Cosmic Ray Muon Identification Techniques in CCD Images (Postprint)

Authors: Feng Haixia, Chen Jianjun, Deng Jianrong, Zhao Yongheng

Date: 2019-07-30T00:00:00+00:00

Abstract

Cosmic rays in CCD (Charge Coupled Device) images are high-energy particles from outer space that traverse the atmosphere, interact with atmospheric particles to form secondary particles, and ultimately deposit on the CCD. Muons constitute the primary component of atmospheric cosmic rays. To study the properties and variation patterns of these muons, it is first necessary to select muons from the images. This paper presents a method for rapidly and effectively selecting muons from CCD images. The approach employs Laplacian edge detection to extract a list of cosmic ray candidate pixels from CCD images, removes bad pixels and noise, and then utilizes an agglomerative hierarchical clustering algorithm to cluster cosmic ray pixels into cosmic ray events. Feature extraction and classification are performed on these events to select cosmic ray muons. Finally, the selection results are summarized and analyzed.

Full Text

Cosmic-Ray Muon Selection Technology in CCD Images

Feng Haixia[1,2,3], Chen Jianjun[1,3*], Deng Jianrong[1], Zhao Yongheng[1,2,3]

[1]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Abstract: Cosmic rays in CCD (Charge Coupled Device) images are high-energy particles from outer space that traverse the atmosphere, interact with atmospheric particles to form secondary particles, and ultimately strike the CCD detector. Muons constitute the primary component of atmospheric cosmic rays. To investigate the properties and variation patterns of these muons, it is first necessary to identify them within the images. This paper presents a rapid and effective method for selecting muons from CCD images. The approach employs Laplacian edge detection to extract candidate cosmic-ray pixel lists from CCD images, removes bad pixels and noise, and then uses agglomerative hierarchical clustering to group cosmic-ray pixels into discrete events. Features are extracted from these events and used for classification to identify cosmic-ray muons. Finally, the selection results are summarized and analyzed.

Keywords: CCD; cosmic-ray muons; Laplacian edge detection algorithm; agglomerative hierarchical clustering; feature extraction

Cosmic rays are high-energy particles originating from outer space, first discovered in 1912 by Austrian physicist Victor Hess through measurements of ionization chamber currents. Cosmic rays consist of 98% protons and helium nuclei, 1% heavy nuclei, 1% electrons, and trace amounts of other particles [1]. When primary cosmic rays enter Earth' s atmosphere, high-energy particles such as protons and helium nuclei interact with atmospheric atomic nuclei, generating extensive air showers that produce numerous secondary particles. During CCD observations of target sources, these particles leave distinct tracks on the detector, including electrons, alpha particles, X-rays, and muons. Near the bottom of the atmosphere, cosmic rays are dominated by secondary muons produced in these showers. Among cosmic rays in CCD images, muons represent the principal component. Charged particles such as muons and electrons leave characteristic tracks on CCDs. Electrons primarily originate from environmental radioisotopes rather than cosmic rays (discussed in detail in reference [2]). Alpha particles and X-rays appear as point-like features in CCD images, making them difficult to distinguish from electronic noise inherent to the CCD chip itself. Therefore, studying cosmic rays in CCD images requires selective identification of muons.

Numerous facilities currently exist for cosmic-ray detection, including the Tibet Yangbajing International Cosmic Ray Observatory, GRAPES-3 in India, the Pierre Auger Observatory in Argentina, and LHAASO (Large High Altitude Air Shower Observatory) in China. These installations employ arrays of detectors and telescopes to observe cosmic-ray particles. This study utilizes CCD images from LAMOST (The Large Sky Area Multi-Object Fiber Spectroscopic Telescope), currently China' s largest-aperture telescope. Reference [3] provides a systematic introduction to LAMOST' s structure, principles, and applications. LAMOST is equipped with 16 spectrographs and 32 4k×4k CCD cameras. Each CCD chip measures 49.2 mm × 49.2 mm with pixel dimensions of 12 m × 12 m, operating at approximately -100°C. Since the pilot survey began in 2011, over one million raw CCD images have been accumulated, with each image containing numerous cosmic-ray events suitable for studying cosmic-ray properties and distributions. The number of cosmic rays in a CCD image correlates with exposure time; when normalized to per-minute rates, CCD images exhibit between 5 and 60 cosmic-ray events per minute.

