Advances in Astronomy

Astronomical Massive Data Processing Technology

Lead Guest Editor: Feng Wang
Guest Editors: Wenda Cao, Denis Bastieri, Junhui Fan, and Chenzhou Cui

Astronomical Massive Data Processing Technology

Advances in Astronomy

Astronomical Massive Data Processing Technology

Lead Guest Editor: Feng Wang
Guest Editors: Wenda Cao, Denis Bastieri, Junhui Fan, and Chenzhou Cui

Contents

Hindawi
Advances in Astronomy
Volume 2020, Article ID 6848935, 2 pages
https://doi.org/10.1155/2020/6848935

Editorial

Astronomical Massive Data Processing Technology

Feng Wang,[1] **Wenda Cao**,[2] **Denis Bastieri**,[3] **Junhui Fan**,[1] **and Chenzhou Cui**[4]

[1]*Center for Astrophysics, Guangzhou University, Guangzhou, China*
[2]*Big Bear Solar Observatory, Big Bear, USA*
[3]*Padova University, Padova, Italy*
[4]*National Astronomical Observatory, Chinese Academy of Sciences, Beijing, China*

Correspondence should be addressed to Feng Wang; fengwang@gzhu.edu.cn

Received 1 April 2020; Accepted 1 April 2020; Published 1 May 2020

Astronomy has become one of the biggest consumers of computing resources in the past 10 years. New computational techniques are emerging and dedicated to the various fields of astronomy research. Thus, we can already see such a development trend that astronomical studies will become larger in all dimensions: larger data collections, more accurate data analysis and processing, and more detailed results. The significant increase of these large-size astronomical studies performed by using instruments built around the world requires not only huge processing power but also clever system design, all the way from hardware construction to software development and deployment. With this special issue, we want to cover most relevant aspects related to new computing technology to illustrate the readers the wide scope opened by these observational techniques in computing astronomy.

There are total 6 papers collected in this special issue, which concerned massive astronomical data archive and management, data analysis, visualization and application, and data reduction in modern telescope. In what follows, we present the main contents of this special issue split by subject:

Two papers of this issue are devoted to massive data archive and management techniques by H. Zhang et al. and X. Ye et al. H. Zhang et al. presented a robust way to transfer massive data. More than 84,000 pulsar observation data files in the PSRFITS format have been released with a maximum file size of 1 GB. X. Ye et al. designed effective mechanisms to ensure the high stability of the system and to control the data transmission when the network is unstable through adaptive algorithms which could transfer the massive data from observatory to its data processing center.

Four papers discussed the massive data analysis, which showed how to apply information processing techniques in current astronomical data processing. S. Shrestha and D.-Y. Choi proposed a technique that provides sufficiently accurate estimation for Ku-band signal attenuation in site specifically whose effectiveness is performed through the statistical analysis against the established rain attenuation models.

In the study of W. Wang and Y. Yan, a new restoration method for solar radio image was proposed. Two images with different weighting functions from the same data are combined. This restoration method has been applied to data processing of Mingantu spectral radioheliograph. M. Zhu et al. extracted filaments based on morphology component analysis from radio astronomical images. The experimental results showed the method proposed by the authors is effective and final images have higher peak signal-to-noise ratio (PSNR). Z. Qiang et al. applied an adaptive background learning method to detect the CMEs (coronal mass ejections). Comparing the result with CDAW, CACTus, SEEDS, and CORIMP catalogs, the new method detected about 73% of the CMEs listed in the CDAW CME catalog, which is identified by human visual inspection.

Conflicts of Interest

All authors declare that there are no conflicts of interest regarding the publication of this article.

Feng Wang
Wenda Cao
Denis Bastieri
Junhui Fan
Chenzhou Cui

Hindawi
Advances in Astronomy
Volume 2019, Article ID 6582104, 14 pages
https://doi.org/10.1155/2019/6582104

Research Article

A CME Automatic Detection Method Based on Adaptive Background Learning Technology

Zhenping Qiang,[1,2] **Xianyong Bai**,[2] **Qinghui Zhang**,[1] **and Hong Lin**[1]

[1]*College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China*
[2]*CAS Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*

Correspondence should be addressed to Zhenping Qiang; qzplucky@163.com

Received 14 March 2019; Revised 22 August 2019; Accepted 8 October 2019; Published 7 November 2019

Guest Editor: Junhui Fan

In this paper, we describe a technique, which uses an adaptive background learning method to detect the CME (coronal mass ejections) automatically from SOHO/LASCO C2 image sequences. The method consists of several modules: adaptive background module, candidate CME area detection module, and CME detection module. The core of the method is based on adaptive background learning, where CMEs are assumed to be a foreground moving object outward as observed in running-difference time series. Using the static and dynamic features to model the corona observation scene can more accurately describe the complex background. Moreover, the method can detect the subtle changes in the corona sequences while filtering their noise effectively. We applied this method to a month of continuous corona images, compared the result with CDAW, CACTus, SEEDS, and CORIMP catalogs and found a good detection rate in the automatic methods. It detected about 73% of the CMEs listed in the CDAW CME catalog, which is identified by human visual inspection. Currently, the derived parameters are position angle, angular width, linear velocity, minimum velocity, and maximum velocity of CMES. Other parameters could also easily be added if needed.

