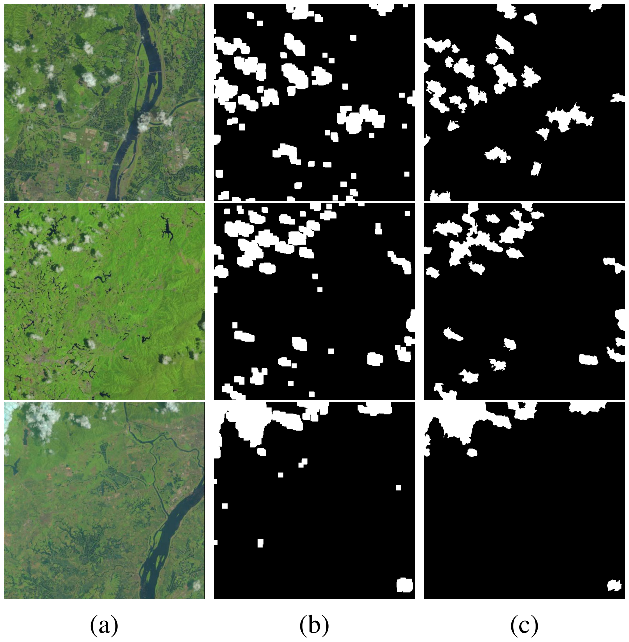


Fig. 9. Comparison of initial cloud and shadow masks. (a) Real cloud images. (b) Mask generated by plain RPCA. (c) Mask generated by GRPCA method.

To test the robustness of the coarse-to-fine framework for cloud removal in real remote sensing sequences, we experimented on three Landsat-8 sites that contained homogeneous and heterogeneous land-covers, and two Sentinel-2 sites located at urban and mountain area, respectively. In the first Landsat-8 site, an urban area containing artifacts and rivers was tested. This scene has some land-cover changes, such as variation of water level and new buildings. As shown in Fig. 10, except for clouds and shadows, two regions of new buildings are falsely detected as outliers marked by red box. The reason is that these kinds of land-cover changes could not fit into a low-rank background model so that they were decomposed into the sparse component. Their magnitude in sparse component was too large to be distinguished from clouds and shadows. That is to say, our method may be prone to reconstructing such land-cover changed regions by mistake with high possibility.

For the second Landsat-8 site, we investigated a mountain area where land-cover was fair homogeneous over the scene and existed few land-cover changes. In such area,

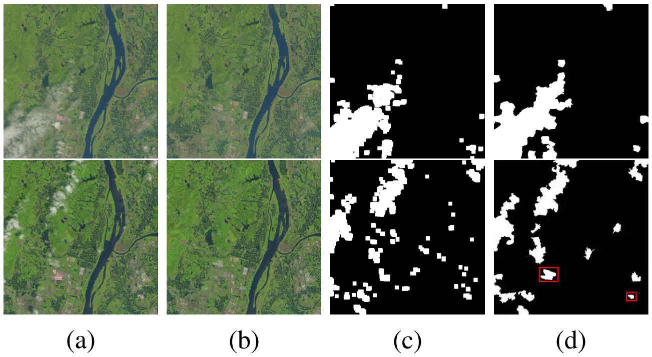


Fig. 10. Experiments on Landsat-8 images site 1 that contains mostly urban area. (a) Real cloud image. (b) Reconstructed cloud-free image. (c) Initial cloud and shadow masks of TRPCA. (d) Initial cloud and shadow masks of GRPCA.

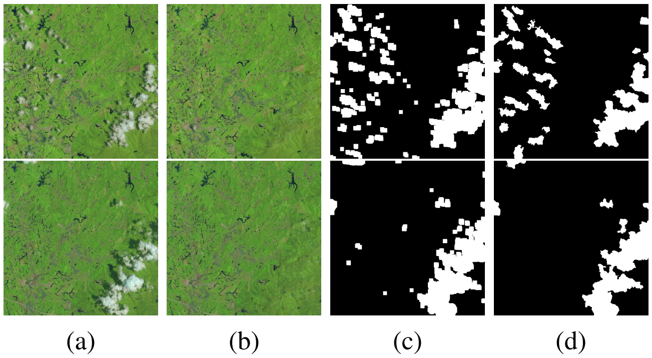


Fig. 11. Experiments on Landsat-8 images site 2 that contains mainly mountain area. (a) Real cloud image. (b) Reconstructed cloud-free image. (c) Initial cloud and shadow masks of TRPCA. (d) Initial cloud and shadow masks of GRPCA.

