ISPRS Journal of Photogrammetry and Remote Sensing 92 (2014) 54–68

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model

Qing Cheng[a], Huanfeng Shen[b,*], Liangpei Zhang[a], Qiangqiang Yuan[c], Chao Zeng[a]

[a] The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei 430079, China
[b] School of Resource and Environmental Science, Wuhan University, Wuhan, Hubei 430079, China
[c] School of Geodesy and Geomatics, Wuhan University, Wuhan, Hubei 430079, China

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 26 February 2014
Accepted 26 February 2014

Keywords:
Cloud removal
Information reconstruction
Spatio-temporal MRF
Similar pixel replacement
Multitemporal
Remotely sensed image

ABSTRACT

Cloud cover is generally present in remotely sensed images, which limits the potential of the images for ground information extraction. Therefore, removing the clouds and recovering the ground information for the cloud-contaminated images is often necessary in many applications. In this paper, an effective method based on similar pixel replacement is developed to solve this task. A missing pixel is filled using an appropriate similar pixel within the remaining region of the target image. A multitemporal image is used as the guidance to locate the similar pixels. A pixel-offset based spatio-temporal Markov random fields (MRF) global function is built to find the most suitable similar pixel. The proposed method was tested on MODIS and Landsat images and their land surface temperature products, and the experiments verify that the proposed method can achieve highly accurate results and is effective at dealing with the obvious atmospheric and seasonal differences between multitemporal images.

1. Introduction

In recent decades, remote sensing has been widely used for many applications, such as terrestrial, oceanic, and atmospheric environmental surveying and parameters inversion. No matter what purpose the remote sensing data are used for, the first thing is to obtain the data required. However, due to the influence of the weather, images acquired from remote sensors are often contaminated by clouds, especially in the humid tropical areas (Tseng et al., 2008). When clouds appear in a region, the true ground information is difficult to obtain, especially in the visible and infrared range, which strongly limits the use of optical images. Although a large number of remotely sensed images are available, there are still big gaps between the data available and the data required. Thus, how to get high-quality image data without cloud contamination in the particular region and time we require is an important issue.

Much research effort has been devoted to the task of cloud removal, to relieve the trouble caused by cloud cover. Cloud removal is essentially an information reconstruction process, and the reconstruction approaches can be grouped into three categories (Lin et al., 2013a, b; Cheng et al., 2013): one category is the non-complementation approaches; the second category is the multi-spectral-complementation based approaches; and the last category comprises the multitemporal-complementation based approaches.

In the non-complementation approaches, without the aid of other complementary data, the information of the cloud-contaminated regions in the remotely sensed image is reconstructed using the remaining parts in the image. The most common method in this category is missing pixel interpolation (Rossi et al., 1994; Van der Meer, 2012); however, most interpolation methods are generally applicable for filling small data gaps and are not appropriate for large-scale cloud removal. In recent years, some new techniques have been introduced to deal with this problem, such as geometry wavelet (Maalouf et al., 2009), maximum a posteriori (MAP) (Shen and Zhang, 2009), and patch filling (Lorenzi et al., 2011). Furthermore, some digital image processing methods can also be used for this problem (Criminisi et al., 2004; Elad et al., 2005; He and Sun, 2012). Generally speaking, for all the methods in this category, the missing regions are synthesized by propagating the geometrical structure from the remaining parts around the missing zone. By using the radiometric information in the image itself, these synthesis methods can yield a visually plausible result, which is appropriate for cloud-free visualization (Lin et al., 2013a, b). However, when the missing regions are large, the disre-

---

* Corresponding author. Tel.: +86 13163235536.
E-mail address: shenhf@whu.edu.cn (H. Shen).

http://dx.doi.org/10.1016/j.isprsjprs.2014.02.015
0924-2716/

gard of information accuracy makes the results unsuitable for data analysis or use in further applications.

In the multispectral-complementation based approaches, multispectral data are utilized for the information reconstruction. The basic idea behind this category of methods is to make use of another complete and clear band of data to restore the contaminated band of data by modeling a relationship between the contaminated band and the auxiliary band. Through an analysis of the visible-band space characterization, Zhang et al. (2002) developed a haze optimized transformation (HOT) method to radiometrically correct the visible-band data contaminated by clouds and haze in Landsat images. Moreover, Li et al. (2012) utilized shortwave infrared imagery to dehaze for visible imagery. These methods are only applicable for the removal of haze and thin clouds, and they do not work for thick cloud removal because thick clouds usually contaminate all of the bands of the acquired images. Shen et al. (2011), Rakwatin et al. (2009), and Gladkova et al. (2012) investigated the reconstruction of the missing data of Aqua MODIS band 6 by using the data in other correlated bands. Similarly, Roy et al. (2008) proposed to use information observed by MODIS to restore Landsat ETM + images. In general, all these approaches are usually constrained by the spectral compatibility (Lin et al., 2013a, b) and tend to have difficulty with thick clouds. Moreover, the multispectral-complementation based approaches are incapable of restoring the quantitative products of remote sensing.

Compared with the non-complementation and multispectral-complementation based methods, the popular multitemporal-complementation based methods are more effective and are able to cope with thick clouds. Since satellite remote sensing systems with a fixed repeat cycle can regularly acquire images in the same area (Zhang et al., 2010), it is easy to get multitemporal images for the same region. The cloud cover area of these images cannot just completely overlap, which is the data source used in the multitemporal image processing methods to reconstruct information. The work of Lee and Crawford (1991), Melgani (2006) and Salberg, 2011 studied the spectro-temporal relationships between the sequences of acquired images for the reconstruction of the areas obscured by clouds or atmospheric disturbance. Image mosaic techniques have also been applied to acquire cloud-free images by mosaicking the cloud-free areas of the multitemporal images (Li et al., 2003; Helmer and Ruefenacht, 2005; Tseng et al., 2008). Chen et al. (2011) developed a neighborhood similar pixel interpolator (NSPI) approach for filling gaps caused by the Landsat ETM + Scan Line Corrector (SLC)-off problem. The NSPI approach was then modified (MNSPI) for the application of thick cloud removal (Zhu et al., 2012). Similarly, some other multitemporal-complementation methods designed for the recovery of missing pixels brought about by sensor failure can also be used for the cloud removal problem, such as the local linear histogram matching (LLHM) approach (Storey et al., 2005) and the weighted linear regression (WLR) approach (Zeng et al., 2013).

In brief, almost all the existing cloud removal methods of the multitemporal-complementation category take the cloud-free regions in the reference image to fill the cloud-contaminated regions in the target image, employing various methods to ensure consistency between the incoming data and the remaining data of the target image. When the multitemporal images do not have big differences, the existing methods can generally work well. However, in many cases, for example when the time interval of the multitemporal image acquisition is not short enough, or when the atmospheric situations are quite different, the multitemporal images may suffer from significant spectral differences, which seriously limits the application of the above methods. In addition, with these methods, the spatial continuity of the ground features may not be preserved in the restored image.

To overcome these problems, this paper presents a new framework for cloud removal, and we merge the ideas of the non-complementation category and the multitemporal-complementation category. A missing pixel is filled only using an appropriate similar pixel within the remaining regions of the target image, and another reference image is used as a guidance to locate the similar pixels. In order to jointly select the most suitable similar pixels in the remaining regions to replace all the missing pixels, a pixel-offset based spatio-temporal MFR function is built, the optimal solution of which is the optimal similar pixels combination. Therefore, the proposed method combines the advantages of the synthesis methods, which reconstruct information using the radiometric information in the image itself to ensure the results keep a high degree of spectral coherence and are visually plausible, and the multitemporal-complementation methods, which guarantee the fine information accuracy of the reconstruction result. Moreover, based on pixel offsets, the proposed spatio-temporal Markov random fields (STMRF) method is able to deal with the obvious differences between multitemporal images.

2. Methodology

2.1. Basic idea

Remotely sensed images are often used for generating a wide range of surface information, so there will always be lots of similar information within an image. This similar information not only includes the local neighborhood similarity, but also includes the non-local similarity (Cheng et al., 2013; Gilboa and Osher, 2008), such as some repeated ground information. For a cloud-contaminated image, some pixel information will be missing. That is to say, we can fill a missing pixel by a similar pixel in the remaining known region. However, since the intensity information of pixels in large cloud regions is totally lost, it is difficult and unreliable to match similar pixels only using the target image information. Therefore, in order to find the positions of similar pixels, we propose to employ another reference image as auxiliary data.

It is a fact that images acquired at different times will usually have some changes because of the different atmospheres or seasonal situations. However, the relative positions of similar pixels within an image will generally be coincident with another scene. The similar pixels are assumed to have similar change trends in multitemporal images (Zhu et al., 2012). For example, in Fig. 1, the locations we picked out with black squares, which are similar pixels in image (a), are also similar pixels in image (b). As another example, a field growing crops will be green in summer but yellow in autumn, and the field will have different spectral characteristics in the images from the two different seasons. However, for each image, the pixels within the field will have the same spectral characteristics. Therefore, another scene can guide us to find the positions of the similar pixels. However, there can also be some other cases, such as an image in early spring before the crop's greenup, in which the crop pixels will show similar spectral characteristics to the bare land. In this case, we may need some other auxiliary information to select similar pixels, such as the spatial neighboring information, structural information, hyperspectral data, or even a geographical map. Our future work will focus on solving this problem.

In Fig. 1, image (c) is the cloud-contaminated image simulated by (b), with some pixels lost in the blank area. We can use image (a) as the reference to find the locations of similar pixels within image (c), and then use a suitable similar pixel to replace the missing pixel. In order to jointly select the most appropriate similar pixels

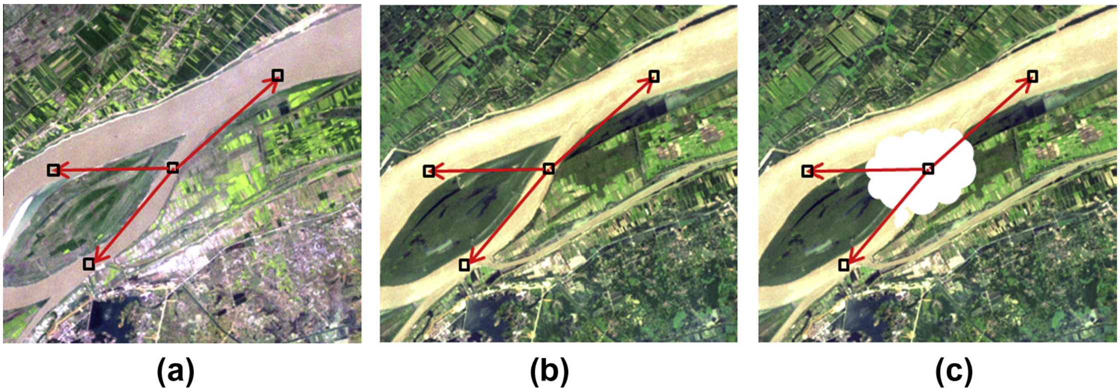

Fig. 1. Two different scenes (a) and (b); (c) is the cloud-contaminated image simulated by (b).

