PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, **120**:814–820, 2008 June

Detection of Cosmic-Ray Hits for Single Spectroscopic CCD Images

ZHANGQIN ZHU AND ZHONGFU YE[1]

Institute of Statistical Signal Processing, University of Science and Technology of China, Hefei 230027, China

Received 2008 March 31; accepted 2008 May 19; published 2008 June 30

ABSTRACT. A new method of detecting cosmic-ray hits from single spectroscopic CCD images is proposed in this paper. First, one-order difference operations in two directions respectively are calculated to detect the candidate points of the cosmic-ray hits by using two thresholds. Second, Bessel curve fitting is adopted to calculate the deviations of the candidate points so as to find the confirmation points of cosmic-ray hits. At the end of this paper, five groups of simulation data and one group of observation data from SDSS are used to demonstrate the effectiveness of the proposed method.

1. INTRODUCTION

The cosmic-ray hits on CCD spectroscopic images have disturbing effects for processing those images. For example, the cosmic-ray hits will weaken the efficiency of compression algorithms on astronomical images (Offenberg, et al. 1999), disturb the operations of autonomous astronomical pipelines (Axelord, et al. 2004; Becker, et al. 2004), and lead to a trouble to the next processing for the 2D spectral data. Therefore, detecting and removing the cosmic-ray hits form spectroscopic CCD images is an important problem.

An effective method was proposed that combines multiple images with the same objects to detect and remove the cosmic-ray hits (Shaw & Horne 1992). Later, a developed method with combining multiple images was used for the *Hubble Space Telescope* (Windhorst et al. 1994, Freudling 1995, Fruchter & Hook 1997). Both methods are based on the positions of cosmic-ray hits, which are changed randomly at each exposure. So these methods can detect the positions of the cosmic-ray hits and remove them fully by combining multiple images of different exposures, and they are widely used for processing images with multi-targets such as the Sloan Digital Sky Survey (SDSS). These methods have some strong constraints that the intensities of sky lines and object spectra must remain unchanged in a short time.

In many conditions, multiple images with the same objects cannot be obtained because of a change in the environments or a lack of adequate time. In fact, when the environment changed over time, the positions and intensities of sky lines and object spectra may be altered (Croke 1995).

Many researchers have noticed the importance of detecting cosmic-ray hits from signal images, and some algorithms have been proposed in recent years. A convolution-based algorithm was proposed (Rhoads 2000), in which a spatial filter was used to identify the positions of cosmic-ray hits on CCD images with the help of a threshold. However, the noise in a CCD image weakens the effectiveness obviously. Van Dokkum used Laplacian edge detection to detect the cosmic-ray hits (Van Dokkum 2001). The method gets a good result except for some cosmic-ray hits with lower energies. To reduce the running time, a fast algorithm based on the histogram of the image data was proposed (Pych 2004). This method does not need any models of objects or high signal-to-noise ratio (S/N), but some spectrum energies of some pixels may be removed incorrectly. Shamir presented an algorithm based on fuzzy logic modeling of human intuition (Shamir 2005), and the algorithm determines whether one is cosmic-ray hit or not based on human sentience. A two-stage scheme method was proposed to detect and remove the cosmic-ray hits (Zhang et al. 2007). Good detecting results were obtained, but the algorithm was very complex.

Based on the project of the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), a new algorithm is proposed in this paper to detect the cosmic-ray hits from single spectroscopic CCD images. First, one-order difference operations in x and y coordinates are done to get the candidate points and part confirmation points of cosmic-ray hits. Then, Bessel curve fitting is used to obtain the confirmation points from the candidate points.

The models of the spectra and the cosmic-ray hits are described in § 2. In § 3, our method is introduced in detail. Experiments and results are given in § 4, and an application of our method for one group of observed data of SDSS is also given in this section. The summary is presented at the end of the article.

2. THE MODELS OF SPECTRA AND COSMIC-RAY HITS

The spectra images contain much information, such as sky lines, object spectra, noise of system and external environments, as well as cosmic-ray hits. In this paper, we assume that the

[1] Corresponding author; yezf@ustc.edu.cn.

images contain only three kinds of signals, cosmic-ray hits, spectra which includes the sky lines and object spectra, and noise. Therefore, the model of the spectroscopic CCD image can be described as follows:

$$I = O + C + N, \tag{1}$$

where O is the spectra, C is the cosmic-ray hits on CCD image, and N is the noise. Furthermore, we assume the values of the noise are much smaller than the spectra, so the noise is neglected in theoretical analysis.