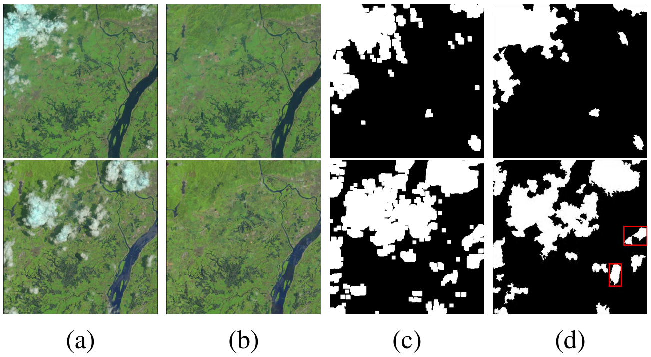


Fig. 12. Experiments on Landsat-8 images site 3 that is located at a suburban area. (a) Real cloud image. (b) Reconstructed cloud-free image. (c) Initial cloud and shadow masks of TRPCA. (d) Initial cloud and shadow masks of GRPCA.

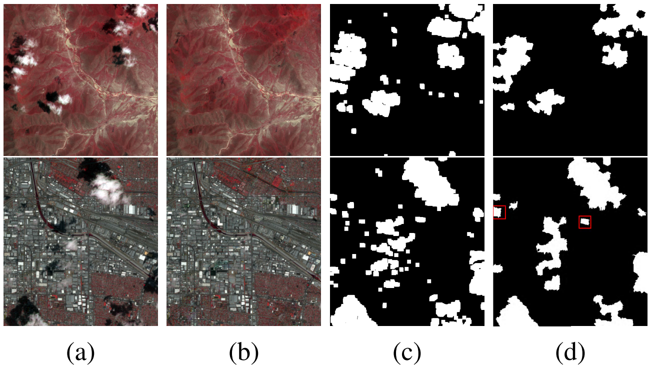


Fig. 13. Experiments on two Sentinel-2 sites located at mountain area and urban area, respectively. (a) Real cloud image. (b) Reconstructed cloud-free image. (c) Initial cloud and shadow masks of TRPCA. (d) Initial cloud and shadow masks of GRPCA.

even though seasonal changes may happen which could turn a bare land to a green land, these changes are usually moderate compared to those of clouds and shadows. Unlike new buildings of which change magnitude is often coincident with clouds, the related values of land-cover changes in mountain area are small in sparse component. It is reasonable to filter them off when performing shrinkage in estimating group-sparse outliers. As we can see in Fig. 11, all clouds and their shadows are included in the initially detected mask, and less redundant cloud-free regions are absorbed in.

The last Landsat-8 site was located in a suburban area that contained mainly farmland and mountain. The seasonal changes may occur in a large region in such scene. Except for unusual sudden changes such as cloud-polluted water or river flood, most of the land-cover changes are often smooth enough to be ignored in the sparse component as discussed in the experiment on the second site. As shown in Fig. 12, crisp cloud and shadow regions are detected through the proposed GRPCA method. However, there are groups of pixels that are falsely detected as clouds, marked in the red box. The reflection of thick cloud on the water surface makes them very different from the general water surface in other images, leading them to be segmented as clouds with high possibility.

On the two Sentinel-2 sites, their results are similar to those of Landsat-8 scenes. As shown in Fig. 13, clouds and shadows are well detected with less unwanted cloud-free pixels compared to the TRPCA method. In the mountain area site, cloud regions are properly over-covered. With regards to the urban area site, two changed regions with large magnitude are falsely detected as discussed before. Above all, the performance of our coarse-to-fine cloud removal framework depends greatly on the first GRPCA step. Given proper cloud and shadow masks, DRPCA can recover cloud contaminated regions with high accuracy and efficiency. However, the low-rank-based method is prone to falsely detect land-cover changes with large magnitude as clouds in some cases.

Finally, we also present several zoomed-in views of the reconstructed results. In most cases, the proposed GRPCA can generate crisp cloud and shadow masks. The shape of initial masks depends greatly on superpixels segmented by SLIC method that can always adapt well to natural objects. As shown in Fig. 14, the initial masks not only cover all cloud and shadow regions but also adapt to their shapes well with little cloud-free pixels being absorbed in. As long as the cloud-contaminated areas are completely included in the detected masks, our method can reconstruct visually plausible cloud-free images.

