Dead Pixel Completion of Aqua MODIS Band 6 Using a Robust M-Estimator Multiregression
Xinghua Li, Huanfeng Shen, Senior Member, IEEE, Liangpei Zhang, Senior Member, IEEE, Hongyan Zhang, and Qiangqiang Yuan

Abstract—The Earth Observing System of the National Aeronautics and Space Administration pays a great deal of attention to the long-term global observations of the land surface, biosphere, atmosphere, and oceans. Specifically, the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the twin satellites Terra and Aqua plays a vital role in the mission. Unfortunately, around 70% of the detectors in Aqua MODIS band 6 have malfunctioned or failed. Consequently, many of the derivatives related to band 6, such as the normalized difference snow index, suffer from the adverse impact of dead or noisy pixels. In this letter, the missing or noisy information in Aqua MODIS band 6 is successfully completed using a robust multilinear regression (M-estimator) based on the spectral relations between working detectors in band 6 and all the other spectra. The experimental results indicate that the proposed robust M-estimator multiregression (RMEMR) algorithm can effectively complete the large areas of missing information while retaining the edges and textures, compared to the state-of-the-art methods.

Index Terms—Aqua, band 6, inpainting, Moderate Resolution Imaging Spectroradiometer (MODIS), reflectance.

I. INTRODUCTION

The Earth Observing System (EOS) contributes to the improved understanding of the Earth as a whole group by means of long-standing global and joint observations of the land, atmosphere, ocean, and ecosphere. The Terra (EOS AM) and Aqua (EOS PM) satellites are both outstanding forerunners of the mission. Aqua’s afternoon observations combined with Terra’s morning observations have the ability to provide important insights into the daily cycles of global precipitation and ocean circulation. Their Moderate Resolution Imaging Spectroradiometer (MODIS) sensors undertake continuous monitoring of the land, ocean, and lower atmosphere every 1–2 days, with a 16-day revisit cycle, by 36 spectral bands ranging from blue to thermal infrared at three different nadir spatial resolutions: 250 m (bands 1 and 2), 500 m (bands 3–7), and 1000 m (bands 8–36) [1]. In terms of the three resolutions, there are 40, 20, and 10 detectors in the corresponding channels [2]. In addition to the short cycle time, another remarkable advantage of MODIS is its free data access. As a result, MODIS has been broadly applied in flood disaster monitoring [3], short-term weather forecasting [4], scientific research, and so on.

Aqua MODIS band 6, in particular, is a spectrum of great importance for many quantitative applications. Taking its reflectance, for example, band 6 can be used for the calculation of the normalized difference snow index (NDSI) [5], the normalized difference infrared index (NDII) [6], and aerosol optical depths (AODs) [7]. Unfortunately, detectors 2, 5, 6, 10, 12–16, and 18–20 do not work properly, and detectors 4 and 17 are affected by serious noise, which means that the data are periodically missing or are so noisy as to be considered unusable [8].

As discussed previously, recovering the large-scale missing or noisy information of Aqua MODIS band 6 is of great significance. A number of researchers have made a contribution to this issue. Wang et al. [2] retrieved the reflectance of Aqua MODIS band 6 using polynomial regression based on the relationship between bands 6 and 7. On this basis, Rakwatia et al. [1] proposed the histogram matching and local least squares fitting (HMLLSF) algorithm, and Shen et al. [9] proposed the within-class local fitting (WCLF) algorithm. They both made improvements to the data preprocessing stage. Additionally, Shen et al. [10] effectively recovered the bad band based on adaptive spectrum-weighted sparse Bayesian dictionary learning with other six bands; however, compared to those of the other methods, the computational cost and complexity are very high. Gladkova et al. [8] utilized the other six bands as supplementary information to band 6 according to multiregression, which is called quantitative image restoration (QIR). QIR gets better results than the off-the-shelf methods using fewer bands; however, some redundant information can be introduced.

In this letter, we propose a simple and robust multiregression according to the Huber M-estimator to complete the missing and noisy pixels in Aqua MODIS band 6 (for brevity, hereafter, we call this algorithm RMEMR), which makes the best use of the spectral correlations between band 6 and the other good bands. This algorithm can adaptively weight the fitting data, depending on the size of the residuals, instead of manually removing the outliers. Taking advantage of the Huber M-estimator, the result is both robust and effective.

The remainder of this letter is organized as follows. Section II describes both the previous and RMEMR algorithms, followed by the experimental results and quantitative evaluations and analyses in Section III. Finally, Section IV gives a summary and the conclusion for this research.

Manuscript received July 20, 2013; accepted August 12, 2013. This work was supported in part by the 973 program under Grant 2013AA12A301 and the 863 program under Grant N2011CB707703.