The astronomical spectra images are the light through the telescope and other instruments, and projecting on the CCD finally. However, the cosmic-ray hits are the high-energy particles in the universe hitting on the CCD directly. Both projecting process can be described as follows:

$$O = o(x, y) \times f(x, y), \qquad C = c(x, y) \times g(x, y), \tag{2}$$

where $o(x, y)$ is the energy of one spectra point before projecting on the CCD, (x, y) is its projecting center position on the CCD, $f(x, y)$ is the 2D point-spread function of spectra; $c(x, y)$ is the energy of one cosmic-ray hit point before projecting on the CCD, (x, y) is its projecting center position on CCD, $g(x, y)$ is the 2D point-spread function of cosmic-ray hits.

The differences between spectra and cosmic-ray hits can be described as follows.

1. In general, the pixels that are covered by cosmic-ray hits have larger energies than neighbors that are only spectra, because these pixels have the energies both of the spectra and the cosmic-ray hits. This character is used to get the candidates points from the image I.

2. The point-spread function of spectra is usually a Gaussian function (Rhoads 2000), but the point-spread function of cosmic-ray hits isnon-Gaussian with sharp contours. The differences of their point-spread functions are the basis to obtain the confirmation points of cosmic-ray hits from the candidate points.

As described in the differences between spectra and cosmic-ray hits, the energy of spectra projecting on CCD is not a point, but it actually expands to its adjacent pixels in the form of a point-spread function. The point-spread function of spectra is a Gaussian function, which is described as

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}}, \qquad \sigma = \frac{\mathrm{FWHM}}{2\sqrt{\ln 4}}, \tag{3}$$

where (x_0, y_0) is the center point of the point-spread function, σ^2 is the variance and $fwhm$ is the full width at half-maximum of the Gaussian function.

The point-spread function of the cosmic-ray hits, $g(x, y)$, is a non-Gaussian function. The cosmic-ray hit projects on the CCD maybe in the form of a point or a line or a spot with sharp contours. In Figure 1, two slices of the spectra and the cosmic-ray hits from a 2D simulation image in spatial orientation are given.

The slices of spectra and cosmic-ray hits in Figure 1 are from a 2D simulation image with many fibers that contains noise and cross-contamination between two neighboring fibers. From these slices, the spectra are approximate Gaussian distributions instead of Gaussian distributions due to the existing of noise, and the cosmic-ray hits are non-Gaussian distribution with sharper contours.

Based on the differences between spectra and cosmic-ray hits in both energies and the forms of the point-spread functions, most of the cosmic-ray hits can be detected from a CCD spectroscopic image.

3. THE DETECTION OF COSMIC-RAY HITS

3.1. Detection of Candidate Points and Some Confirmation Points

Generally, the pixels that are covered by cosmic-ray hits have two kinds of energies, both the cosmic-ray hits and spectra, whose values are larger than those of the neighbors with only spectra energies. So one-order difference operations in x and y coordinates are used to obtain the candidate points of the cosmic-ray hits by comparing the results of difference with thresholds. The one-order difference operations in x and y coordinates are described as follows:

$$\begin{aligned} \Delta o_r(x, y) &= O(x, y) - O(x, y-1), \\ \Delta o_c(x, y) &= O(x, y) - O(x-1, y), \end{aligned} \tag{4}$$

where $O(x, y)$ is the value of the original image at point (x, y).

After the operations of difference on the original image, two new images, O_r and O_c, are obtained by

$$O_r(x, y) = \Delta o_r(x, y), \qquad O_c(x, y) = \Delta o_c(x, y). \tag{5}$$

For all the pixels of one original CCD image, two thresholds T_1 and T_2 are used here to divide them into the set of candidate points, the set of part confirmation points, and the set of spectra. The choice of the values T_1 and T_2 has an effect on the running time and the effectiveness of the algorithm. An approximate method will be given to estimate the values of the two thresholds later.

We assume that T_1 is larger than T_2 and that the values of the two thresholds are under the following constraints:

1. If $O_c(x, y) > T_1$ or $O_r(x, y) > T_1$, then $O_c(x, y)$ or $O_r(x, y)$ is covered by a cosmic-ray hit point. The corresponding pixel is a confirmation point of cosmic-ray hits, and the corresponding position (x, y) is labeled as the confirmation point.