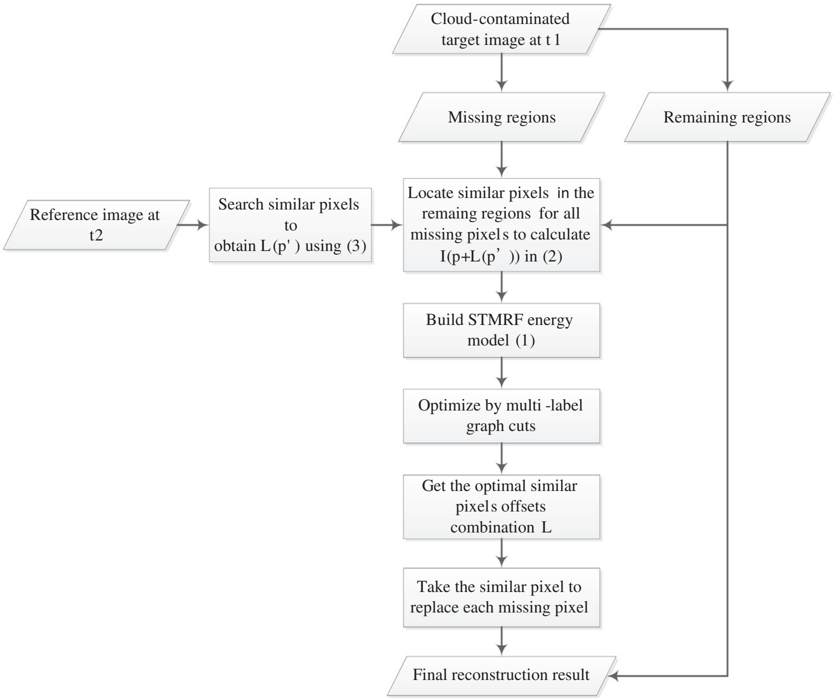

Fig. 2. The flowchart of the STMRF cloud removal method.

for all the missing pixels, the STMRF global model is built to achieve this goal.

2.2. Spatio-temporal Markov random fields (STMRF) model for cloud removal

Markov random fields (MRF) are commonly used to model the contextual correlations among image pixels in terms of conditional prior probabilities of individual pixels, given their neighboring pixels (Li, 2001). The joint distribution of these conditional prior probabilities modeled by MRFs allows us to model the complex global contextual relationship of an entire image by using the local pixel neighborhoods of the MRFs, which makes MRF a very popular contextual model. Furthermore, when multitemporal image series are observed, MRF can be extended to include the temporal dependencies of pixels (Melgani and Serpico, 2003; Liu et al., 2006).

In our method, the relationship between an input image (observed cloud-contaminated target image) $I(x, y)$ and the corresponding reconstructed image $R(x, y)$ is defined by an offset map $L(x, y) = (s_x, s_y)$. A reconstructed pixel $R(x, y)$ will be derived from the input pixel I $(x + s_x, y + s_y)$. Therefore, our goal is to calculate a suitable offset map $L(x, y)$ for all the pixels in the missing region. The optimal offset map minimizes the following spatio-temporal MRF energy function:

$$E(L) = \sum_{p \in \Omega} E_d(L(p)) + \alpha \sum_{p \in \Omega} E_t(L(p), L(p')) + \beta \sum_{(p,p'') \in N} E_s(L(p), L(p'')) \tag{1}$$

Here, Ω is the missing region (with boundary conditions). $p = (x,y)$ represents a pixel in the target image (cloud-contaminated image), p' is the pixel of the reference image in the same position as p, N represents the spatial 4-connected neighbor system of the images, and p'' is a 4-connected neighbor of p in the target image. L is the labeling, where the labels represent the pixel offsets. "$L(p) = (s_x, s_y)$" means that we copy the pixel at $(x + s_x, y + s_y)$ to the location (x,y).

The first term E_d is a data term. E_d is 0 if the label is valid (i.e., $(x + s_x, y + s_y)$ is a known pixel, not located at the missing region); otherwise it is $+\infty$. The second term E_t is a temporal smoothness term. We define E_t as:

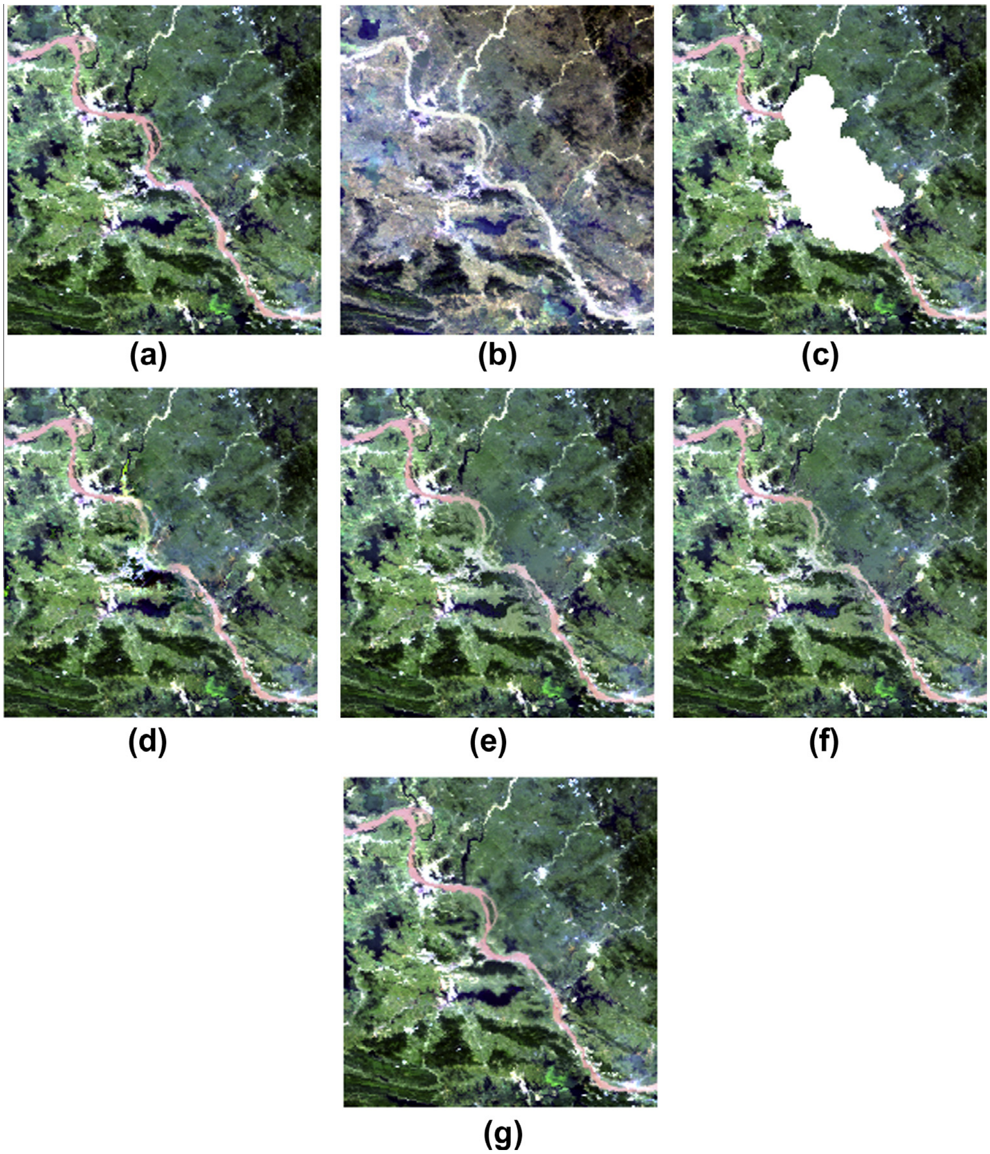

Fig. 3. MODIS images for the first experiment: (a) original image acquired on August 2, 2010; (b) reference image acquired on March 11, 2010; (c) cloud-contaminated image simulated by (a); (d)–(g) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

$$E_t(L(p), L(p')) = \|I(p + L(p)) - I(p + L(p'))\|^2 \tag{2}$$

This temporal smoothness is in consideration of the same change trend for similar pixels as time goes on. Therefore, the relative positions of similar pixels generally remain the same in multitemporal images. This term is used to enforce the offset $L(p)$ to be similar to the offset $L(p')$, which ensures that we use a similar pixel to replace the missing pixel. $L(p')$ is calculated in advance, and its computing method is matching similar pixels to p' in the reference image and obtaining their offsets.

$$L(p') = \underset{L(p')}{\operatorname{argmin}} \|B(p' + L(p')) - B(p')\|^2 \tag{3}$$

Here, $B(p')$ is a block centered at p' , and in most of our experiments we used the block size as 7 × 7 pixels. The similarity is measured by the sum of the squared differences between two blocks. In this temporal smoothness term, in order to reduce the error brought about by anomalous changes (Stein et al., 2002) between the target image and the reference image, we obtain eight pixels that are the most similar to p', so we get eight offsets $L(p')$, then we take the average of the eight values of $I\ (p + L(p'))$ in the calculation.

The third term E_s is a spatial smoothness term. We define E_s as:

$$E_s(L(p), L(p'')) = \|I(p + L(p)) - I(p + L(p''))\|^2 + \|I(p'' + L(p)) - I(p'' + L(p''))\|^2 \tag{4}$$

This smoothness term takes into account the consistency of the spatial neighbors between the missing region and the remaining region in the target image. An offset map discontinuity would occur between two neighboring locations p and p'' if their offsets were different: $L(p) \neq L(p'')$. In this case, a seam will appear between p and p'' in the recovery image. Thus, (4) penalizes such a seam so that the two offsets $L(p)$ and $L(p'')$ are similar near this seam. This smoothness term is similar to those defined in shift-map (Pritch et al., 2009; He and Sun, 2012) or graphcut textures (Kwatra et al., 2003).