1. Introduction

A coronal mass ejection (CME) is a significant release of plasma and accompanying magnetic field from the solar corona. It often follows solar flares and is normally present during a solar prominence eruption. The plasma is released into the solar wind and can be observed in coronagraph imagery [1–3]. CME is the most energetic and important solar activity and is a significant driver of space weather in the near-Earth environment and throughout the heliosphere. When the ejection is directed towards and reaches the Earth as an interplanetary CME (ICME), ICME can cause geomagnetic storms that may disrupt Earth's magnetosphere, damage satellites potentially, induce ground currents, and increase the radiation risk for astronauts [4]. Thus, CME detection is an active area of research.

CME was first observed coincided with the first-observed solar flare on 1 September 1859, and it has been studied extensively since it was first reported [5] more than four decades ago. Along with the continuous progress of space observations of the corona, a series of satellites with coronal imaging observation ability such as OSO-7, P78-1, Skylab, SMM, and SOHO were launched; especially over the past 24 years, coronal mass ejections have been detected routinely by visual inspection of each image from the Large Angle Spectrometric Coronagraph (LASCO) onboard SOHO [6]. To further understand CME, especially its three-dimensional properties which can be observed by the Sun-Earth Connection Coronal and Heliospheric Investigation (SECCHI) [7], SECCHI flew aboard NASA's recently launched Solar Terrestrial Relations Observatory (STEREO).

In contrast to this huge amount of observations of CMEs, the identification and cataloging of CMEs are important tasks that provide the basic knowledge for further scientific studies. There are two main categories of methods used to detect the CMEs. One category is the manual detection method, with the LASCO instrument coronagraphs. Currently, there exists a manual catalog which is the

Coordinated Data Analysis Workshop Data Center (CDAW) catalog [8] to catalog observed CMEs. This catalog is compiled by observers who look through the sequences of LASCO coronagraph images. But this human-based process is tedious and subjected to observers' biases. To promote the detection of CMEs, another category is the automatic detection method, which detects and characterizes CMEs in coronagraph images.

The Computer Aided CME Tracking software package is the first automatic detection method introduced in 2004 [9]. It utilizes the Hough transform to identify CMEs. In 2005, Boursier et al. proposed a method named the Automatic Recognition of Transient Events and Marseille Inventory from Synoptic maps (ARTEMIS) [10]; it utilizes LASCO C2 synoptic maps and based on an adaptive filtering and segmentation to detect CMEs. In [11], Olmedo et al. presented the Solar Eruptive Event Detection System (SEEDS) which used image segmentation techniques to detect CMEs. In [12], Young and Gallagher described and demonstrated a multiscale edge detection technique that addresses the CME detection and tracking, which could serve as one part of an automated CME detection system. In 2009, Goussies et al. developed an algorithm based on level set and region competition methods to characterize the CME texture, and by using the texture information in the region competition motion equations to evolve the curve, to this end, segmentation of the leading edge of CMEs is performed on individual frames [13]. In the same year, Byrne et al. adopted a multiscale decomposition technology to extract structure of the processing image and used an ellipse parameterization of the front to extract the kinematics (height, velocity, and acceleration) and morphology (width and orientation) change to detect the CMEs [14]. In [15], Gallagher et al. developed an image processing technique to define the evolution of CMEs by texture and used a supervised segmentation algorithm to isolate a particular region of interest based upon its similarity with a prespecified model to automatically track the CMEs. In 2012, Zhao-Xian et al. [16] presented a method to detect CMEs by analyzing the sudden change of frequency spectrum in the coronagraph. In 2014, Bemporad et al. [17] described the onboard CME detection algorithm for the Solar Orbiter-METIS coronagraph. The algorithm is based on the running differences between consecutive images to get significant changes and to provide the CME first detection time. In 2017, Zhang et al. [18] proposed a suspected CME region detection algorithm by using the extreme learning machine (ELM) method which takes into account the features of the grayscale and the texture. In 2018, based on the intensity thresholding followed by the area thresholding in successive difference images spatially rebinned to improve signal-to-noise ratio, Patel et al. [19] proposed a CME detection algorithm for the Visible Emission Line Coronagraph on ADITYA-L1. Recently, machine learning has been used in solar physics. Dhuri et al. [20] used machine learning to classify vector magnetic field observations from flaring ARs. Huang et al. [21] applied a deep learning method to flare forecasting. Very recently, Wang et al. [22] even proposed an automatic tool for CME detection and tracking with machine learning techniques.