2. If $O_c(x, y) < T_2$ or $O_r(x, y) < T_2$, then $O_c(x, y)$ or $O_r(x, y)$ is a spectrum point. The corresponding pixel is not be detected in the next steps, and it is not labeled.

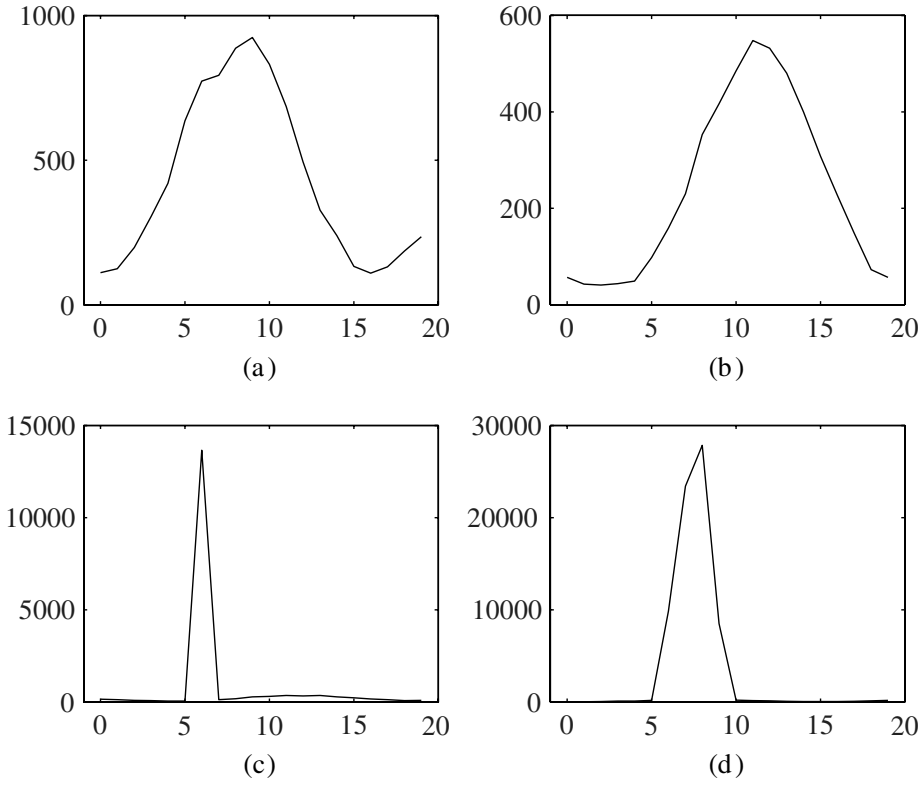

FIG. 1.—Slices of spectra and cosmic-ray hits in spatial orientation.

3. If $T_2 \le O_c(x, y) \le T_1$ or $T_2 \le O_r(x, y) \le T_1$, then $O_c(x, y)$ or $O_r(x, y)$ is a candidate point of cosmic-ray hits and the corresponding position is labeled as the candidate point.

After comparing with the two thresholds, the difference images O_c and O_r can be divided into three sets. The next step is combing the two group sets as follows:

For every pixel:

1. If it belongs to both the set of part confirmation points determined by O_c and the set of part confirmation points determined by O_r, it is considered a confirmation point of the cosmic-ray hits.

2. If it belongs to both the set of spectra points determined by O_c and the set of spectra points determined by O_r, it is considered a spectrum point.

3. The remaining points are considered the candidate points of the cosmic-ray hits.

It is obvious that processing once is enough to detect a one-pixel cosmic-ray hit from the spectroscopic CCD image. If the cosmic-ray hit is a curve or a line, processing once is also enough. If the cosmic-ray hit is a spot, the processing for obtaining the candidate points is more complicated. After processing once, some confirmation points and some candidate points have been detected, and the values of these points are substituted by the averages of their neighbors that are not detected as the confirmation points or the candidate points. Then a new image O_{T_2} is obtained. The image O_{T_2} is treated as the new original image and it will be processed repeatedly. The iterative processing must be done enough to ensure that all the cosmic-ray hits can be detected.