X. Li and H. Shen are with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China (e-mail: lixinghua5540@sina.com; shenhf@whu.edu.cn).

L. Zhang and H. Zhang are with the Remote Sensing Group, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp02@public.wh.hb.cn; zhanghongyan@whu.edu.cn).

Q. Yuan is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China (e-mail: yqiangu86@gmail.com).

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/LGRS.2013.2278626
II. METHODOLOGY

To date, there has not been a lot of research into recovering the incomplete information of Aqua MODIS band 6. Although the official National Aeronautics and Space Administration (NASA) Web site has released an algorithm called bilinear interpolation to resolve this serious problem, the ensuing result is unacceptable, with many obvious distortions. In fact, the method proposed by Wang et al. [2], based on the relation between bands 6 and 7, lays a solid foundation for most of the existing algorithms. Gladkova et al. pointed out that the other bands are able to provide additional information for restoring the damaged band 6 [8]. Accordingly, we make a robust generalization of the restoration function, depending on the correlations between band 6 and all the other six bands. Our proposed method consists of data preprocessing and completion stages (shown in Fig. 1), which are described as follows.

A. Data Preprocessing

In general, other than band 6 of MODIS, some quality problems also exist in the other bands, including nonexistent pixels, odd–even pixels (also called subframe stripes in [11]), and stripes. The nonexistent pixels mean what have empty values or not a number. The odd-even pixel problem indicates an offset between the neighboring two columns/samples [see Fig. 3(a)]. Additionally, stripes are also a common phenomenon, as shown in Fig. 4(a). These quality problems have a harmful effect on the accuracy of the reconstruction. Therefore, the data preprocessing stage aimed at these issues is required.

First, check out the nonexistent pixels in the other bands. Once a nonexistent pixel is detected by the values themselves, it is replaced by the average of the neighborhood with an adaptive size. However, supposing that the suitable size is not found successfully when the size is larger than a threshold, this pixel will be discarded temporarily until other nonexistent pixels are substituted successfully.

In this second step, the odd-even pixel problem is resolved. This issue is, in fact, the pixel difference or offset between the adjacent odd sample and the even sample in the charge-coupled device (CCD), as described in [12]. It is corrected by the statistics of the even sample minus the neighboring two odd samples in each CCD [13]. As shown in Fig. 2, the difference \( \text{Dif}_{2i} \) between even and odd samples in each block is calculated by

\[
\text{Dif}_{2i} = \frac{M_{2i} - M_{2i-1} + M_{2i+1}}{2}, \quad i \in N^+ \tag{1}
\]

where \( M_{2i} \) represents the mean value of the even sample \( 2i \) in each block (\( n \) lines) and \( M_{2i-1} \) and \( M_{2i+1} \) denote the same meaning for the odd samples. In general, we assume that the even sample will be revised. \( S_{2i} \), the pixel values of sample \( 2i \) in each block, is then corrected as follows:

\[
S_{2i} = S_{2i} - \text{Dif}_{2i}, \quad i \in N^+. \tag{2}
\]

Fig. 3 shows the contrast of odd–even pixel correction based on this method with one-line blocks. It can be seen that this correction is an effective method. Nevertheless, under this correction (mainly for narrow blocks), some small individual objects may be cleared away or introduced over heterogeneous areas. Fortunately, this phenomenon is not serious.

The third step is destriping. Fig. 4(a) shows two kinds of stripes in MODIS data. “A” represents one-line stripes, and “B” represents more-than-one-line stripes. The latter are common in a homogeneous area (e.g., desert and sea). Recently, there have been a lot of destriping algorithms which get effective results. For simplicity, the simple methods are used in our case. We undertake destriping with different methods according to the types. First, the one-line stripes are replaced by the average of the two lines which are closest to the stripe line. Although the averaging method is not sufficient for quantitative restoration, it is just a preprocessing step for fitting samples. On the other hand, the more-than-one-line stripes (if they exist) are destriped by wavelet–Fourier filtering [14]. Note that MODIS band 6 is not processed in this step. Fig. 4 shows the destriping outcome.
B. Completion Algorithm

After the preprocessing stage, the completion stage begins. In terms of Aqua MODIS L1B 500-m-resolution data, we call band 6 the out-of-order band and the other six bands the working bands. Furthermore, in the out-of-order band, the pixels corresponding to the normal running detectors belong to $\Omega$, and the remainder belong to $\Omega^-$. Let $R_{i,j,k}$ be the gray intensity (e.g., reflectance) in line $i$, sample $j$, and band $k$.