α and β in (1) are the parameters weigh the influence of the different information sources on the decision process. There are some

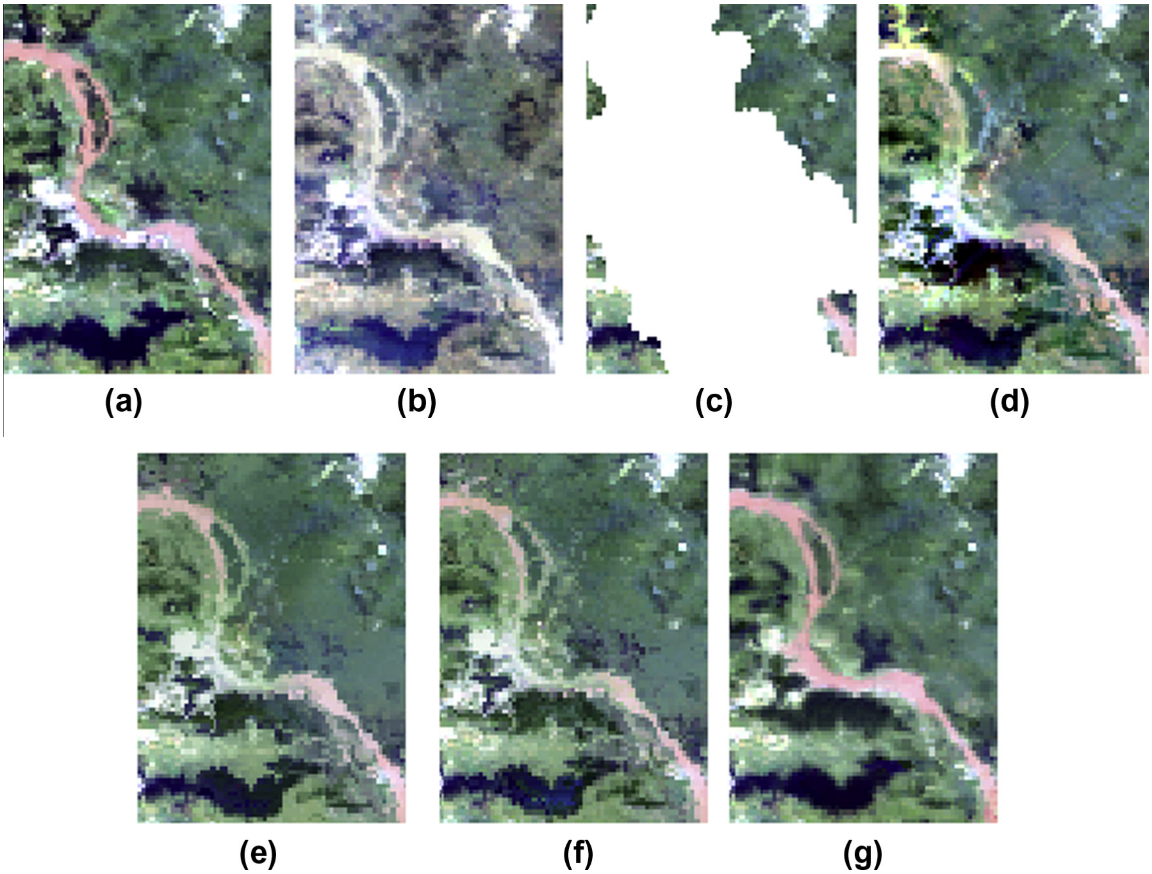

Fig. 4. (a)–(g) Detailed regions cropped from Fig. 2(a)–(g).

Table 1
The NMSE, ARE, and CC values of the four sets of experimental results.

		LLHM	MNSPI	WLR	STMRF	IR (%)
Fig. 3	NMSE	0.10364	0.06701	0.06751	0.05214	22.19
	ARE	0.08097	0.01202	0.01188	0.01056	11.11
	CC	0.81618	0.83849	0.83804	0.86527	3.21
Fig. 5	NMSE	0.04139	0.01724	0.01573	0.01064	32.36
	ARE	0.04271	0.00085	0.00094	0.00052	38.82
	CC	0.88497	0.93460	0.93884	0.96197	2.46
Fig. 7	NMSE	0.00078	0.00049	0.00053	0.00042	14.29
	ARE	0.00199	0.00132	0.00164	0.00104	21.21
	CC	0.91864	0.93816	0.93248	0.95187	1.46
Fig. 9	NMSE	0.00204	0.00148	0.00179	0.00095	35.81
	ARE	0.03414	0.02633	0.02919	0.01891	28.18
	CC	0.71535	0.79003	0.76721	0.81659	3.36

methods for optimizing parameters in MRF, such as "minimum perturbation" method (Melgani and Serpico, 2003) and genetic method (Liu et al., 2008). However, for this spatio-temporal MRF model we proposed above, it can be inferred that the temporal dependence is the main information source for the selection of similar pixels; the spatial dependence is relatively weak, and used to alleviate the spatial discontinuity of the results. In most of the experiments, we used weights of 1 and 0.5 to balance the contribution of the temporal and spatial terms. If the spatial weight significantly increases, the reconstruction accuracy will decline sharply; if the spatial weight significantly decreases, the reconstruction results will have an obvious spatial discontinuity.

We optimize the energy function (1) using multi-label graph cuts (Boykov et al., 2001). The code of the multi-label graph cuts can be downloaded at the following address: http://vision.csd.uwo.ca/code/. A flowchart of the proposed cloud removal method is presented in Fig. 2.

3. Experimental results

3.1. Evaluation of the proposed method

In this part, we conduct four groups of experiments to test and quantitatively evaluate the efficacy of the proposed method. In these four experiments, the cloud-contaminated regions are simulated on the original clear images. The recovered images are then compared with the original clear images from both visual and quantitative fidelity aspects.

In the first experiment, the proposed method is tested on MODIS images with a resolution of 500 m. We use the three bands of band 1 (red), band 4 (green), and band 3 (blue) to undertake the test. The original and reference true color composite images are shown in Fig. 3(a) and (b), acquired on August 2, 2010, and March 11, 2010, respectively, in southeast of China. The cloud-contaminated image simulated by Fig. 3(a) is shown in Fig. 3(c). To make a comparative analysis, the proposed STMRF method is compared with the methods of local linear histogram matching (LLHM) (Storey et al., 2005), modified neighborhood similar pixel interpolation (MNSPI) (Zhu et al., 2012), and weighted linear regression (WLR) (Zeng et al., 2013). The recovery results of each method are shown in Fig. 3(d)–(g).

From Fig. 3 and its zoomed detailed regions in Fig. 4, it can be seen that the image reconstructed by the proposed STMRF method (Fig. 3(g)) appears much closer to the original clear image (Fig. 3(a)), compared to the results of the other methods. For the

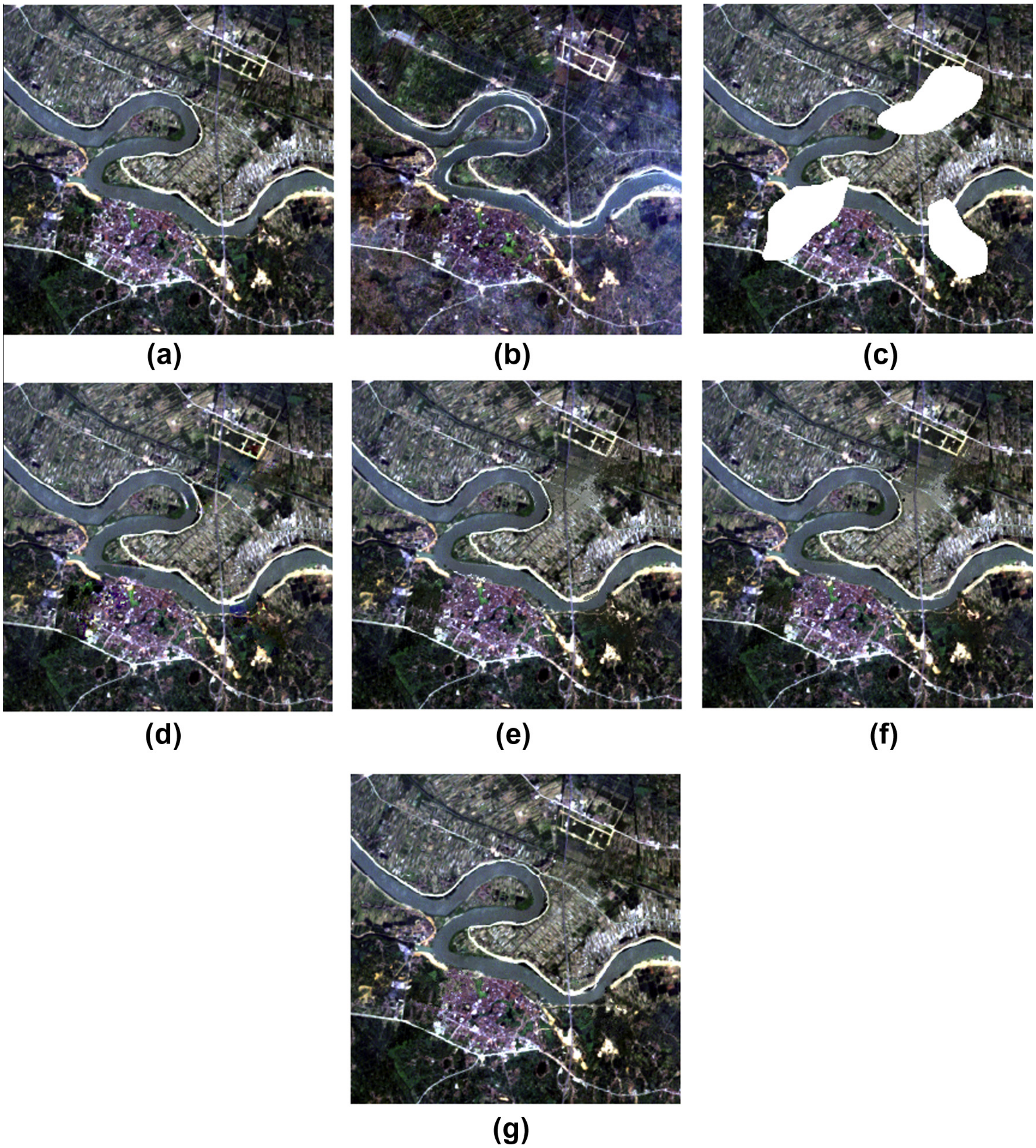

Fig. 5. Landsat TM images for the second experiment: (a) original image acquired on July 9, 2002; (b) reference image acquired on October 13, 2002; (c) cloud-contaminated image simulated by (a); (d)–(g) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

LLHM method (Fig. 3(d)), the recovery result of the cloud-contaminated region has a serious spectral distortion, especially in the river region. This result indicates that when the two images contain a complex terrain and have large spectral differences, a simple histogram matching method is unsatisfactory. The results of the MNSPI and WLR methods (Fig. 3(e) and (f)) appear similar, most of the ground features are well recovered, but in the river region, the spectral characteristic is still different from that in the remaining river region, which may also result from the large spectral differences between the target image and the reference image. For the STMRF method, it shows the most plausible visual result, some of the detailed information is well recovered, and the reconstructed region is consistent with the original image; moreover, some unnecessary noise is effectively suppressed in the result of the proposed method. It is worth noting that the two input images (the original and the reference images) were acquired in different seasons, and the spectral characteristics of the ground features have been significantly changed. The LLHM, MNSPI, and WLR methods cannot deal with this change very well, leading to more errors in their results. Through the guidance of similar pixel offsets and using the radiometric information of the image itself to fill the missing region, the proposed STMRF method is better able to address this issue. The quantitative assessment in Table 1 also shows the clear superiority of the proposed STMRF method.