However, the one-order difference operations in two orientations not only can detect the cosmic-ray hits but also mistake some spectra points for candidate points. One simple detection result is shown in Figure 2. Considering this situation, obtaining the confirmation points of cosmic-ray hits from the candidate points is necessary.

3.2. Detection of Confirmation Points from the Candidate Points

The candidate points of cosmic-ray hits have been obtained by the use of the algorithm of one-order difference with two thresholds. In this section, an algorithm will be introduced to obtain the confirmation points of cosmic-ray hits from the set of the candidate points. We know that the point-spread function of cosmic-ray hits, $g(x, y)$, is different from the point-spread function of spectra, and the most important difference is that the cosmic-ray hits have much sharper contours than the spectra.

Bessel curve fitting is a mathematical method to fit out a smooth curve based on several control points, and it can be described as follows:

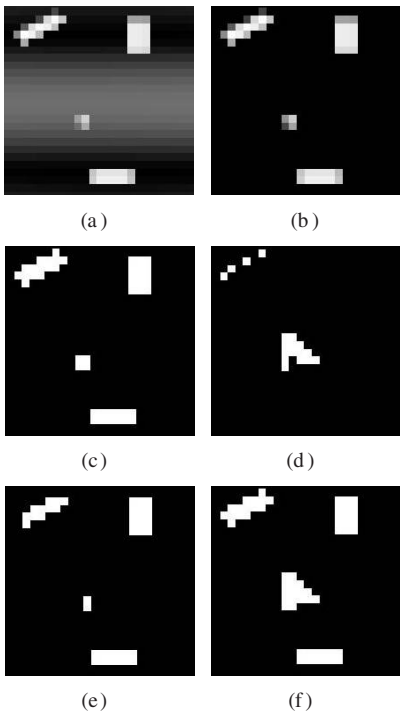

FIG. 2.—Detection of the candidate points and some confirmation points by one-order difference operations. (*a*) The spectra with the cosmic-ray hits. (*b*) The cosmic-ray hits of the original image. (*c*) The binary image of the cosmic-ray hits. (*d*) The binary image of the candidate points. (*e*) The binary image of the confirmation points detected. (*f*) The binary image of the confirmation and the candidate points.

$$p(u) = \sum_{i=0}^{n} p_i C(n,i) u^i (1-u)^{n-i}, \qquad C(n,i) = \frac{n!}{i!(n-i)!}, \tag{6}$$

where p_i is the ith control point, $C(n,i)$ is the polynomial coefficient, $p(u)$ is the curve after Bessel curve fitting, such that $p(0)$ is the first point after curve fitting and $p(1)$ is the last point after curve fitting. The number of the control point is $n+1$.

Bessel curve fitting is used here to detect cosmic-ray hits in the original image. The point-spread function of spectra is a Gaussian function, so it is smooth. Because a Bessel fitting curve based on the points of a smooth curve is also a smooth curve just like the original one, the deviations between the Bessel fitting curve and the spectra are much smaller than those of cosmic-ray hits. The deviation between the original value and the fitting value at (x,y) is defined as

$$D(x,y) = \frac{O(x,y) - O_b(x,y)}{O(x,y)}, \tag{7}$$

where $D(x,y)$ is the deviation value, $O(x,y)$ is the original value, and $O_b(x,y)$ is the new value after Bessel curve fitting.

In order to simplify calculation, a fast algorithm of Bessel curve fitting is used here instead of the original Bessel curve fitting:

$$\begin{cases} p_i^r(u) = (1-u)p_i^{r-1} + up_{i+1}^{r-1} \\ p_i^0(u) = p_i \end{cases}, \qquad r = 0,1,\cdots,n; i = 0,1,\cdots,n-r, \tag{8}$$

where u is a constant between 0 and 1, r is the iterative time, and p_i is the ith control point. Furthermore, when u takes a fixed value of 0.5, $p_i^r(u=0.5)$ is a new control point just at the center between p_i^{r-1} and p_{i+1}^{r-1}.

The Bessel fitting curves by the fast algorithm have some features:

1. The first point and the last point keep the same values after each curve fitting.
2. At each fitting, one new point is determined by the control points. After each fitting, the new points are treated as new control points, and the number of the control points reduces one.
3. The final fitting curve is a smooth curve surrounded by the control points.
4. The number of the control points can be chosen according to actual needs, but it should cover the size of any cosmic ray in spatial orientation. The control points are the basic points of the Bessel curve fitting.