CCD operation relies on the photoelectric effect to convert photons into electrons, which are collected by capacitors, transformed into voltage, and finally output as digital signals. Muons carry either positive or negative charge with magnitude 1 and have a mass of approximately 106 MeV, about 207 times heavier than electrons. When traversing a CCD, muons ionize silicon atoms along their path. Due to their greater mass, they experience less Coulomb scattering and bremsstrahlung than electrons, leaving identifiable straight tracks on the detector.

1. Extraction of Cosmic-Ray Candidate Pixel Lists Using Laplacian Edge Detection

In LAMOST data, single-exposure CCD images contain hundreds of fiber spectra with strong flux values. These images include not only the observed target information but also various noise sources (primarily cosmic rays and bad pixels). To identify muons in CCD images, the first step involves extracting candidate cosmic-ray pixel lists.

Several methods exist for detecting cosmic rays in CCD images, including classical median filtering [4], Laplacian edge detection [5], histogram-based fast algorithms [6], and universal noise elimination algorithms [7]. This work employs an improved Laplacian edge detection method [8] to extract cosmic-ray candidate pixel lists from CCD images. Compared to other approaches, this method effectively detects cosmic rays in single-exposure multi-fiber spectral images even when background flux values are strong. The improvement builds upon the algorithm described in reference [5], enabling identification of cosmic rays that fall within fiber spectral profiles.

The improved Laplacian edge detection algorithm proceeds as follows: (1) Preprocess the original image and convolve it with a 3×3 Laplacian template to obtain image I. Construct a noise model N for cosmic-ray rejection based on Poisson noise and readout noise. Generate an initial cosmic-ray candidate pixel list based on the ratio relationship between I and N. (2) Remove the candidate pixels identified in step (1) from the original image to create a new image. Process each fiber spectrum individually by grouping pixels within 8 pixels of the fiber trace center into a module. Fit the image profiles of these modules in both spatial and dispersion directions (excluding cosmic rays) to obtain fitted image I1. (3) Construct a noise model N1 for cosmic-ray rejection from the fitted image I1 based on Poisson noise and readout noise. Finally, compare residuals between the original and fitted images with noise model N1 to generate the final cosmic-ray candidate pixel list.

Figure 1 [Figure 1: see original paper] illustrates the results of cosmic-ray pixel extraction using the Laplacian edge detection method. Figure 1(a) shows a portion extracted from a CCD image, while Figure 1(b) displays the cosmic-ray pixels extracted from Figure 1(a) using the Laplacian edge detection method, with a black background corresponding to zero flux values. The figure demonstrates that this algorithm can effectively extract cosmic-ray pixels from CCD images.

2. Processing of Bad Pixels and Noise Points

The cosmic-ray candidate pixel list extracted using the Laplacian edge detection method includes actual cosmic rays, bad pixels on the CCD, and noise artifacts generated by the algorithm itself. Bad pixels are inherent defects in the CCD that appear consistently at the same positions across images from the same batch. These can be identified and removed using multiple exposure images of the same target. When fiber trace centers shift in the image, the Laplacian edge detection method produces significant errors during profile fitting, resulting in fitting biases that manifest as algorithmic noise artifacts. These artifacts appear as several columns of dotted lines in the image, with pixel counts far exceeding those of actual cosmic rays in any given column (ranging from 150–500 pixels for artifacts versus 8–50 pixels for cosmic rays). Leveraging this characteristic, anomalous columns can be identified in CCD images to flag algorithmic noise points, which are then removed.

3. Clustering Cosmic-Ray Events Using Agglomerative Hierarchical Clustering

After extracting cosmic-ray pixels from CCD images, an agglomerative hierarchical clustering algorithm groups these pixels into individual cosmic-ray events. The algorithm employed in this work represents an improvement upon the classical agglomerative hierarchical clustering method.

Hierarchical clustering attempts to partition datasets at different levels, forming tree-like cluster structures. Dataset partitioning can employ either a "bottom-up" agglomerative strategy or a "top-down" divisive strategy [9]. Agglomerative hierarchical clustering (AGENS) adopts the "bottom-up" approach, treating each sample in the dataset as an individual cluster initially, then calculating distances between clusters, merging the two nearest clusters, and repeating this process until reaching a predetermined number of clusters.

In classical agglomerative hierarchical clustering, the key aspects are the method for calculating inter-cluster distances and determining the number of clusters. Each cluster is a set of samples, and inter-cluster distances can be computed using various distance metrics such as minimum distance, maximum distance, or average distance. To cluster cosmic-ray pixels into independent events, minimum distance is employed. The minimum distance calculation formula is shown in equation (1), where Ci and Cj represent given clusters, {ma}a=1,2,⋯,k denotes the sample point set of Ci, {mb}b=1,2,⋯,l denotes the sample point set of Cj, and dist is the distance function for calculating distances between ma and mb sample points.