These CME automatic detection methods that mentioned above are mainly based on three kinds of strategies: (i) enhance the coronagraph images and describe the kinematics and morphology features (edge, luminance, shape, etc) of the processing images and then use these features to determine the occurrence of CME; (ii) establish the CME evolution models according to the historical CMEs' dynamic evolution characteristics, use the same feature extraction method to extract dynamic evolution characteristics of the processing sequences, and then compare the extracted characteristics to the model to determine the occurrence of CME; and (iii) apply supervised classification problem in machine learning to detect CME.

The coronagraph data can be considered as a three/four-dimensional dataset with two/three spatial and one temporal dimension. The key to automatic detection methods is how to distinguish CME regions from other parts of the image. These methods do not utilize time dimension information adequately. To make full use of time-domain information, we can use video processing technology for CME detection. In fact, we can consider a coronagraph image sequence as a video and regard CMEs as abnormal events in the video. The detection process of CME can use the video surveillance technology, which includes change detection, background model, foreground detection, and object tracking. Further, considering the coronagraph image sequence itself is a dynamic scene, and the CME is also a dynamic process, so the CME detection methods must adapt to the scene change. Inspired by these ideas, in this paper, we attempt to detect CMEs based on adaptive background learning technology. The method consists of three main modules described below:

(1) Adaptive background module: this module is mainly implemented to maintain the background model of the coronagraph image sequence

(2) Candidate CME area detection module: this module is used to detect the foreground areas of the coronagraph images

(3) CME detection module: this module is based on the candidate areas to identify the CME event

The remaining of the paper is organized as follows. In Section 2, we first give a specification about the adaptive background module. Then in Section 3, we will formulate the background and foreground classification problem and propose a method of candidate CME area detection. Section 4 describes an algorithm for CME detection based on candidate CME area detection module. The experimental results and validation on LASCO C2 data are presented in Section 5. The paper is concluded in Section 6.

2. Adaptive Background Module

In coronagraph image sequence, the background environments always change; for example, small moving objects such as stars and cosmic rays can make the background change. So, the background representation model must be more robust and adaptive, and the background module must be continuously updated to represent the change of the

scene. To solve the strong chaotic interference in the background, several methods have been proposed to adapt to variety of background situations. Among them, mixture of Gaussians (MoG) [23] is considered as a promising method. In the video monitoring, because of the high frame rate, the MoG can achieve good results in the gradual change scene, but for CME detection, the interference changes significantly, so it needs better method to model the dynamic scene. Li et al. proposed a statistical modeling [24] that used the co-occurrence of color characteristics of two consecutive frames to model the dynamic scene. By using this statistical modeling, this method can represent nonstatic background objects, so it has good robustness for the existence of dynamic background periodic interference. The statistical modeling is very suitable for CME detection which is often associated with other forms of solar activities. We apply this method to model the background, namely, employing the color feature to describe the static background and co-occurrence color features to describe the moving background, and then use a Bayes decision rule for classification of background and foreground.

2.1. Formulation of the Classification Rule Based on Bayes. In the method of automated detection of CMEs based on an adaptive background module, each pixel in the coronagraph image is divided into two categories: background pixels and foreground pixels (candidate CME area pixels). Therefore, using the Bayes rule, the feature vector distribution probability of each pixel satisfies the following equation:

$$P_s(v) = P_s(v \mid b)P_s(b) + P_s(v \mid f)P_s(f), \tag{1}$$

where $s = (x, y)$ indicates pixel position, v is the statistical feature vector, $P_s(v \mid b)$ is the probability of the feature vector v being observed as a background at s, $P_s(b)$ is the prior probability of the pixel s belonging to the background, and $P_s(v)$ is the prior probability of the feature vector v being observed at the position s. Similarly, f denotes the foreground (or candidate CME area). By using the Bayes decision rule, the pixel can be classified as background if the feature vector satisfies the following equation:

$$P_s(b \mid v) > P_s(f \mid v), \tag{2}$$

and by using Bayesian conditional posterior probability,

$$P_s(C \mid v) = \frac{P_s(v \mid C)P_s(C)}{P_s(v)}, \quad C = b \text{ or } f, \tag{3}$$

and substituting (1) and (3) into (2), it becomes

$$2P_s(v \mid b)P_s(b) > P_s(v), \tag{4}$$

that is, if we obtained the prior probability $P_s(b)$, $P_s(v)$, and conditional probability $P_s(v \mid b)$ at the moment t, the pixel s with the feature vector v can be classified as background or foreground based on formula (4).

2.2. Description of the Feature Vector. In formula (4), the probability functions $P_s(v)$ and $P_s(v \mid b)$ are all associated with the feature vector v. For the coronagraph images, the most prominent feature is the luminance characteristics and takes into account the dynamic disturbance; we must increase the feature vector to characterize the dynamic properties. In this paper, we adopt the luminance features and co-occurrence luminance features to model the background.