The second experiment is performed on Landsat TM images with six bands (bands 1, 2, 3, 4, 5, and 7) and a 30-m resolution. Fig. 5(a) and (b) shows the original and the reference TM images (true color composite R = band 3, G = band 2, B = band 1), acquired on July 9, 2002, and October 13, 2002, respectively, in Hubei. Fig. 5(c) is the cloud-contaminated image simulated by Fig. 5(a). Fig. 5(g) shows the reconstruction result of the proposed method. The results of the other comparative methods are listed in

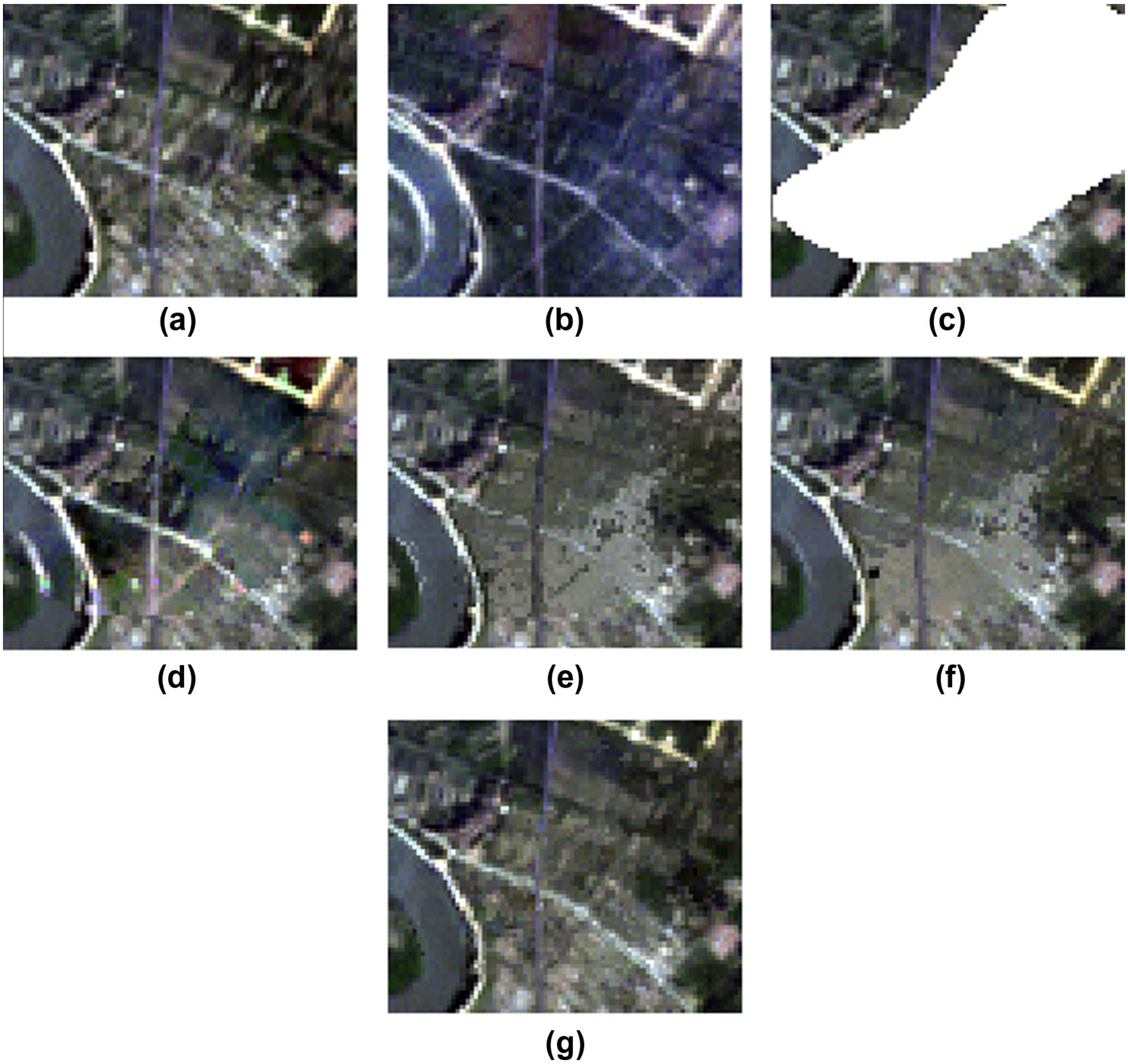

Fig. 6. (a)–(g) Detailed regions cropped from Fig. 5(a)–(g).

Fig. 5(d)–(f). The zoomed regions cropped from Fig. 5(a)–(g) are shown in Fig. 6(a)–(g).

It can be seen that there are some significant changes between the two input images since their acquisition times were not close enough. With a high spatial resolution, the ground features are more complex in these images than those in the first experiment. The results in Fig. 5(d)–(g) indicate that the proposed STMRF method can again achieve a better performance than the other three methods. In the result of the LLHM method (Fig. 6(d)), some obvious spectral distortion still exists. For the MNSPI method (Fig. 6(e)), a lot of the detailed information is lost, causing obvious blurring. The same defect also occurred in the result of WLR method (Fig. 6(f)). For the STMRF method (Fig. 6(g)), the result appears more spectrally and spatially consistent with the surrounding features, and we can see that it is the closest to the original clear image.

Remotely sensed land surface temperature (LST) data from satellite images play an important role in a variety of environmental and ecological applications; however, clouds and other atmospheric disturbances often obscure parts of these satellite images. Our third experiment involves the reconstruction of MODIS LST data. The proposed method is tested on MODIS daily LST images with a resolution of 1 km. Fig. 7(a) and (b) shows the original and the reference LST images, acquired on March 21, 2012, and March 16, 2012, respectively. The image with missing pixels simulated by Fig. 7(a) is shown in Fig. 7(c). Fig. 7(g) shows the reconstruction result of the proposed method. The results of the other comparative methods are shown in Fig. 7(d)–(f). From Fig. 7, we can see that the proposed STMRF method can provide a better reconstruction result for the LST data, and the result is closer to the original clear image data than the results of the other three methods. LST data usually change quickly over a short time interval, which is the reason why most of the existing methods cannot work well for this kind of data. However, the proposed STMRF method is better able to deal with this change.

In order to further compare the results of the four methods, we randomly select a line of pixels in the recovered region of the result image and display their temperature values, as shown in Fig. 8. From Fig. 8, we can see that the temperature curve of the STMRF result is closer to the curve of the original image than the other three curves. This again suggests that the proposed STMRF method can achieve a result that is closer to the original clear image.

The fourth experiment is performed on some other MODIS LST daily data images. Fig. 9(a) and (b) shows the original and reference LST images, acquired on March 12, 2012, and March 14, 2012, respectively. Fig. 9(c) is the cloud-contaminated image simulated by Fig. 9(a). The reconstruction results are listed in Fig. 9(d)–(g). Here, we can see that the two input LST images have

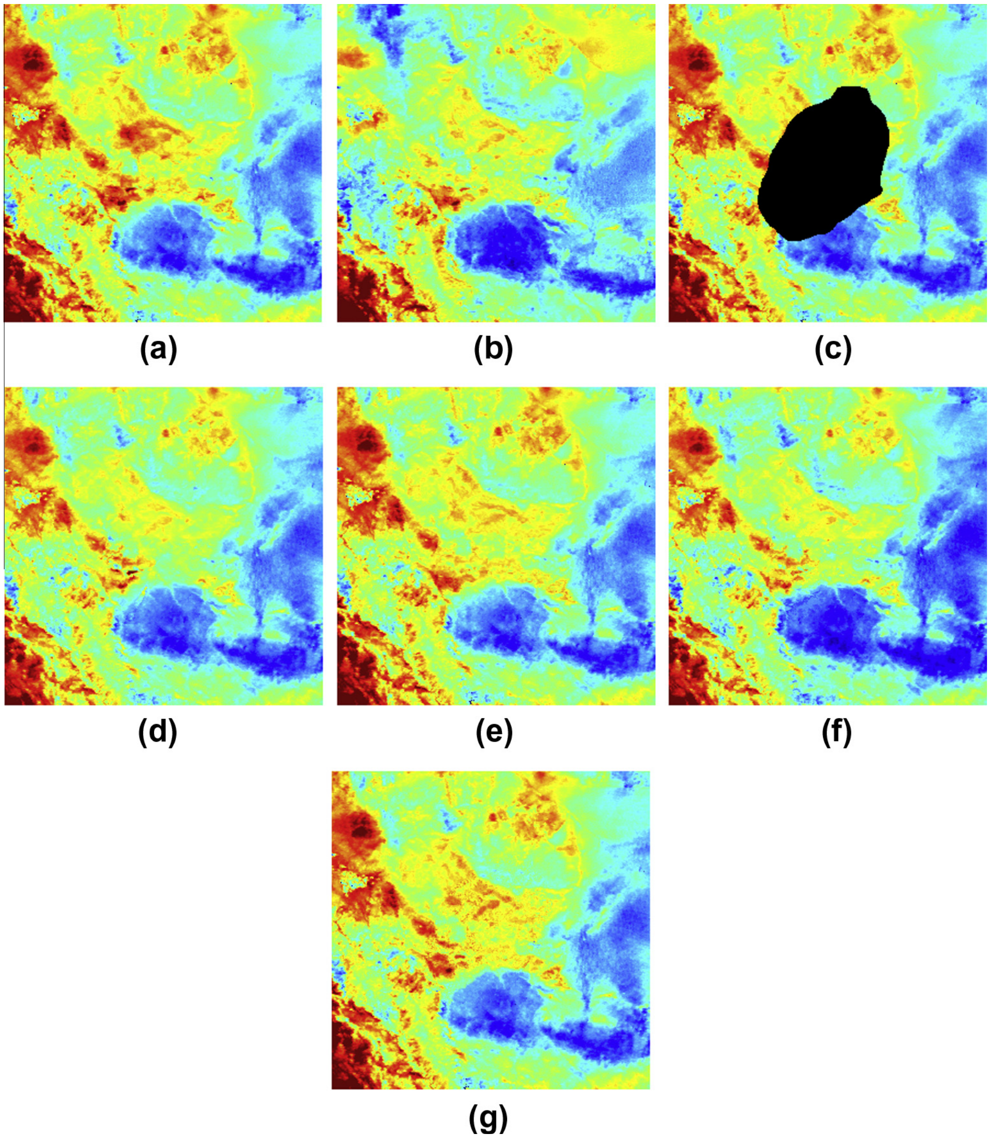

Fig. 7. MODIS LST images for the third experiment: (a) original image acquired on March 21, 2012; (b) reference image acquired on and March 16, 2012; (c) image with missing pixels simulated by (a); (d)–(g) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

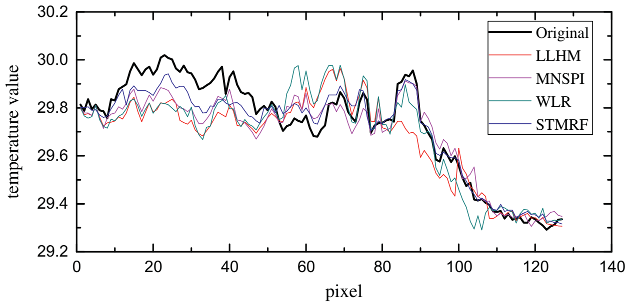

Fig. 8. The temperature values of the pixels in a recovered line for the third experiment.

changed more than in the third experiment. From Fig. 9, it can be seen that the result using the proposed STMRF method (Fig. 9(g)) can still give a better result which is closer to the original image. The results of the other three methods suffer from more errors since they cannot adapt to this large degree of change very well. In order to further compare the results of the four methods, we display the temperature values of a line of pixels in the recovered region, as with the third experiment. The temperature curves are shown in Fig. 10. From Fig. 10, we can see that the temperature curve of the STMRF result is closer to the curve of the original image than the other three curves.