When processing CCD images, the number of cosmic-ray events is unknown beforehand, making it impossible to preset the number of clusters. Instead, an inter-cluster distance threshold dlim is set as the termination condition. The specific algorithm is as follows: (1) Let dataset D consist of pixels corresponding to cosmic rays in the CCD image, treating each pixel as an initial cluster corresponding to C1, C2, ···, Cn. (2) Calculate inter-cluster distances for all clusters using equation (3), which employs Euclidean distance and minimum distance. When both clusters contain only a single sample point, the inter-cluster distance equals the Euclidean distance between the two sample points; otherwise, it equals the minimum distance between the two clusters. If the inter-cluster distance between two clusters is not greater than the threshold dlim, merge these two clusters into a single cluster.

The experimental data are two-dimensional images. In equation (3), Ci and Cj represent any two clusters in the dataset; {ma}a=1,2,···,k denotes the sample point set of Ci; {mb}b=1,2,···,l denotes the sample point set of Cj; xa and ya correspond to the positional information of ma in the image; xb and yb correspond to the positional information of mb in the image. (3) Update the dataset. (4) Repeat steps (2) and (3) until all inter-cluster distances in the dataset exceed the threshold, then terminate.

When numerous cosmic-ray events appear in a CCD image, calculating distances between clusters individually dramatically increases computational load and slows processing speed. Leveraging the characteristic that cosmic-ray events consist of contiguous pixels with counts, the algorithm can be optimized by checking whether other clusters exist within distance dlim from the outermost elements of a cluster. If other clusters exist, they are merged with the current cluster; if not, the process repeats for unprocessed clusters until all inter-cluster distances exceed dlim. When clustering cosmic-ray pixels into events, if the minimum distance between two cosmic-ray events exceeds dlim, they are considered independent events; if the minimum distance does not exceed dlim, they are considered the same event. During cosmic-ray pixel extraction, when cosmic-ray flux values are close to background values, extracted pixels may occasionally be discontinuous. Although rare, to minimize errors from missing pixels and considering the random sparse distribution of cosmic rays in CCD images, setting dlim to 3 effectively clusters cosmic-ray events.

4. Feature Extraction

After clustering cosmic-ray pixels into events, features are extracted based on the morphological characteristics of muons in the images. The primary components of this process are principal component analysis and Pearson correlation coefficient calculation. Muons are identified by obtaining characteristic values through principal component analysis and determining linearity using Pearson correlation coefficient.

4.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA), also known as Karhunen-Loeve expansion, is a classical technique for feature extraction and data representation [10]. Reference [11] provides a comprehensive overview of PCA. PCA can also be used for dimensionality reduction, aiming to find an optimal orthonormal basis that maximizes the variance of projected data. Cosmic-ray events are two-dimensional data. Assuming a cosmic-ray event contains n pixels, its sample set is M = {m1, m2, ···, mn}. All samples are centered as shown in equation (4). The scatter matrix St of the sample set is defined as: where the sample set mean is: Assuming the projection matrix from high-dimensional to low-dimensional space is w, the optimization objective is to maximize the variance of projected sample points. Performing eigenvalue decomposition on St yields eigenvalues , which are sorted in descending order, selecting the largest eigenvalue. The eigenvector corresponding to the largest eigenvalue provides the principal component solution, identifying the optimal orthonormal basis and obtaining characteristic vectors for cosmic-ray events. The projection of cosmic-ray events onto the eigenvector direction serves as the feature. Two eigenvectors correspond to two eigenvalues: one corresponds to the major axis of the cosmic ray (the axis representing the straight track on the CCD), and the other corresponds to the minor axis (the width of the cosmic ray in the CCD image). Muons produce straight tracks in CCD images with indeterminate length but limited width, which is a small quantity. The feature representing width is selected as the muon characteristic and denoted as "projection width."

4.2 Pearson Correlation Coefficient

Correlation coefficients measure the linear relationship between variables. The most commonly used is Pearson correlation coefficient, which describes whether a linear relationship exists between two variables. Reference [12] provides a detailed introduction to Pearson correlation coefficient. For two variables x and y, the Pearson correlation coefficient r is: where x and y correspond to the means of variables x and y. The value of r ranges from -1 to 1; as |r| approaches 1, the linear correlation between the two variables becomes more significant. When $r > 0$, the variables are positively correlated; when $r < 0$, they are negatively correlated; when $r = 0$, they are independent.