The coronal image's luminance level is high, if calculating and recording all the luminance feature vectors' probability is unrealistic. Fortunately, at the same location of the coronagraph image, the luminance change is not very big. So for each pixel, it will be enough to record a small subspace feature vectors as the background model. An example of the principal feature representation with luminance and co-occurrence luminance in LASCO C2 pseudocolor coronagraph images in the year of 2014 is shown in Figure 1. The left image (a) shows the position of the selected pixel, and the right image (b) and image (c) are the histograms of the statistics for the most significant color and co-occurrence color. The histogram of the color features shows that only the first thirty color distributions account for 68.38% of all color feature space, and the first thirty co-occurrence color distributions account for 79.51% of all co-occurrence color feature spaces.

Therefore, as shown in Figure 1, we can represent $P_s(v)$ and $P_s(v \mid b)$ well by selecting a small number of feature vectors. In the experiments of this paper, the color feature vector is quantized for 128 levels and recorded the first 25 feature vectors and the co-occurrence color feature vector is quantized for 64 levels and recorded the first 40 feature vectors.

2.3. Background Model and Parameters. In this paper, we focus on the effective detection method of CME, and the data object processed is pseudocolor coronagraph images. So we use statistical features in pseudocolor coronagraph images to model the background, in particular including the prior probability of feature vectors belonging to the background, color, and co-occurrence color feature vectors statistics list information. Suppose at the time t, at pixel point s, the color is $c_t = [\, r_t \;\; g_t \;\; b_t \,]^T$, the previous frame's luminance is $c_{t-1} = [\, r_{t-1} \;\; g_{t-1} \;\; b_{t-1} \,]^T$, and the co-occurrence color feature vector can be defined as $cc_t = [\, r_{t-1} \;\; g_{t-1} \;\; b_{t-1} \;\; r_t \;\; g_t \;\; b_t \,]^T$. For each pixel, the background model includes the following:

(1) The prior probabilities $p_{b,c}^{s,t}$ and $p_{b,cc}^{s,t}$, $p_{b,c}^{s,t}$ indicate that color feature vector belongs to the background of the time t at pixel point s, and $p_{b,cc}^{s,t}$ indicates that the co-occurrence color feature vector belongs to the background of the time t at pixel point s

(2) Color feature vector statistics list of the time t at pixel point s, $S_v^{c,s,t,i}$, $i = 1, \ldots, \mathrm{Nc}$:

$$S_v^{c,s,t,i} = \begin{cases} p_{v_c}^{s,t,i} = P\left(v_c^{t,i} \mid s\right), \\ p_{v_c,b}^{s,t,i} = P\left(v_c^{t,i} \mid b, s\right), \\ v_c^{t,i} = \left[\, r_t^i \;\; g_t^i \;\; b_t^i \,\right]^T, \end{cases} \tag{5}$$

where Nc is the recording number of statistical color feature vectors, $p_{v_c}^{s,t,i}$ is the statistical