The original image is then divided into a series of $20 \times 20 \times 7$ overlapping cubic patches $\{1, 2, \ldots, p, \ldots, P\}_{p=1}^P$, sliding from left to right and from up to down with a fixed step size. In our experiments, usually, it is ten. In every patch, all the pixels in the out-of-order band have an identical function relationship to those of the working bands, as follows:

$$R_{i,j,6} = f_p(R_{i,j,1}, R_{i,j,2}, R_{i,j,3}, R_{i,j,4}, R_{i,j,5}, R_{i,j,7}) + \varepsilon_{i,j},$$

where $f_p(\cdot)$ means the function relationship, $\varepsilon_{i,j}$ represents the residual, and $\Omega_p$ and $\Omega^p$ are the corresponding $\Omega$ and $\Omega^-$ in patch $p$. The reason why we choose the patch size of $20 \times 20 \times 7$ is that there are 20 detectors in total in the out-of-order band, and the correlation is closer in most cases of this patch size. Afterward, the pixels of the working detectors are exploited as the training samples for exploring the function in the given patch. In other words, there exists an independent function in every patch. At the same time, in the traditional framework of the least squares solver, the sum of squared residuals $\varepsilon_{i,j}$ of every patch should be minimized as

$$\min \sum_{i,j} \left(R_{i,j,6} - \hat{R}_{i,j,6}\right)^2, \quad (i,j) \in \Omega_p$$

where $\hat{R}_{i,j,6}$ represents the estimated value of $R_{i,j,6}$. In general, there may be some outliers in the fitting samples which have an adverse effect on the regression model. Aiming at this imperfection, the weighted least squares estimator is introduced to weaken the influence of the outliers and strengthen the robustness of the model. The basic idea is to minimize the sum of weighted and squared residuals $\varepsilon_{i,j}$ of every patch of the form

$$\min \sum_{i,j} w_{i,j} \left(R_{i,j,6} - \hat{R}_{i,j,6}\right)^2, \quad (i,j) \in \Omega_p$$

where $w_{i,j}$ denotes the corresponding weight. Briefly speaking, $w_{i,j}$ trades off the importance of every fitting sample in the current patch. The larger the residual is, the smaller the weight is. In our method, $w_{i,j}$ is calculated by the Huber M-estimator [15]

$$w_{i,j} = \begin{cases} 1, & |u_{i,j}| \leq c \\ \frac{c}{|u_{i,j}|}, & |u_{i,j}| > c \end{cases}$$

where $c$ is a given segmentation threshold (usually 1.345) and $u_{i,j}$ is a normalized standard index of the residual defined by

$$u_{i,j} = \frac{r_{i,j}}{s} = \frac{r_{i,j}}{1.48 \times \text{med}(|r_{i,j} - \text{med}(r_{i,j})|)}$$

where $s$ signifies the scale of the residual (we choose it as in [16]) and $\text{med}(\cdot)$ represents the median. The weights are iteratively adapted until their variations are less than a given tolerance of 0.0001. Due to the adaptive weights of every fitting sample, the robustness of the multiregression is effectively enhanced. The minimizing objective function of (5) also corresponds to

$$\min \sum_{i,j} \rho(r_{i,j}), \quad (i,j) \in \Omega_p$$

$$\rho(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq c \\ c|x| - \frac{c^2}{2}, & |x| > c \end{cases}$$

where $c$ is a constant equivalent to (6). In other words, (6) is the weight function, and (9) is the objective function.

Finally, with the help of the aforementioned fitting, the functions $f_p(\cdot)$ that will be applied to the pixels in $\Omega_p^-$, according to (11), are gained. In other words, once the functions are fitted, the information in $\Omega^-$ is restored

$$\hat{R}_{i,j,6} = f_p(R_{i,j,1}, R_{i,j,2}, R_{i,j,3}, R_{i,j,4}, R_{i,j,5}, R_{i,j,7}), \quad (i,j) \in \Omega_p^-.$$  

In fact, $f_p(\cdot)$ has various forms. Generally, the multilinear regression in (12) gets the stable and satisfactory result with low time cost and computational cost. In practical applications, it is the first choice without any beforehand checkpoints

$$R_{i,j,6} = \sum_k a_k R_{i,j,k} + b, \quad k = 1, 2, 3, 4, 5, 7; \quad (i,j) \in \Omega_p^-.$$  

As far as every patch is concerned, the previous process is undertaken repeatedly. Furthermore, in the whole image, it is a sliding process. Once a missing pixel gets into the current sliding patch, it will be restored by the independent fitting function. In general, a missing pixel may be completed many times according to different patches. The final result will be the average of all the completion results. The times have a big deal with its specific location in the image. Generally speaking, it increases from the edge to the center and remains unchanged while reaching the maximum. In addition, the shorter the sliding step size is, the higher the number is and the higher the operation cost is.
III. EXPERIMENTS