Cosmic-ray muons produce straight tracks on CCDs, indicating a linear relationship between their horizontal and vertical coordinates in CCD images. Pearson correlation coefficient is used to characterize this feature. Since muon positions in images may show either positive or negative correlation, the absolute value is taken to indicate the degree of linear correlation. The Pearson correlation coefficient r mentioned hereafter refers to the absolute value.

5. Experimental Results and Analysis

Muons are distinguished from other particles in CCD images based on morphological differences. The most common particles in CCD images are muons and "worms" (multi-scattered electrons, detailed in reference [2]). Alpha particles rarely appear in CCD images. Some cosmic-ray events with unusual shapes cannot be definitively identified and are not discussed here. Figure 2 [Figure 2: see original paper] shows images obtained from LAMOST data, with Figure 2(a) displaying a cosmic-ray muon and Figure 2(b) showing a cosmic-ray "worm." As illustrated in Figure 2(b), "worms" exhibit a certain curvature.

Using the processing steps described in Sections 1 through 3, over 200 CCD images from LAMOST in 2015 were randomly selected, yielding more than 30,000 cosmic-ray events. The positional index information of these events serves as sample set C1. From this set, 1,000 muons and 1,000 "worms" were manually identified to create data sample 1. The "projection width" and Pearson correlation coefficient of these events were statistically analyzed, with results shown in Table 1 . In Table 1, the mean of cosmic-ray event "projection width" is denoted as wp and standard deviation as wp; the mean of Pearson correlation coefficient is r and standard deviation as r.

Table 1 clearly shows the distinction between muons and "worms."For "projection width," muons exhibit smaller characteristic values while "worms" show larger values. For Pearson correlation coefficient, muons produce straight tracks in CCDs, resulting in more significant linear correlation between horizontal and vertical coordinates and values closer to 1. "Worms" undergo multiple scattering in CCDs, producing curved tracks with lower Pearson correlation coefficient values.

From sample set C1, 2,500 muons and 2,500 non-muon cosmic-ray events (including "worms," X-rays, possible alpha particles, and other particles) were randomly selected through manual identification to create data sample 2. The distribution of "projection width" and Pearson correlation coefficient for these 5,000 events is shown in Figure 3 Figure 3: see original paper. The two features effectively separate most muons from other cosmic rays. In the region where Pearson correlation coefficient exceeds 0.8 and "projection width" is less than 1, muons and other cosmic-ray events cannot be clearly distinguished. Examining these events with similar characteristics reveals they contain very few pixels. The relationship between Pearson correlation coefficient and pixel count is shown in Figure 3 Figure 3: see original paper. Cosmic-ray events that are difficult to differentiate in Figure 3(a) can be effectively distinguished by pixel count in Figure 3(b).

For cosmic-ray events with few pixels, relying solely on "projection width" and Pearson correlation coefficient introduces some error in muon selection. Pixel count must also be considered. Figure 4 [Figure 4: see original paper] shows statistical results for pixel counts in data sample 2. The horizontal axis represents the number of pixels in cosmic-ray events, and the vertical axis represents the number of events with corresponding pixel counts. Figure 4(a) shows results for the entire dataset, demonstrating that events with more pixels are less common. Based on Figure 4(a), events with no more than 30 pixels were selected for further analysis, shown in Figure 4(b). Muon events contain more than 8 pixels, while other cosmic-ray events contain more than 5 pixels. Among non-muon cosmic rays, nearly half have fewer than 8 pixels. With pixel counts greater than 8, muons outnumber other cosmic rays. Therefore, a pixel count threshold of 8 is set to classify muons versus other cosmic-ray events.

Based on the different distributions of characteristic values, various parameters are set to select muons. From sample set C1, positional index information for 2,500 muons and 2,500 non-muon cosmic rays in CCD images was randomly selected as test dataset C2. Different characteristic parameters were applied to select muons from cosmic-ray events, with results shown in Table 2 . Table 2 displays selection accuracy, with the first column showing feature parameter settings (mw corresponds to the mean of muon "projection width," mw to its standard deviation, mr to the mean of muon Pearson correlation coefficient, and mr to its standard deviation). The second column shows the number of correctly identified muons, the third column shows cosmic-ray events incorrectly classified as muons, the fourth column shows selection accuracy, and the final column shows selection sensitivity, measuring the algorithm' s ability to identify muons. Assuming M muons and F non-muon cosmic-ray events exist in the complete dataset, with TM correctly identified muons and TF correctly identified non-muon events in the selection results, the accuracy formula is shown in equation (9) and sensitivity formula in equation (10).