The iteration will end when the control point is only one. The values of the pixels in spatial orientation are Gaussian distribution approximately, so the control points are selected in spatial orientation.

Some examples of the original curves of spectra and cosmic-ray hits and the fitting curves are given in Figure 3.

From Figure 3, we can see that the deviations of cosmic-ray hits are much larger than the deviations of the spectra, so only one threshold is used here to separate the cosmic-ray hits from other candidate points easily. The choice of the threshold is introduced later.

3.3. Summary of Our Algorithm

The whole algorithm proposed in this paper is summarized as follows:

1. The images of one-order difference in x and y coordinates are obtained from the original image.
2. Comparing with the two thresholds, three sets are obtained: the set of candidate points, the set of part confirmation points, and the set of spectra.
3. The values of the points that have been detected are substituted by their neighbors. The image are restored as a new image.
4. Treat the new image as the original image, go to step 1 until enough times of iterative processing.

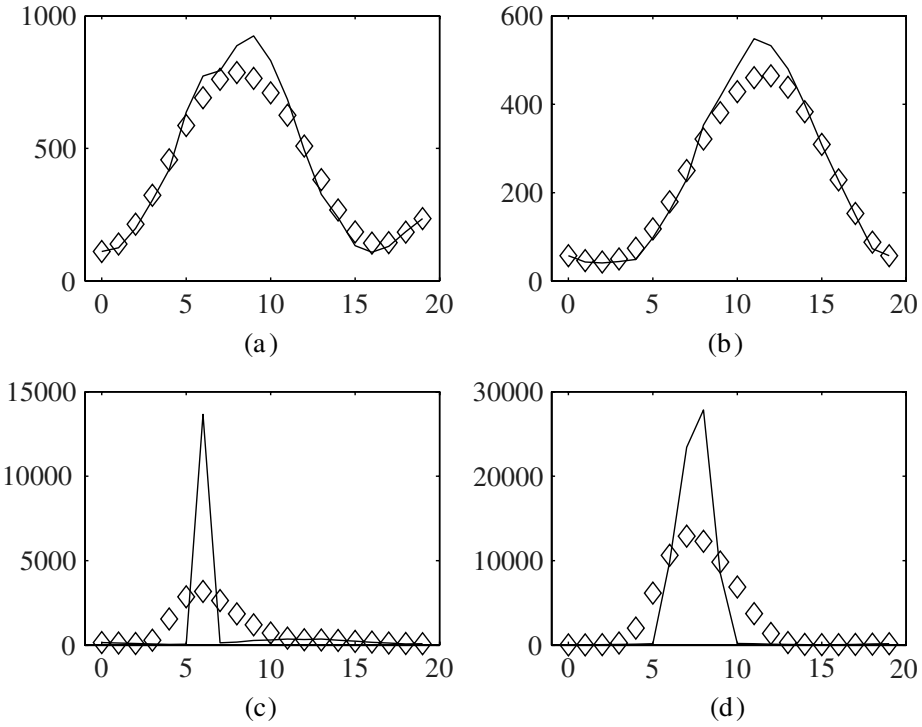

FIG. 3.—Original curves and the Bessel fitting curves. The line shows the original spectra, and the diamond points are the results after Bessel fitting.

5. Get the candidate points and part of confirmation points of cosmic-ray hits. Substitute the values of the part of confirmation points with the average of their neighbors' spectra, and get an image that does not contain the confirmation points detected.

6. Scan each column of the image obtained at step 5 to find the candidate points, and take out several points in spatial orientation continuously as the control points. The candidate point should not be the first point in these control points, because the algorithm keeps the value of the first and the last one.

7. Run the Bessel curve fitting, and obtain new values of these control points.

8. Calculate the deviations of these points, and the candidate points with larger deviations than the threshold are considered as the cosmic-ray hits. Substitute the values of these points with the average of their neighbors' spectra.

9. For the remaining candidate points, repeat the Bessel curve fitting and the calculating of the deviations once, in order to obtain the more confirmation points.

10. Get the resulting image, in which the cosmic-ray hits has been detected and removed.

3.4. Estimation of the Thresholds

Three thresholds are used in our whole algorithm, and the values of the three thresholds will affect the effectiveness of the algorithm. Two approximate methods to obtain the three thresholds are introduced here.