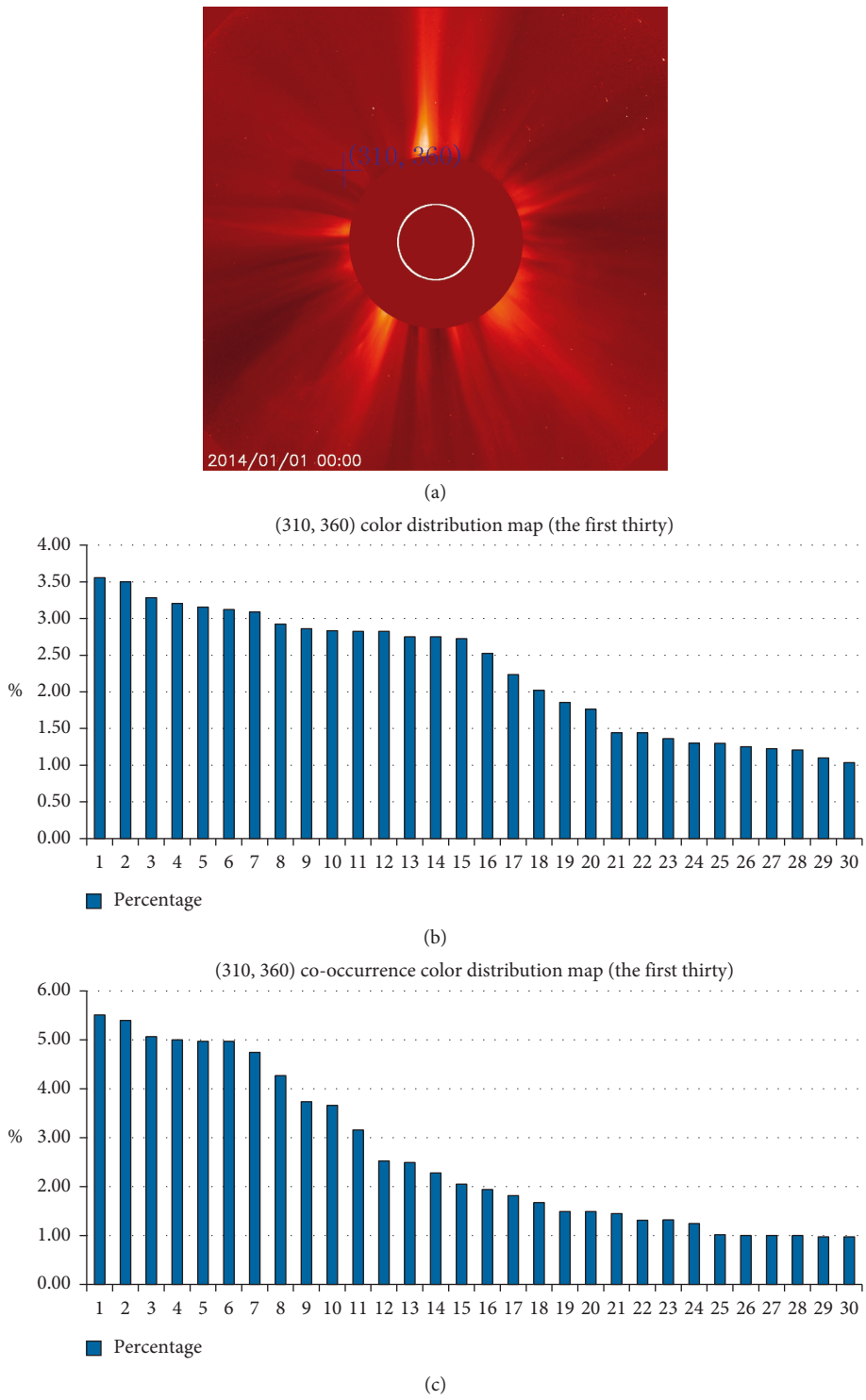


Figure 1: One example of learned principal features of LASCO C2 pseudocolor coronagraph images in the year of 2014. (a) The position of the selected pixel, (b) significant color histogram, and (c) significant co-occurrence color histogram.

probability of ith color feature vector v_c at position s until time t, and $p_{v_c,b}^{s,t,i}$ is the probability of ith color feature vector v_c at position s which was judged as the background

(3) Co-occurrence color feature vector statistics list of the time t at pixel point s, $S_v^{cc,s,t,i}$, $i = 1, \ldots, \mathrm{Ncc}$:

$$S_v^{cc,s,t,i} = \begin{cases} p_{v_{cc}}^{s,t,i} = P\left(v_{cc}^{t,i} \mid s\right), \\ p_{v_{cc},b}^{s,t,i} = P\left(v_{cc}^{t,i} \mid b, s\right), \\ v_{cc}^{t,i} = \left[\, r_{t-1}^i \;\; g_{t-1}^i \;\; b_{t-1}^i \;\; r_t^i \;\; g_t^i \;\; b_t^i \,\right]^T, \end{cases} \tag{6}$$

where Ncc is the recording number of statistical co-occurrence color feature vectors, $p_{v_{cc}}^{s,t,i}$ is the statistical probability of ith co-occurrence color feature vector v_{cc} at position s until time t, and $p_{v_{cc},b}^{s,t,i}$ is probability of ith co-occurrence color feature vector v_{cc} at position s which was judged as the background

According to the distribution of the feature vectors, as shown in Figure 1, the first N_1 elements of the list are enough to cover the majority part of the feature vectors from the background. Therefore, in the case $p_{v_c}^{s,t,i} \approx p_{v_c,b}^{s,t,i}$ or $p_{v_{cc}}^{s,t,i} \approx p_{v_{cc},b}^{s,t,i}$, $p_{v_c}^{s,t,i}$ and $p_{v_{cc}}^{s,t,i}$can be used to represent the background. Otherwise, when $p_{v_c}^{s,t,i} \gg p_{v_c,b}^{s,t,i}$ or $p_{v_{cc}}^{s,t,i} \gg p_{v_{cc},b}^{s,t,i}$, it indicates that this feature vector is corresponding to the foreground. This is the foundation that we used to detect the CMEs.

3. Candidate CME Area Detection Module

The candidate CME area detection module is based on the established background model we discussed in 2.3 and the formulation of background and foreground classification we discussed in 2.1. The candidate CME area detection module consists of three parts: change detection, change classification, and candidate CME area segmentation. In the first step, nonchange pixels are filtered out by using the background difference and frame difference, which will improve the computing speed; in the meantime, the detected change pixels are separated as pixels belonging to stationary and moving scene according to interframe changes. In the second step, based on the learned statistics information of color feature vectors and co-occurrence color feature vectors, the pixels associated with stationary or moving scene are further classified as background or candidate CME area by using the Bayes decision rule. In the third step, candidate CME areas are segmented by the morphological processing based on the classification results. The process is the basis of the algorithm [25], and the block diagram of the candidate CME area detection is shown in Figure 2.