The effectiveness of the proposed STMRF method can also be illustrated by the quantitative assessment. Here, the normalized mean square error (NMSE) index, the average relative error (ARE) index, and the correlation coefficient (CC) index are used to give a quantitative evaluation of the results of the above experiments. Furthermore, an improvement ratio (IR) index (Shen et al. 2013) is defined to evaluate the degree of the improvement of the proposed method with respect to the three comparative methods. The definitions of these evaluation indices are as follows:

$$\mathrm{NMSE} = \frac{\sum_{j=1}^{M}(I_{Oj} - I_{Rj})^2}{\sum_{j=1}^{M}(I_{Oj})^2} \tag{5}$$

$$\mathrm{ARE} = \left(\sum_{j=1}^{M}(|I_{Oj} - I_{Rj}|/I_{Oj})\right)/M \tag{6}$$

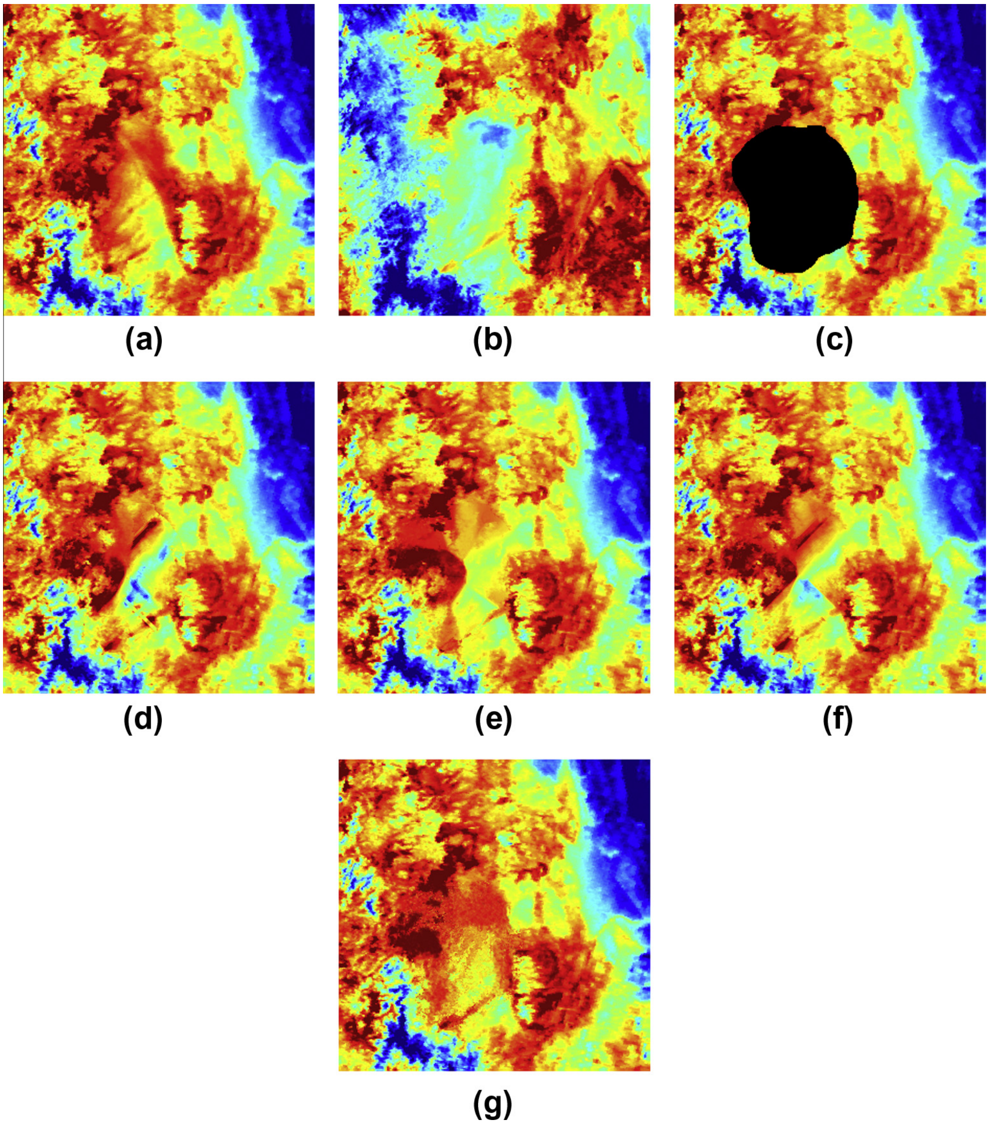

Fig. 9. MODIS LST images for the fourth experiment: (a) original image acquired on March 12, 2012; (b) reference image acquired on March 14, 2012; (c) image with missing pixels simulated by (a); (d)–(g) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

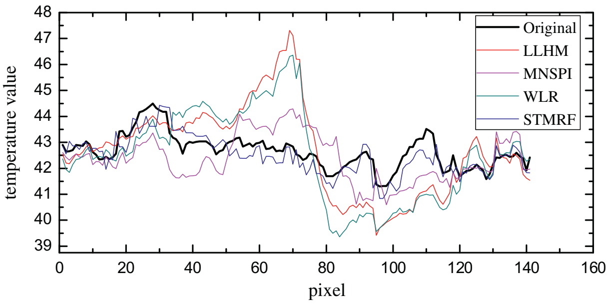

Fig. 10. The temperature values of the pixels in a recovered line for the fourth experiment.

$$\mathrm{CC}=\frac{\sum_{j=1}^{M}(I_{Oj}-\bar{I}_O)(I_{Rj}-\bar{I}_R)}{\sqrt{\sum_{j=1}^{M}(I_{Oj}-\bar{I}_O)^2\sum_{i=1}^{M}(I_{Rj}-\bar{I}_R)^2}} \tag{7}$$

$$\mathrm{IR}=\frac{|V_{existing}-V_{proposed}|}{V_{existing}}\times 100\% \tag{8}$$

where M is the total number of missing pixels, I_{Oj} and I_{Rj} are the original and the recovered values of the jth missing pixel, respectively, and $\bar{I}_O$ and $\bar{I}_R$ are the mean values. $V_{existing}$ is the best evaluation value in the three comparative methods. $V_{proposed}$ is the evaluation value of the proposed method. The smaller the NMSE and ARE values are, the better the recovery is. A larger CC indicates a closer consistency between the groups of data, and the data will be identical when CC equals 1.

The NMSE, ARE, and CC indices are all calculated on the cloud-contaminated regions, and we compute their mean values of all the bands for the four experiments. The results are listed in Table 1. It can be seen from Table 1 that, for the result using the proposed STMRF method, the NMSE and ARE values are both lower, and the CC value is higher than for the other methods. The same improvements are evident in all the four groups of experimental results. Moreover, from the quantitative assessment of Figs. 7 and 9 shown in Table 1, we can see that the superiority of the proposed STMRF method is more distinct in Fig. 9 than in Fig. 7. This is because the changes in the input multitemporal images in Fig. 9 are greater than in Fig. 7, and the proposed STMRF method is much better at handling these big changes. Overall, the four groups of

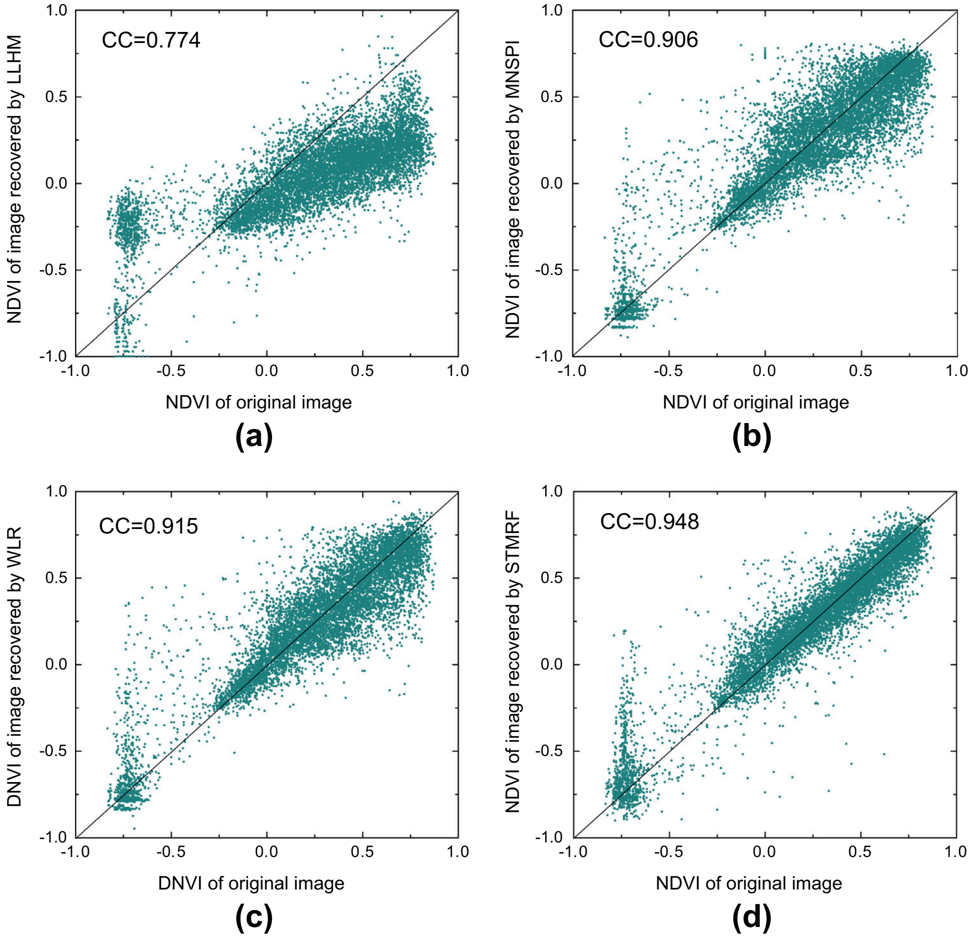

Fig. 11. The scatterplots of the real and the recovered NDVI values: (a) the LLHM method; (b) the MNSPI method; (c) the WLR method; and (d) the proposed STMRF method.