First, the values of T_1 and T_2 are chosen according to the one-order difference image of the original image, which is needed to be processed in x coordinate or in y coordinate. T_1 and T_2 can be estimated just as follows:

First, the number of the positive values of the one-order difference image is accounted. The average of larger 5% pixels of these positive values is $C_{\max}$. The average of the lesser 50% pixels of these positive values is $C_{\min}$:

$$T_1 \approx C_{\max}, \qquad T_2 \approx C_{\min}. \tag{9}$$

Second, the threshold of the deviation that is used to separate the cosmic-ray hits from the candidate points is chosen with an approximate method. The point-spread function of spectra is a Gaussian function with the full width at half maximum is known. The largest deviation of spectra without noise can be estimated. The point with the largest deviation is also the point with the largest energy. The largest approximate deviation is obtained as

$$\begin{aligned} D_{O_{\max}} &= 1 - \frac{1}{4}\{e^{-\frac{[(x_0-1)-x_0]^2}{2\sigma^2}} + 2e^{-\frac{(x_0-x_0)^2}{2\sigma^2}} + e^{-\frac{[(x_0+1)-x_0]^2}{2\sigma^2}}\}, \\ \sigma &= \frac{\mathrm{FWHM}}{2\sqrt{\ln 4}}. \end{aligned} \tag{10}$$

When FWHM is 8 pixels, the value of $D_{O_{\max}}$ is about 0.02 without noise, but the noise can affect the shape of the observation data, so the value of the largest spectra deviation should be increased properly. For the SDSS data, the FWHM is 2 pixels, and the value of $D_{O_{\max}}$ is about 0.25 without noise, which is sufficiently less than the deviations of cosmic-ray hits that cosmic-ray hits can almost be detected. Due to the effect of the noise, if the full width at half-maximum of the point-spread function is narrow, the effectiveness of detection is slightly degraded. When FWHM is 1 pixel, the value of $D_{O_{\max}}$ is about 0.4688, by which it is difficult to separate spectra from cosmic-ray hits correctly.

4. EXPERIMENTS AND RESULTS

To demonstrate the effectiveness of our method, five groups of simulation data and one group of observation data are used in this section. To contrast the experiment results, the Laplacian method (Van Dokkum 2001) is used as a comparing method. To measure the effectiveness of the cosmic-ray hits detection, two statistics, the probability of false alarm and the probability of detection assigned as P_{fa} and P_d, are defined here.

We assume that the number of pixels that are not points of cosmic-ray hits but are detected as cosmic-ray hits is assigned as K. The number of pixels that are points of cosmic-ray hits and detected as the cosmic-ray hits is assigned as M. The number of pixels that are cosmic-ray hits in the original image is N. Then, the probability of false alarm and probability of detection are defined as

$$P_{fa} = K/N, \qquad P_d = M/N.$$

The model of the five groups of simulation data is based on the LAMOST. The parameters of the model are selected according to the measured values or the theoretical values of LAMOST. The apparent magnitudes of object stars are between 16.5 mag and 18.5 mag, and the magnitude of sky is 16 mag. The point-spread function of spectra is a 2D Gaussian function, and its full width at half-maximum is 8 pixels.

The parameters of our method are set as (i) The two thresholds T_1 and T_2 are set to 1300 and 100. (ii) The times of the iterative operation of one-order difference is 6. (iii) The threshold of the deviation is 0.30. (iv) The number of the control points is 20. In order to illustrate clearly, all the figures given in this section are only part figures.

TABLE 1

FIVE GROUPS OF EXPERIMENT RESULTS

Image no.	P_{fa} (Laplacian)	P_d (Laplacian)	P_{fa} (Our method)	P_d (Our method)
1	0.03825	0.73620	0.04022	0.97792
2	0.02630	0.78205	0.05480	1
3	0.01579	0.78829	0.05343	0.99943
4	0.04808	0.75437	0.06643	1
5	0.10804	0.80757	0.10292	0.95110

TABLE 2

STATISTICS CHANGES WITH THRESHOLD OF DEVIATION

Threshold of deviation	P_{fa}	P_d
0.31	0.01072	0.95491
0.33	0.00614	0.94932
0.35	0.00321	0.93796
0.40	0.00073	0.91157

Figure 4 contains one group of the experiment results. More results are given in Table 1.