Table 2 demonstrates that different thresholds for "projection width," Pearson correlation coefficient, and pixel count yield different accuracies. Appropriate threshold settings can effectively select muons from cosmic-ray events. Parameter selection in Table 2 must incorporate the means and standard deviations of different muon features from Table 1. By statistically analyzing muon "projection width" and Pearson correlation coefficient, specific numerical parameters and their variations are obtained. Combining these parameter fluctuations with morphological differences between muons and other cosmic-ray events in the dataset enables final selection. When appropriate cosmic-ray feature thresholds are set, accuracy reaches approximately 0.98, indicating effective muon selection. The method demonstrates high muon identification capability but may misclassify non-muon cosmic-ray events with similar features as muons.

Although the algorithm introduces some errors during data processing, analyzing muon intrinsic characteristics combined with morphological differences between muons and other cosmic-ray events in the dataset still enables effective selection. The core of this method lies in feature parameter setting. Appropriate parameter settings yield better results, and parameters can be optimized by training on large datasets. Using the optimal selection conditions from Table 2 (feature parameters corresponding to the second row), images from LAMOST' s CCD camera 6 red channel taken in November 2017 were processed. The number of muons in CCD images with 600.0s exposure time was counted throughout the month, with results shown in Figure 5 [Figure 5: see original paper]. The horizontal axis represents the modified Julian minute (mjm) when CCD images were taken, and the vertical axis shows the number of muons in each CCD image. Blue vertical lines correspond to dates in the secondary axis, visually displaying muon fluctuations during the month.

This paper explores a simple and effective method for selecting muons in single-exposure CCD images, achieving efficient data processing by leveraging morphological features of cosmic rays in images. The method uses Laplacian edge detection to extract cosmic-ray pixels, employs clustering algorithms to group pixels into cosmic-ray events, and selects muons through feature extraction and morphological differences between events. Selection accuracy depends on feature parameter settings, which can be improved through continuous statistical analysis and testing to establish better parameters and additional features.

During training and test sample collection, manual identification to differentiate muons from other cosmic-ray events introduces some bias into the dataset, affecting subsequent statistical analysis and selection results.

Acknowledgments: We thank Professor Hu Hongbo from the Institute of High Energy Physics, Chinese Academy of Sciences, for helpful discussions.

References

[1] O.C. Allkofer. *Introduction to Cosmic Radiation* [M]. C.X. Xu & Q.Q. Zhu, Trans. Beijing: Science Press, 1987:34.

[2] Don Groom. Cosmic Rays and Other Nonsense in Astronomical CCD Imagers [J]. *Experimental Astronomy*, 2002, 14(1): 45-55.

[3] Cui, X.Q., Zhao, Y.H., Chu, Y.Q., et al. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) [J]. *RAA*, 2012, 12(9): 1197-1242.

[4] Gonzalez RC, Woods RE. *Digital Image Processing* [M]. Q.Q. Ruan & Y.Z. Ruan, Trans. Beijing: Electronic Industry Press, 2007:185-186.

[5] Pieter G. van Dokkum. Cosmic-Ray Rejection by Laplacian Edge Detection [J]. *PASP*, 2001, 113: 1420-1427.

[6] Pych W. A Fast Algorithm for Cosmic-ray Removal from Single Images [J]. *PASP*, 2004, 116(816): 148-153.

[7] Garnett R, Huegerich T, Chui C, et al. A Universal Noise Removal Algorithm with An Impulse Detector [J]. *IEEE Transaction On Image Processing.* 2005, 14(11): 1747-1754.

[8] Zhongrui Bai, Haotong Zhang, et al. Cosmic Ray Removal in Fiber Spectroscopic Image [J]. *PASP*, 2017, 129(972): 024004(11pp).

[9] Zhou Zhihua. *Machine Learning* [M]. Beijing: Tsinghua University Press, 2016: 214.

[10] C. Li, Y. Diao, H. Ma and Y. Li, A Statistical PCA Method for Face Recognition [J]. *Intelligent Information Technology Application*, 2008, pp. 376-380.

[11] Sasan Karamizadeh, Shahidan M. Abdullah. An Overview of Principal Component Analysis [J]. *Journal of Signal and Information Processing*, 2013, 4: 173-175.

[12] Leo Egghe, Loet Leydesdorff. The relation between Pearson' s correlation coefficient r and Salton' s cosine measure [J]. *Journal of the American Society for Information Science & Technology*, 2009, 5: 1027-1036.

Note: Figure translations are in progress. See original paper for figures.