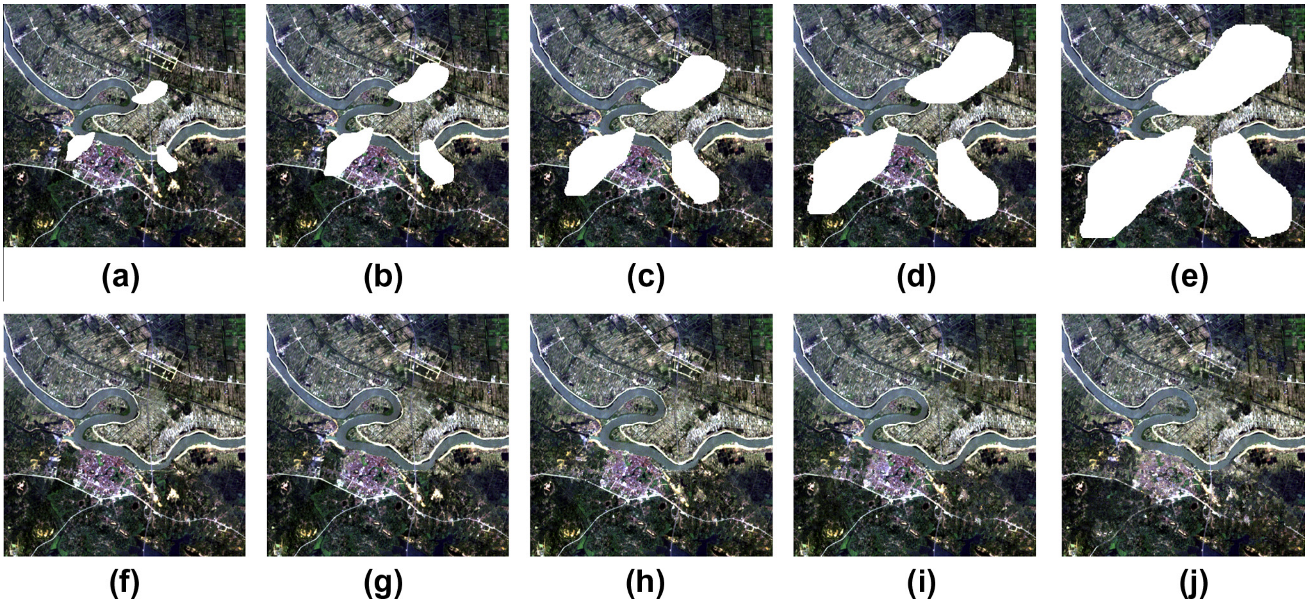

Fig. 12. The simulated clouds with different sizes, and the results of the STMRF method. The length of the major axis for the lower-left cloud is: (a) 50 pixels; (b) 100 pixels; (c) 150 pixels; (d) 200 pixels; and (e) 250 pixels. (f)–(j) are the corresponding results of the STMRF method.

experiments confirm that the proposed STMRF method can provide a more accurate and robust reconstruction result.

In order to assess the usability of the reconstructed data for further applications, the normalized difference vegetation indices (NDVI) of the recovered images in the second experiment are compared. Before the calculation of the NDVI, an atmospheric correction using a dark object subtraction method (Chavez, 1988) is implemented on the original and recovered images. According to

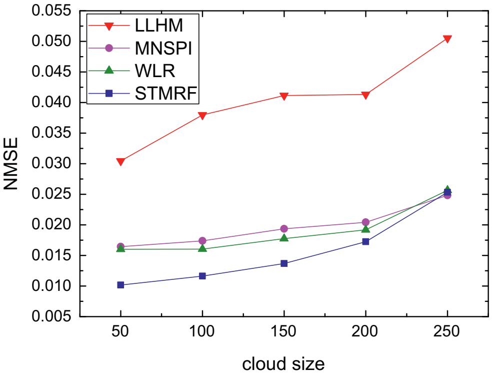

Fig. 13. The NMSE values of the reconstruction results for various cloud sizes.

the scatterplots listed in Fig. 11, it is suggested that: the LLHM method estimate the values in the cloud-contaminated regions with larger errors; the results of the MNSPI and WLR methods have a better agreement with the original image; and the proposed STMRF method still gives the best performance. As shown in Fig. 11, the CC value of the LLHM result is 0.774 and, which is unsatisfactory; the MNSPI and WLR results are 0.906 and 0.915, respectively, which have an obvious increase; and the result of the proposed STMRF method gives the highest value of 0.948. These results demonstrate that the cloud removal results derived from the proposed STMRF method can provide a better support for further applications.

In this part, we analyze the sensitivity of the proposed method to cloud size. The clouds are simulated at five different levels, as shown in Fig. 12. The image in this test is of 400 × 400 pixels size. The length of the major axis for the simulated cloud in the lower-left corner of the image is about 50 pixels in Fig. 12(a), 100 pixels in Fig. 12(b), 150 pixels in Fig. 12(c), 200 pixels in Fig. 12(d), and 250 pixels in Fig. 12(e), and the length of the minor axis is half of the length the major axis. The other two clouds are just a little bit smaller than the lower-left one. The NMSE values of the five groups of reconstruction results are calculated, as shown in Fig. 13. Through the contrast and comparison of the NMSE values for the different methods with different cloud sizes, we can observe the following phenomenon: (1) the smaller the cloud size is, the more obvious the advantages of the proposed STMRF method are; (2) when the cloud size grows to the 200 pixels level (about 35% of the test image), the proposed STMRF method still shows a better reconstruction result than the other three methods; and (3) when the cloud size grows to the 250 pixels level (about 55% of the test image), the four methods have a similar poor performance. These phenomenon can be explained as follows: for the LLHM, MNSPI, and WLR methods, they all generally use the local neighborhood information to predict the missing information. However, as the cloud size increases, the valid local neighborhood information decreases. For the proposed STMRF method, it not only uses the local information but also the non-local information, and is able to achieve a better result. Despite considering the 1st geographic law (pixels with a greater distance between them may be less similar), the combination of local and non-local information is still

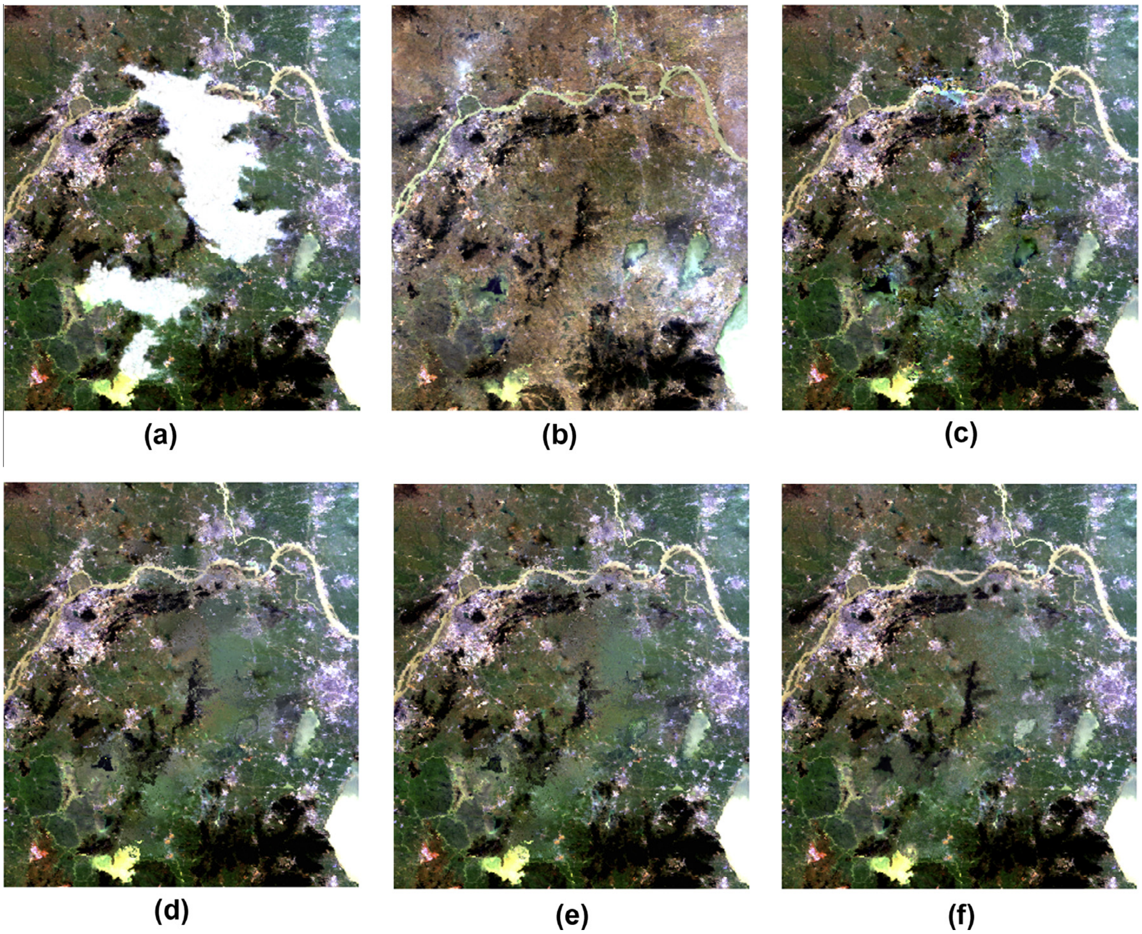

Fig. 14. MODIS images for the first real data experiment: (a) original cloud-contaminated image acquired on March 20, 2008; (b) reference image acquired on November 21, 2007; (c)–(f) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

more effective than only using the local information. Besides, the integration of spatio-temporal information into the reconstruction may also is a reason to improve the accuracy values.

3.2. Real data experiments

In this part, we conduct three groups of real data experiments with different image data. For the first real data experiment, the proposed method is tested with cloud removal for a MODIS image. In Fig. 14(a), the original image is partially contaminated by cloud, which leads to parts of the ground information being lost in this image. This cloud-contaminated image was acquired on March 20, 2008, and the reference image was acquired on November 21, 2007, as shown in Fig. 14(b). The cloud removal results of the proposed STMRF method and the other comparative results are shown in Fig. 14(c)–(f). The zoomed regions cropped from Fig. 14(a)–(f) are shown in Fig. 15(a)–(f).