3.1. Change Classification. Candidate CME area detection is based on two-class background features (color and co-occurrence color); first of all, each coronagraph image's change must be classified into two types. As shown in Figure 2, the change classification gets through temporal differences and background differences. The temporal difference binary image is denoted by $\mathrm{Ftd}(s,t)$, and the background difference binary image is denoted by $\mathrm{Fbd}(s,t)$. If $\mathrm{Fbd}(s,t) = 1$ (no matter the result what $\mathrm{Fbd}(s,t)$ is) is detected, the pixel s is classified as a change pixel. If $\mathrm{Fbd}(s,t) = 1$ and $\mathrm{Fbd}(s,t) = 0$ are detected, the pixel s is classified as a stationary pixel. They are further classified as background or candidate CME area separately, the change pixel will be classified by co-occurrence color features, and the stationary pixel will be classified by color features.

3.2. Pixel's Classification. For the current processing coronagraph image's each pixel point s, at first to extract the feature vector v^t (color feather vector and co-occurrence color vector which are based on pixel change classification) and match with each vector in the pixel's feature vector statistics list by using formula (7), the sum of all matched ($M = 1$) prior probability and condition probability in the statistical list was further calculated to obtain the prior probability $P_s(v^t)$ and conditional probability $P_s(v^t \mid b)$ of the pixel's vector v^t. Meanwhile, the prior probability $P_s(b)$ is obtained, which was maintained in the background model. Finally, by substituting $P_s(v^t)$, $P_s(v^t \mid b)$, and $P_s(b)$ into formula (4), the pixel point s can be classified as background or candidate CME area:

$$M(v_1, v_2) = \begin{cases} 1, & \forall i \left| v_1(i) - v_2(i) \right| \le \delta, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $\delta = 2$ is chosen so that if the similar features are quantized into neighboring vectors, the statistics can still be retrieved. If no element in the pixel's feature vector statistics list is matched, $P_s(v^t)$ and $P_s(v^t \mid b)$ are set to 0.

3.3. Candidate CME Area Segmentation. It is obvious that, after the pixel's classification, only a small percentage of the background pixels are wrongly classified as candidate CME ones. There are many isolated points, so the morphological operation (a pair of open and close) is applied to remove the scattered error points and connect the candidate CME area points. Finally, the candidate CME area detection module will output a binary image $O(s,t)$.

3.4. Adaptive Background Learning. The coronagraph image sequence is a gradually changing scene, so the background model must be maintained to adapt to the various changes over time. In practice, the background model's probability information and a reference background image must be updating.

3.4.1. Updating to Background Model's Probability Information. Based on the previous obtained binary image $O(s,t)$, the pixel s with the feature vector v^t is classified as candidate CME area or background. The prior probability and the conditional probability associated with the color feature are gradually updated by formula (8), and the updating of the prior probability and the conditional probability associated with the co-occurrence color feature is similar:

$$\begin{cases} p_{b,c}^{s,t+1} = (1-\alpha_1) p_{b,c}^{s,t} + \alpha_1 M_{b,c}^{s,t}, \\ p_{v_c}^{s,t+1,i} = (1-\alpha_1) p_{v_c}^{s,t,i} + \alpha_1 M_{v_c}^{s,t}, \\ p_{v_c,b}^{s,t+1,i} = (1-\alpha_1) p_{v_c,b}^{s,t,i} + \alpha_1 \left(M_{b,c}^{s,t} \wedge M_{v_c}^{s,t} \right), \end{cases} \tag{8}$$

for $i = 1, \ldots, \mathrm{Nc}$, where α_1 is a learning rate which controls the speed of feature learning; in the experiment, we set $\alpha_1 = 0.005$; $M_{b,c}^{s,t} = 1$ when s is labeled as the background at time t from $O(s,t)$; otherwise, $M_{b,c}^{s,t} = 0$. $M_{v_c}^{s,t} = 1$ when $v_c^{t,i}$ in the color feature vector statistics list $S_v^{c,s,t,i}$ in formula (5)