From Table 1, we can see that the probability of detection of our method is larger than the probability of the Laplacian method, but the probability of false alarm of our method is slightly larger than that of the Laplacian method. The fifth group data in Table 1 shows that when the probability of false alarm is close, the detection of our method is better than the Laplacian method.

To demonstrate the robustness of our method, Table 2 gives the different thresholds of deviation with the same experiment data and the corresponding probability of false alarm and detection.

From Table 2, we can see that when the threshold changes in a small range, the probability of detection almost stays the same. Furthermore, the probability of detection decreases as the probability of false alarm decreases.

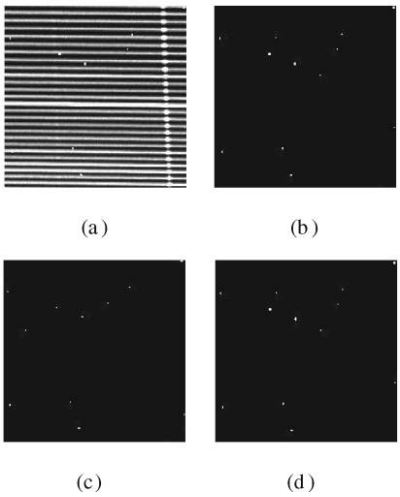

FIG. 4.—Results of one group of simulated data from LAMOST. (*a*) The original image with cosmic-ray hits. (*b*) The cosmic-ray hits of the original image. (*c*) The cosmic-ray hits detected by the Laplacian method. (*d*) The cosmic-ray hits detected by our method.

FIG. 5.—Results of one group of observed data from SDSS. (*a*) The original image. (*b*) The cosmic-ray hits detected by the Laplacian method. (*c*) The cosmic-ray hits detected by our method.

Our method is also applied to one group of observation data of SDSS.[2] Because the full width at half-maximum of SDSS data is 2 pixels, the threshold of deviation is set to 0.50. The results are shown in Figure 5.

Because the distribution of cosmic-ray hits is unknown to the SDSS data, only the results of the detection are given here.

5. CONCLUSION

To detect the cosmic-ray hits from signal spectroscopic CCD images, one new method has been proposed, based on the differences of the contours and the energies between spectra and cosmic-ray hits. One-order difference operations in two orientations are used to get the candidate points of cosmic-ray hits, and Bessel curve fitting is used to get the confirmation points from the candidate points. Then five groups of simulation data and one group of observation data are used in experiments. The results of these experiments show that our method detects better than the Laplacian method.

REFERENCES

Axelrod, T., Connolly, A., Ivezic, Z., Kantor, J., Lupton, R., Plante, R., Stubbs, C., & Wittman, D. 2004, in AAS Meeting 205, 108.11
Becker, A. C., Rest, A., Miknaitis, G., Smith, R. C., & Stubbs, C. 2004, BAAS, 36, 1529
Croke, B. F. W. 1995, PASP, 107, 1255
Freudling, W. 1995, PASP, 107, 85
Fruchter, A. S., & Hook, R. N. 1997, Proc. SPIE, 3164, 120
Offenberg, J. D., Sengupta, R., Fixsen, D. J., Stockman, H. S., Nieto-Santisteban, M. A., Stallcup, S., Hanisch, R. J., & Mather, J. C. 1999, in ASP Conf. Ser. 172, Astronomical Data Analysis Software and Systems VIII , ed. D. M. Mehringer, R. L. Plante, & D. A. Roberts (San Francisco: ASP), 141
Pych, W. 2004, PASP, 116(816), 148
Rhoads, J. E. 2000, PASP, 112(771), 703
Shamir, L. 2005, Astron. Nachr., 326(6), 428
Shaw, R. A., & Horne, K. 1992, in ASP Conf. Ser. 25, Astronomical Data Analysis, Software, and Systems, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes (San Francisco: ASP)
Van Dokkum, P. G. 2001, PASP, 113 (789), 1420
Windhorst, R. A., Franklin, B. E., & Neuschaefer, W. L. 1994, PASP, 106, 798
Zhang, L., Bai, Z. R., Wang, C., & Ye, Z. F. 2007, J. USTC, 37(6), 688–694

[2] SDSS observation data are available at http://das.sdss.org/DR6/data/spectro/ss_25/2045/.