From Figs. 14 and 5, we can see that the result of the LLHM method (Fig. 15(c)) suffers from serious spectral distortion, as with the results of the simulated experiments. For the results of the MNSPI and WLR methods, the spectral consistency is generally kept well; however, there is some obvious noise in the restored region, as shown in Fig. 15(d) and (e). For the result of the proposed STMRF method, as we can see in Fig. 15(f) that the restored region

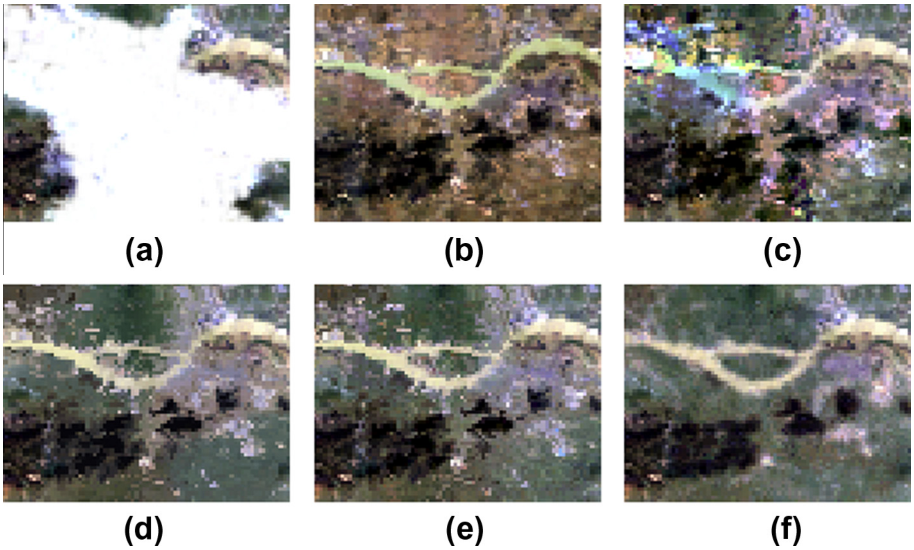

Fig. 15. (a)–(f) Detailed regions cropped from Fig. 14(a)–(f).

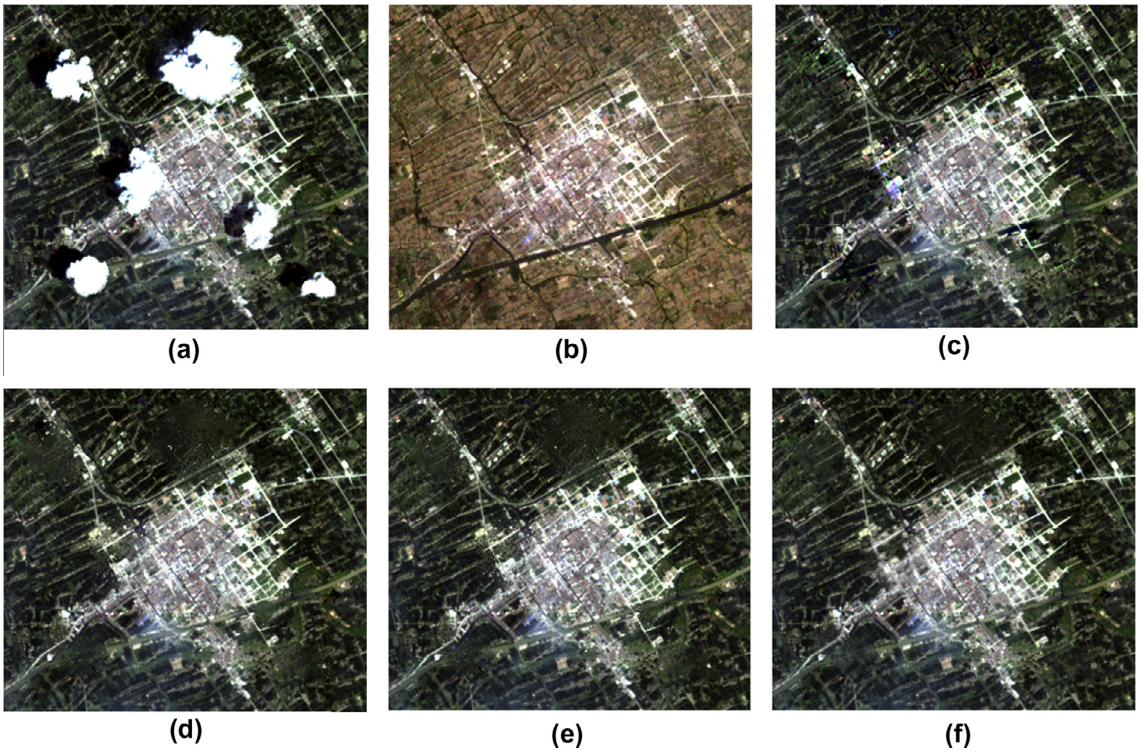

Fig. 16. Landsat TM images for the second real data experiment: (a) original cloud-contaminated image acquired on August 3, 2000; (b) reference image acquired on February 10, 2000; (c)–(f) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

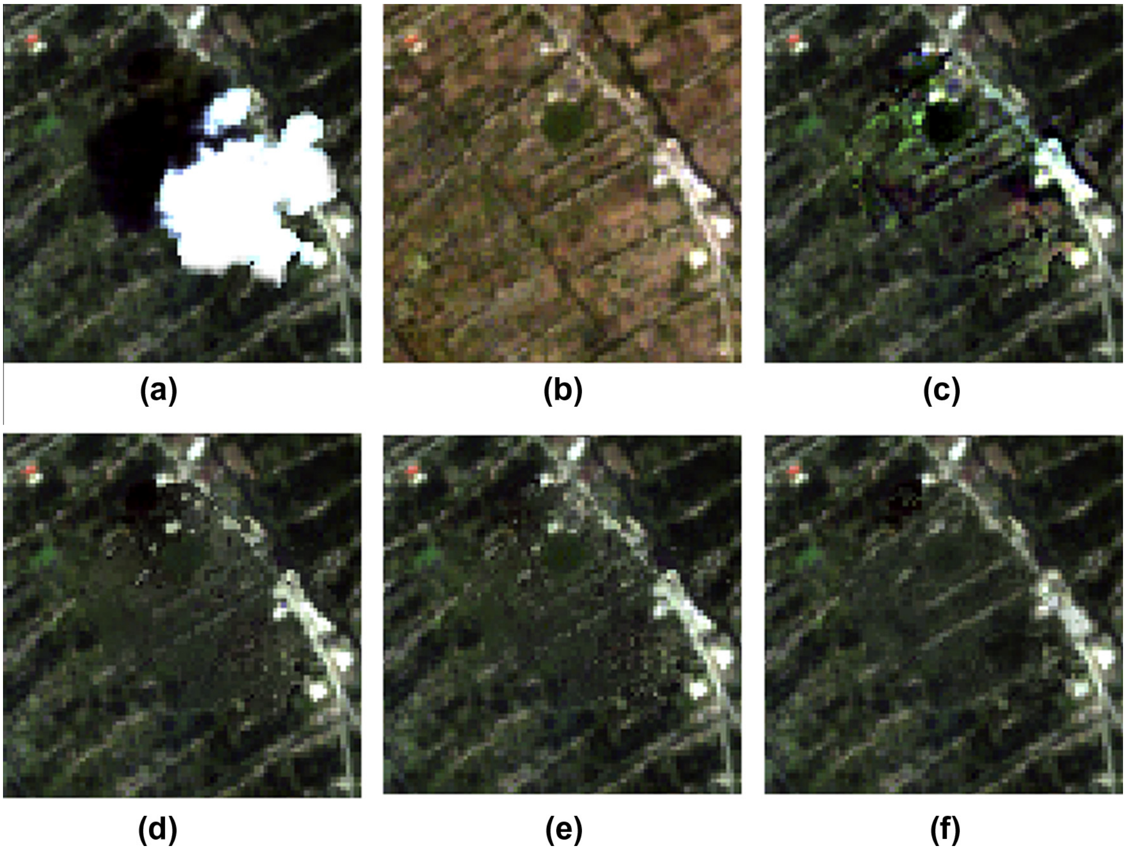

Fig. 17. (a)–(f) Detailed regions cropped from Fig. 15(a)–(f).

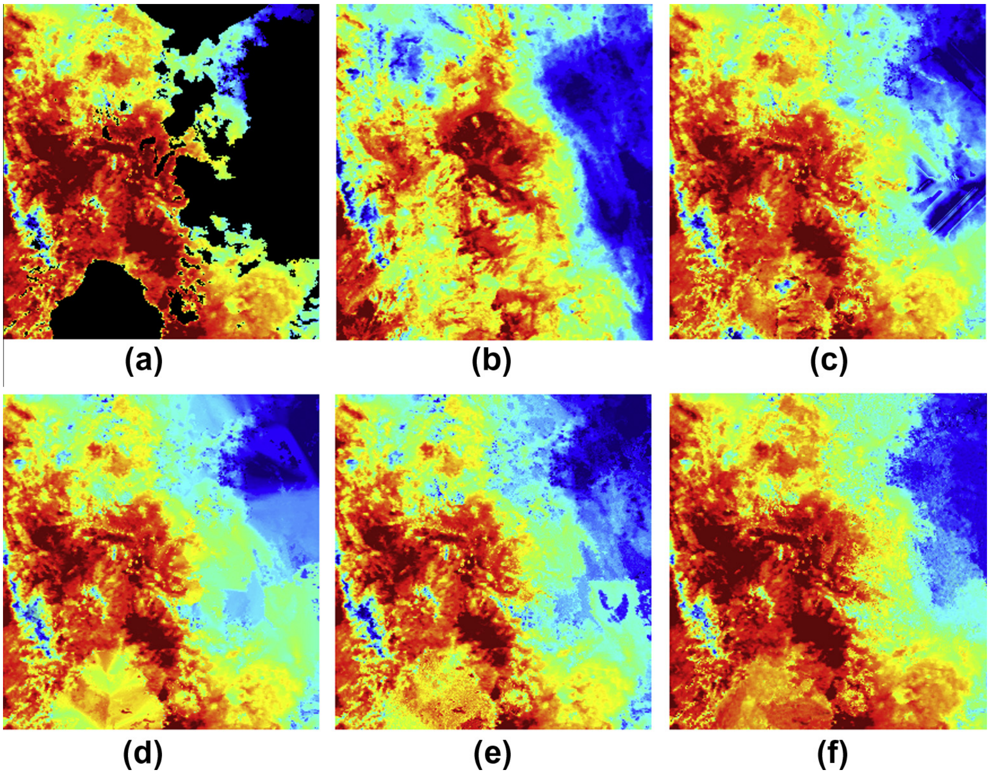

Fig. 18. MODIS LST images for the third real data experiment: (a) original image acquired on March 21, 2012; (b) reference image acquired on and March 7, 2012; (c)–(f) are the results recovered by LLHM, MNSPI, WLR, and the proposed STMRF method, respectively.

is more spectrally and spatially continuous with the remaining region, and the noise is suppressed to some degree in this result.