In order to comprehensively evaluate the advantages and disadvantages of the proposed algorithm over the state-of-the-art algorithms, synthetic and real experiments are conducted in this section. For simplicity, we just process the reflectances of MODIS L1B 500-m-resolution data, downloaded from the NASA Web site and not georectified. Note that the bow tie effect is not our interest, which can be kept and independently processed after our completion. In the absence of any special instructions, all images are of the size of $400 \times 400 \times 7$, and the sliding step size is ten pixels. The experimental platform is a personal computer whose CPU is Intel at 3.4 GHz and whose memory is 8 GB. The qualitative and quantitative evaluations of the experimental results are given as follows.

A. Synthetic Experiments

The MODIS sensors on board the Aqua and Terra satellites have analogous design patterns, which makes it possible to consider the recovery result of Terra MODIS as the approximate evaluation approach. As stated previously, Aqua MODIS band 6 has dead and noisy detectors; for brevity, we also consider the noisy detectors as dead ones and set the values as zeros. To get a quantitative evaluation, the peak signal-to-noise ratio (PSNR) [10], structural similarity (SSIM) index [17], and mean absolute difference (MAD) are selected as the indicators. The results include the synthetic recoveries of Terra MODIS band 6 by the four methods: HMLLSF [1], WCLF [9], QIR [8], and the proposed RMEMR. From the qualitative and visual perspective, these approaches can achieve impactful outcomes with indistinguishable dissimilarities. As a result, we primarily make the quantitative evaluations as shown in Table I. Fig. 5 shows the results of RMEMR. Note that RMEMR formulates a number of linear functions using (12), so as to reasonably compare QIR and RMEMR in this experiment. In terms of the three indices, Table I shows that QIR gets a better result than HMLLSF and WCLF. This is because the latter two methods just utilize the information of band 7 to restore the missing pixels of band 6; however, QIR exploits the extra five bands in the given granule. That is to say, the more the useful complementary information is, the better the result is. In addition, RMEMR significantly outperforms QIR. Although QIR takes the spectral correlations between band 6 and all other bands, and the spatial correlations within every band, into consideration, it can also draw some redundant information into the recovery model. At the same time, the simple least squares fitting cannot effectively counteract the influence of existing outliers, and it gives rise to an inferior recovery to RMEMR. With the robust M-estimator, RMEMR outperforms QIR. Nevertheless, compared to HMLLSF and WCLF, this also brings additional time cost for determining the weights. This experiment verifies that RMEMR is a more effective instrument to resolve the problem of Aqua MODIS band 6.

To further perceive these algorithms intuitively, we show the scatter plots between the synthetic recoveries and the originals in Fig. 5, as shown in Figs. 6 and 7, respectively. They show that the scatter plots of HMLLSF and WCLF are similar to each other and are distributed more discretely from the diagonal than QIR and RMEMR. In terms of Fig. 6, section A in Fig. 6(c) is more discrete than that in Fig. 6(d); in contrast, section B in Fig. 6(c) is a little more concentrated. In fact, section A represents the low-value region (dark region) in Fig. 5(b), and section B represents the region of high value. In the low-value region, the reflectance approximates to zero and varies slightly, and the proposed method cannot get a better result; however, in the higher reflectance region, the proposed method reveals an advantage, which is also demonstrated in Fig. 7 (the majority of which is a high-value region). Moreover, Fig. 7(d) is a little more concentrated than Fig. 7(c) in sections C and D, which are both higher value regions, which means that the RMEMR recovery has closer correlations than QIR. Basically, the scatter plot results are consistent with the results shown in Table I.
real experiments, due to the absence of the ideal Aqua MODIS band 6 data.

IV. Conclusion

This letter has proposed a robust multiregression algorithm using a Huber M-estimator (RMEMR) to complete the dead and noisy pixels of Aqua MODIS band 6 by revealing the latent spectral correlation between band 6 and all the other working bands (for L1B 500-m-resolution data). The experimental results demonstrate that RMEMR outperforms the state-of-the-art methods, as confirmed by the objective measurements of SSIM, MAD, and PSNR in the simulated experiments. It is a competitive algorithm that can get improved quantitative results, which will contribute to the quantitative applications of MODIS, such as the calculation of NDSI, NDII, and AOD. However, as far as the other quality problems of MODIS granules and adaptive spectral weights of other bands are concerned, they are deserving of further consideration.

REFERENCES