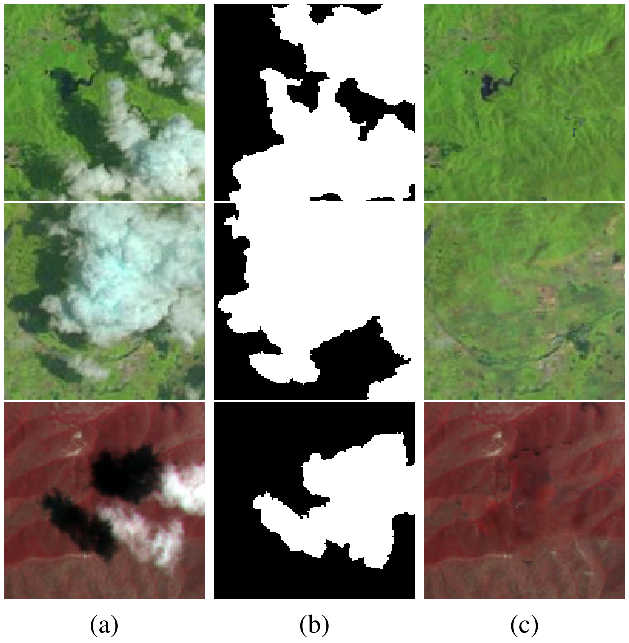


Fig. 14. Zoomed-in view of the reconstruction results of real images. (a) Zoomed-in view of real cloud images. (b) Zoomed-in view of initial mask of clouds and shadows. (c) Zoomed-in view of reconstructed cloud-free images.

V. Conclusion

In this paper, we have developed a novel coarse-to-fine framework to remove clouds and shadows in a remote sensing image sequence. The proposed method takes spatial coherence into consideration and adopts superpixel to cluster object pixels. Group-sparsity-constrained RPCA is proposed to detect initial cloud and shadow regions without any further post-processing. Specifically, we apply nonoverlapping groups and design groupwise weights to facilitate segmentation between cloud and cloud-free groups. Significant improvement of the proposed group-sparse RPCA has been demonstrated in real image experiments compared to the plain RPCA. Moreover, we have enabled our method to handle misaligned sequence images, which has never been discussed by other methods, and experiments also verify the feasibility of such extension. Significantly superior to the available methods, neither cloud-free reference image(s) nor specific operations of cloud and shadow detection are required in our method. Quantitative experiments demonstrate that the proposed method recovers the cloud images with high accuracy compared to several state-of-the-art representative methods. In addition, essentially different from other methods, our method processes a batch of images and exhibits a tremendous efficiency improvement than all other methods. On the other hand, the proposed method can generate a crisp cloud and shadow detection results, which can also be used for other applications. That is to say, we combine two traditional works (i.e., cloud and shadow detection and image restoration) into a whole, showing great potential to be applied in remote sensing image processing.

Meanwhile, there is still some room for improvement. As discussed in the experiments part, our method is prone to falsely detect land-cover changes with large magnitude as clouds. The reason is that most of these changes have large values in the sparse component, which cannot be filtered off through our $\|\cdot\|_\infty$ norm even if introducing a ratio index. Thus, more insightful contextual information or spectral information should be incorporated for more robust cloud region detection in future works. In addition, due to the limitation of data acquirement, we have not tested our method on real misaligned image sequences. In the future, based on 2-D transformation, it is worth extending the method to process images from different optical sensors with similar resolution.

Acknowledgment

The authors would like to thank Dr. B. Chen and Dr. X. Li for providing experiment results of STWR method and PM-MTGSR method, respectively. They would also like to thank two anonymous reviewers and the editor for providing valuable suggestions and comments.

References

[1] J. Ju and D. P. Roy, “The availability of cloud-free landsat ETM+ data over the conterminous United States and globally,” *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1196–1211, Mar. 2008.

[2] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, “Cloud removal based on sparse representation via multitemporal dictionary learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2998–3006, May 2016.

[3] F. Chen, Z. Zhao, L. Peng, and D. Yan, “Clouds and cloud shadows removal from high-resolution remote sensing images,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2005, pp. 4256–4259.

[4] A. Maalouf and P. Carre, B. Augereau, and C. Fernandez-Maloigne, “A bandelet-based inpainting technique for clouds removal from remotely sensed images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.

[5] L. Lorenzi, F. Melgani, and G. Mercier, “Inpainting strategies for reconstruction of missing data in VHR images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 914–918, Sep. 2011.

[6] D.-C. Tseng, H.-T. Tseng, and C.-L. Chien, “Automatic cloud removal from multi-temporal SPOT images,” *Appl. Math. Comput.*, vol. 205, no. 2, pp. 584–600, Nov. 2008.

[7] P. Pérez, M. Gangnet, and A. Blake, “Poisson image editing,” *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, Jul. 2003.

[8] C. H. Lin, P. H. Tsai, K. H. Lai, and J. Y. Chen, “Cloud removal from multitemporal satellite images using information cloning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 232–241, Jan. 2013.