The second real data experiment is cloud removal for a Landsat TM image. Fig. 16(a) and (b) shows the original cloud-contaminated image and the reference TM image, acquired on August 3, 2000, and February 10, 2000, respectively. Since the images were acquired in different seasons, the spectral characteristics of the ground features are significantly different between the two input images. Again, we can see that the weaknesses of the LLHM, MNSPI, and WLR methods are clearly shown in their results (Fig. 16(c)–(e)). The result of the proposed STMRF method is shown in Fig. 16(f). From Fig. 16(c), it can be see that obvious spectral distortion exists in the result of the LLHM method. For the results of the MNSPI method (Fig. 16(d)), a lot of detailed information is not recovered well, and serious noise arises in the restored region. The result of WLR method (Fig. 16(e)) suffers from the same defect with the MNSPI method. For the result of the STMRF method, although the detailed spatial information is not fully recovered, the visual quality is more convincing, and the spectral information is consistent with the remaining region. For the convenience of visual judgment, a series of detailed regions cropped from Fig. 16(a)–(f) are shown in Fig. 17(a)–(f).

The third real data experiment is reconstruction of MODIS LST data. Fig. 18(a) and (b) shows the original and the reference daily LST image, acquired on March 21, 2012, and March 7, 2012, respectively. The recovery results using the proposed method and the comparative methods are listed in Fig. 18(c)–(f). In Fig. 18(c), for the result of LLHM, obvious artifacts can be found. In Fig. 18(e), for WLR, the temperature values are not continuous, with obvious jump, which does not make sense. For STMRF, it provides a more convincing recovery result, as shown in Fig. 18(f). Another competitive result for this test is that of the MNSPI method (Fig. 18(d)).

4. Discussion and conclusion

When affected by cloud cover, much ground information usually cannot be acquired in optical remotely sensed images, which significantly limits the application of these images. Although a large amount of remotely sensed images are available nowadays, a high-quality image which can provide complete, clear ground information is still what we strive for. Therefore, this paper proposes a new and effective method to remove clouds and precisely reconstruct the ground information. A missing pixel is replaced using a similar pixel within the remaining regions of the cloud-contaminated image. Another complementary temporal image is used as a reference to locate the similar pixels. In order to select the most appropriate similar pixels to replace the missing pixels, a pixel-offset based spatio-temporal MFR function model is built, and we use a multi-label graph cut algorithm to get its optimal solution. The optimal solution is the optimal similar pixels combination to replace all the missing pixels.

The experimental results show that the integration of the spatio-temporal image information by Markov random fields for image reconstruction clearly improves the accuracy values over the conventional methods. Moreover, by using the radiometric information in the image itself to reconstruct the missing information, the proposed method is effective and robust to the significant changes brought about by various atmospheric situations and the seasonal changes in multitemporal images. In general, the proposed STMRF method is capable of achieving temporally, spatially, and spectrally coherent reconstruction.

There are, however, some limitations to the proposed method. The method is good at dealing with the significant changes in multitemporal images; however, when the time interval is very short and the atmospheric situation is similar, the acquired multitemporal images will be very similar, or their changes will be simple and linear. In this case, some traditional methods such as fitting, matching, and regression (Zeng et al. 2013) are capable of getting excellent reconstruction results, and their calculations are very fast. Therefore, comparing with these conventional methods, the method we proposed will not have obvious advantage, and will be time consuming. Therefore, in order to remedy this defect, we thus consider combining the proposed method with the traditional methods to deal with the different levels of changes in multitemporal images in our future work.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China (41271376), in part by Wuhan Science and Technology Program (2013072304010825), and in part by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1278).

References

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23 (11), 1222–1239.

Chavez Jr, P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens. Environ. 24 (3), 459–479.

Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 115 (4), 1053–1064.

Cheng, Q., Shen, H.F., Zhang, L.P., Li, P.X. 2013. Inpainting for remotely sensed images with a multichannel non-local total variation model. IEEE Trans. Geosci. Remote Sens. (doi: 10.1109/TGRS.2012.2237521).

Criminisi, A., Pérez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. 13 (9), 1200–1212.

Elad, M., Starck, J.-L., Querreb, P., Donoho, D.L., 2005. Simultaneous cartoon and texture image inpainting using Morphological Component Analysis (MCA). Appl. Comput. Harmonic Anal. 19 (3), 340–358.

Gilboa, G., Osher, S., 2008. Nonlocal operators with applications to image processing. Multiscale Model. Simul. 7 (3), 1005–1028.

Gladkova, I., Grossberg, M.D., Shahriar, F., Bonev, G., Romanov, P., 2012. Quantitative restoration for MODIS band 6 on aqua. IEEE Trans. Geosci. Remote Sens. 50 (99), 2409–2416.

He, K., Sun, J., 2012. Statistics of patch offsets for image completion. In: Computer Vision–ECCV 2012, Firenze, Italy, Springer, pp. 16–29.

Helmer, E., Ruefenacht, B., 2005. Cloud-free satellite image mosaics with regression trees and histogram matching. Photogramm. Eng. Remote Sens. 71 (9), 1079.

Kwatra, V., Schödl, A., Essa, I., Turk, G., Bobick, A., 2003. Graphcut textures: image and video synthesis using graph cuts. ACM Transactions on Graphics (TOG) 22 (3), 277–286.

Lee, S., Crawford, M., 1991. An adaptive reconstruction system for spatially correlated multispectral multitemporal images. IEEE Trans. Geosci. Remote Sens. 29 (4), 494–508.

Li, H., Zhang, L., Shen, H., Li, P., 2012. A variational gradient-based fusion method for visible and SWIR imagery. Photogramm. Eng. Remote Sens. 78 (9), 947–958.

Li, M., Liew, S.C., Kwoh, L.K., 2003. Producing cloud free and cloud-shadow free mosaic from cloudy IKONOS images. In: 2003 IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, IEEE, vol. 6, pp. 3946–3948.

Li, S.Z., 2001. Markov random field modeling in image analysis. Springer.

Lin, C.H., Tsai, P.H., Lai, K.H., Chen, J.Y., 2013a. Cloud removal from multitemporal satellite images using information cloning. IEEE Trans. Geosci. Remote Sens. 51 (1), 232–241.

Lin, C. H., Lai, K. H., Chen, Z. B., Chen, J. Y., 2013b. Patch-Based Information Reconstruction of Cloud-Contaminated Multitemporal Images. IEEE Trans. Geosci. Remote Sens. (doi: 10.1109/TGRS.2012.2237408).

Liu, D., Kelly, M., Gong, P., 2006. A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. Remote Sens. Environ. 101 (2), 167–180.

Liu, D., Song, K., Townshend, J.R., Gong, P., 2008. Using local transition probability models in Markov random fields for forest change detection. Remote Sens. Environ. 112 (5), 2222–2231.

Lorenzi, L., Melgani, F., Mercier, G., 2011. Inpainting strategies for reconstruction of missing data in VHR images. IEEE Geosci. Remote Sens. Lett. 8 (5), 914–918.

Maalouf, A., Carré, P., Augereau, B., Fernandez-Maloigne, C., 2009. A bandelet-based inpainting technique for clouds removal from remotely sensed images. IEEE Trans. Geosci. Remote Sens. 47 (7), 2363–2371.

Melgani, F., 2006. Contextual reconstruction of cloud-contaminated multitemporal multispectral images. IEEE Trans. Geosci. Remote Sens. 44 (2), 442–455.

Melgani, F., Serpico, S.B., 2003. A Markov random field approach to spatio-temporal contextual image classification. IEEE Trans. Geosci. Remote Sens. 41 (11), 2478–2487.

Pritch, Y., Kav-Venaki, E., Peleg, S., 2009. Shift-map image editing. In: 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, IEEE, pp. 151–158.

Rakwatin, P., Takeuchi, W., Yasuoka, Y., 2009. Restoration of aqua MODIS band 6 using histogram matching and local least squares fitting. IEEE Trans. Geosci. Remote Sens. 47 (2), 613–627.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens. Environ. 112 (6), 3112–3130.

Rossi, R.E., Dungan, J.L., Beck, L.R., 1994. Kriging in the shadows: geostatistical interpolation for remote sensing. Remote Sens. Environ. 49 (1), 32–40.

Salberg, A.B., 2011. Land cover classification of cloud-contaminated multitemporal high-resolution images. IEEE Trans. Geosci. Remote Sens. 49 (1), 377–387.

Shen, H., Wu, P., Liu, Y., Ai, T., Wang, Y., Liu, X., 2013. A spatial and temporal reflectance fusion model considering sensor observation differences. Int. J. Remote Sens. 34 (12), 4367–4383.

Shen, H., Zeng, C., Zhang, L., 2011. Recovering reflectance of AQUA MODIS band 6 based on within-class local fitting. IEEE J. Selected Top. Appl. Earth Observ. Remote Sens. 4 (1), 185–192.

Shen, H., Zhang, L., 2009. A MAP-based algorithm for destriping and inpainting of remotely sensed images. IEEE Trans. Geosci. Remote Sens. 47 (5), 1490–1500.

Stein, D.W., Beaven, S.G., Hoff, L.E., Winter, E.M., Schaum, A.P., Stocker, A.D., 2002. Anomaly detection from hyperspectral imagery. IEEE Signal Process Mag. 19 (1), 58–69.

Storey, J., Scaramuzza, P., Schmidt, G., Barsi, J. 2005. Landsat 7 scan line corrector-off gap-filled product development. In: Proceedings of Pecora 16 Global Priorities in Land Remote Sensing, Sioux Falls, South Dakota, October 23–27.

Tseng, D.C., Tseng, H.T., Chien, C.L., 2008. Automatic cloud removal from multi-temporal spot images. Appl. Math. Comput. 205 (2), 584–600.

Van der Meer, F., 2012. Remote-sensing image analysis and geostatistics. Int. J. Remote Sens. 33 (18), 5644–5676.

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

Zhang, X., Qin, F., Qin, Y., 2010. Study on the thick cloud removal method based on multi-temporal remote sensing images. In: 2010 International Conference on Multimedia Technology (ICMT), Ningbo, China, IEEE, pp. 1–3.

Zhang, Y., Guindon, B., Cihlar, J., 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. Remote Sens. Environ. 82 (2), 173–187.

Zhu, X., Gao, F., Liu, D., Chen, J., 2012. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images. IEEE Geosci. Remote Sens. Lett. 9 (3), 521–525.