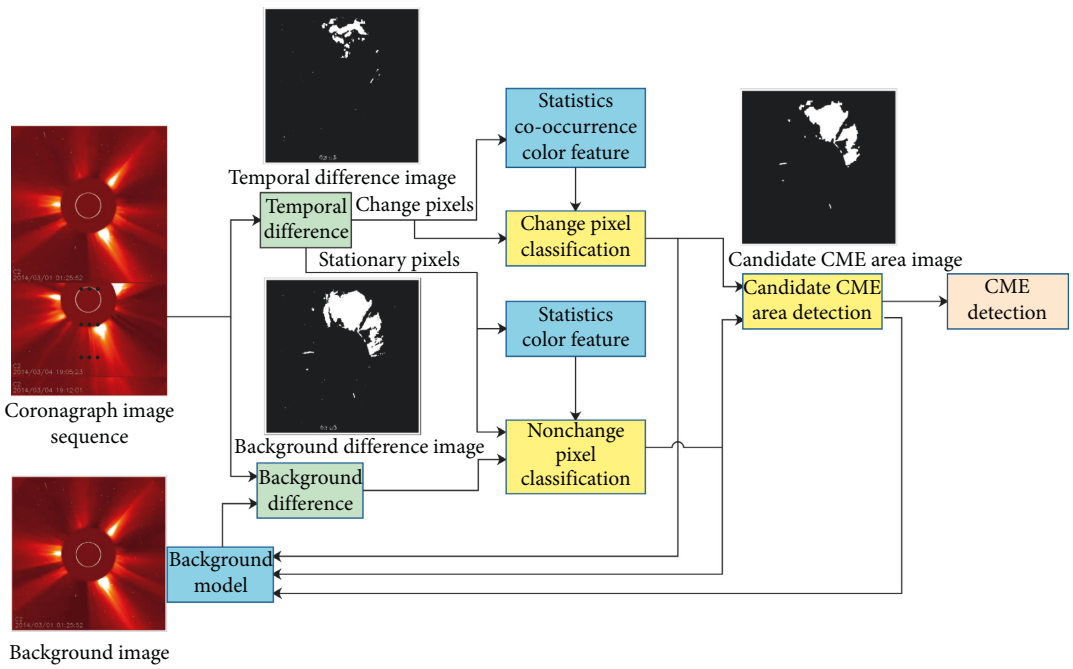


Figure 2: The block diagram of the candidate CME area detection.

matches v_c^t best and $M_{v_c}^{s,t} = 0$ for others. In more detail, the above updating can be stated as follows:

(a) If the pixel s is labeled as a background point at time t by color feature, $p_{b,c}^{s,t+1}$ is slightly increased from $p_{b,c}^{s,t}$ due to $M_{b,c}^{s,t} = 1$. Meanwhile, the probability of the matched feature is also increased due to $M_{v_c}^{s,t} = 1$. If $M_{v_c}^{s,t} = 0$, then the statistics for the unmatched features are gradually decreased. If there is no matched feature between $v_c^{t,i}$ and the elements of the feature vector recording list $S_v^{c,s,t,i}$, the Ncth element in the list is replaced by a new feature vector by formula (9). If the number of the elements is smaller than Nc, a new feature vector by formula (9) is added:

$$\begin{aligned} p_{v_c}^{s,t+1,Nc} &= \alpha_1, \\ p_{v_c,b}^{t+1,Nc} &= \alpha_1, \\ v_c^{t,Nc} &= v_c^t. \end{aligned} \tag{9}$$

(b) If the pixel s is labeled as a foreground point at time t by color feature, $p_{b,c}^{s,t+1}$ and $p_{v_c,b}^{s,t+1,i}$ are slightly decreased due to $M_{b,c}^{s,t} = 0$. However, the probability of the matched feature is increased.

To ensure that the element that is replaced is the lowest probability one, updated elements in the feature vector statistics list $S_v^{c,s,t,i}$ are resorted to a descending order according to $p_{v_c}^{s,t+1,i}$.

3.4.2. Updating to the Reference Background Image. In the candidate CME area detection process, it need to use background difference to classify the change, so a reference background image that represents the most recent appearance of the sense must be maintained at each time step. An infinite impulse response (IIR) is used to update the gradual changes for stationary background sense. If the pixel s is classified as a change point in the change classification step and the candidate CME area segmentation result $O(s,t) = 1$, the reference background image is updated as

$$B_c(s,t+1) = (1-\alpha_2)B_c(s,t) + \alpha_2 I_c(s,t), \tag{10}$$

where α_2 is a parameter of the IIR filter and $c \in \{r,g,b\}$ is the color information of the process point. A small positive number of α_2 is selected to smooth out the disturbances caused by image noise, and in the experiment, we set $\alpha_2 = 0.1$.

If $\mathrm{Fbd}(s,t) = 1$ and $\mathrm{Ftd}(s,t) = 1$, but $O(s,t) = 0$. This means that there is a significant change, but in the end, it was not classified as candidate CME area; it indicates a background change is detected. So the processed pixel s's color information should replace the reference background, that is, $B_c(s,t+1) = I_c(s,t)$.

Through this operation, the reference background image can be a good representation of coronal scene change.

4. CME Detection Module

Based on the candidate CME areas, we can detect the CME according to the morphological and dynamic characteristics of the candidate CME area. For example, to identify a newly emerging CME, it must be seen to move outward in at least two running-difference images. This condition is set by Robbrecht and Berghmans [9] and Olmedo [26] to define a newly emerging CME.