[9] 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 J. Photogramm. Remote Sens.*, vol. 92, pp. 54–68, Jun. 2014.

[10] X. Zhu, F. Gao, D. Liu, and J. Chen, “A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 521–525, May 2012.

[11] 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.

[12] L. Lorenzi, F. Melgani, and G. Mercier, “Missing-area reconstruction in multispectral images under a compressive sensing perspective,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3998–4008, Jul. 2013.

[13] 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 Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7086–7098, Nov. 2014.

[14] 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 Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.

[15] F. Wen, Y. Zhang, Z. Gao, and X. Ling, “Two-pass robust component analysis for cloud removal in satellite image sequence,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1090–1094, Jul. 2018.

[16] Z. Li, H. Shen, H. Li, G. Xia, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.*, vol. 191, pp. 342–358, Mar. 2017.

[17] K. Tan, Y. Zhang, and X. Tong, "Cloud extraction from chinese high resolution satellite imagery by probabilistic latent semantic analysis and object-based machine learning," *Remote Sens.*, vol. 8, no. 11, p. 963, Nov. 2016.

[18] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 303–312, 2003.

[19] 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.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

[21] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2016.

[22] K. Jia, T.-H. Chan, and Y. Ma, "Robust and practical face recognition via structured sparsity," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, Oct. 2012, pp. 331–344.

[23] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *J. Mach. Learn. Res.*, vol. 12, pp. 2777–2824, Feb. 2011.

[24] J. Mairal, R. Jenatton, F. R. Bach, and G. R. Obozinski, "Network flow algorithms for structured sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1558–1566.

[25] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2502–2514, Aug. 2015.

[26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[27] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.

[28] Z. Lin, M. Chen, and Y. Ma. (Sep. 2010). "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices." [Online]. Available: https://arxiv.org/abs/1009.5055

[29] Z. Gao, L.-F. Cheong, and Y.-X. Wang, "Block-sparse RPCA for salient motion detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1975–1987, Oct. 2014.

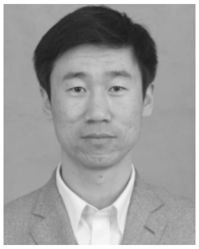
[30] G. Tang and A. Nehorai, "Robust principal component analysis based on low-rank and block-sparse matrix decomposition," in *Proc. 45th Annu. Conf. Inf. Sci. Syst.*, Mar. 2011, pp. 1–5.

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

[32] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[33] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.

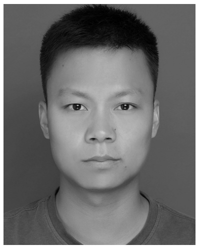
[34] J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *J. Vis. Commun. Image Represent.*, vol. 6, no. 4, pp. 348–365, Dec. 1995.



Yongjun Zhang received the B.S., M.S., and Ph.D. degrees from Wuhan University (WHU), Wuhan, China, in 1997, 2000, and 2002, respectively.

He is currently a Professor of photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, WHU. He has been supported by the Chang Jiang Scholar Program from the Ministry of Education of China in 2017, the China National Science Fund for Excellent Young Scholars in 2013, and the New Century Excellent Talents in University from the Ministry of Education of China in 2007. His research interests include space, aerial, and low-attitude photogrammetry, image matching, combined bundle adjustment with multisource data sets, and 3-D city reconstruction.

Dr. Zhang was a recipient of the National Science and Technology Progress Award (second place) in 2017.



Fei Wen received the B.S. degree from the School of Geosciences and Info-Physics, Central South University, Changsha, China, in 2014. He is currently pursuing the Ph.D. degree in the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

His research interests include remote sensing image processing and machine learning.

Zhi Gao received the B.Eng. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

Since 2008, he has been a Research Fellow (A) and Project Manager with the Interactive and Digital Media Institute, National University of Singapore (NUS), Singapore. He is currently a Research Scientist (A) with the Temasek Laboratories, NUS. He has authored or co-authored several research papers on top journals and conferences, such as IJCV, PAMI, TITS, TCSVT, ACM journals, ECCV, ACCV, and so on. His research interests include computer vision, machine learning, remote sensing and their applications, especially in UAV-based surveillance research and applications.

Xiao Ling was born in 1989. He received the B.S., M.S., and Ph.D. degrees from Wuhan University, Wuhan, China, in 2012, 2014, and 2017, respectively.

He is currently a Post-Doctoral Researcher with the Future Cities Laboratory, Singapore-ETH Center, Singapore. His research interests include photogrammetry, computer vision, camera calibration, and image matching.