The CME detection method we proposed is based on a continuous frame processing approach, so after the detection of the candidate CME area, we set two conditions as

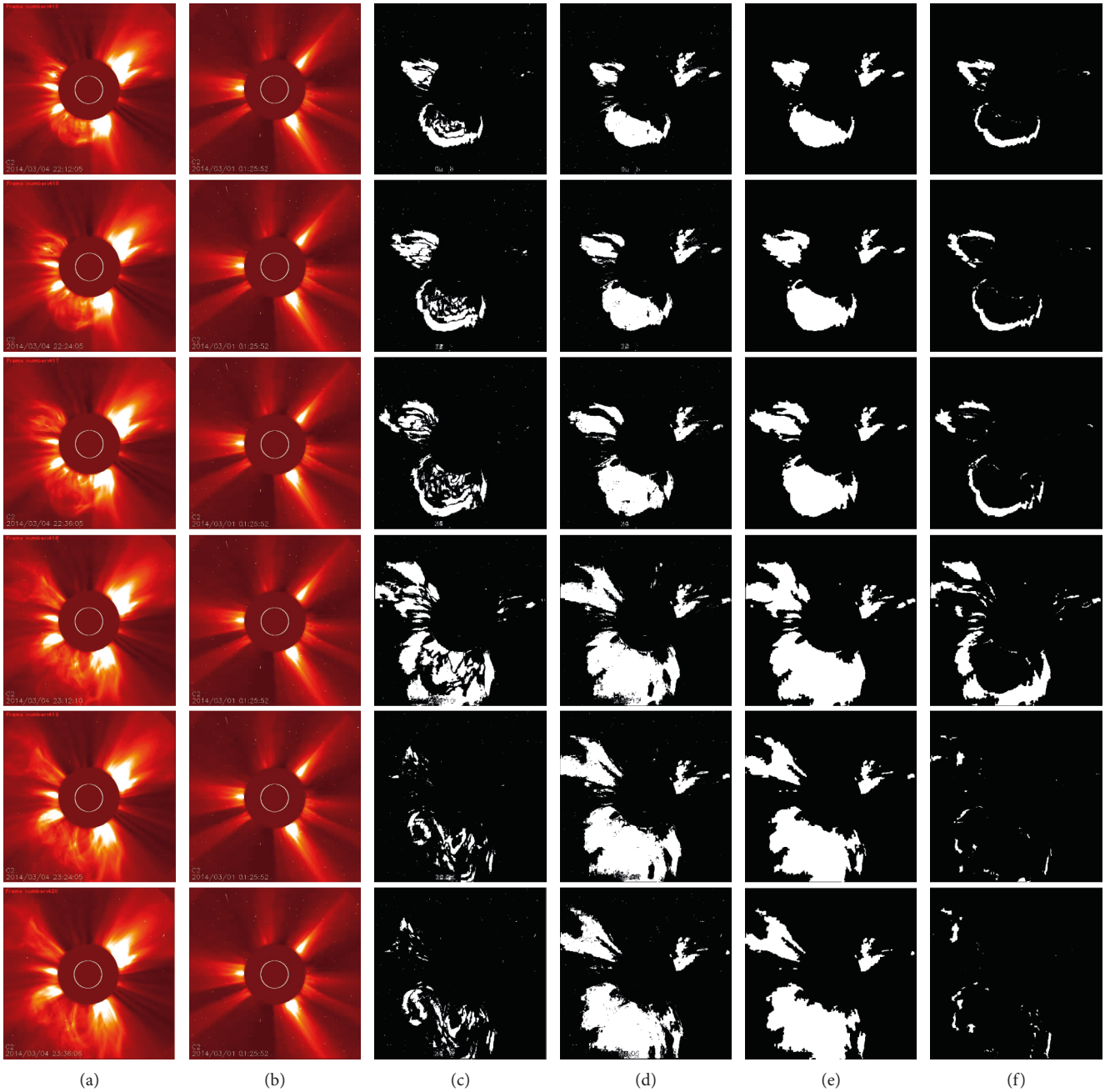


Figure 3: An experimental process graph of CME candidate area segmentation based on scene modeling. (a) LASCO C2 pseudocolor coronagraph images; (b) the reference background images; (c) the difference images between the two sequential frames; (d) the difference images between the current image and reference background; (e) the final candidate CME area images; (f) the changing region images of the candidate CME areas.

criterion of CME event: (1) the CME candidate region of two consecutive frames detected must be extended from the heliocentric; (2) since the start of the CME candidate region is detected, the region has enlarged gradually.

Besides, considering the angle range of the CMEs, we set the minimum angle threshold filtering noise. And the features of interest are intrinsically in polar coordinates owing to the spherical structure of the Sun. A polar transformation is applied to each candidate CME area image: the $[x, y]$ field of view (FOV), starting from the North of the Sun going counterclockwise, becomes a $[\theta, r]$ FOV, with θ the poloidal angle around the Sun and r the radial distance measured from the limb. This kind of transformation has been used in other CME detection algorithms [9, 11]. While transforming, we also rebin, from 1024×1024 pixels for the $[x, y]$ FOV to 360×360 pixels for the $[\theta, r]$ FOV. Through the appropriate r − range selection, the dark occulter and corner regions can easily be avoided. The radial FOV in polar coordinate image corresponds to 360 discrete points between 2.2 and 6.2 solar radii. We set the minimum detection

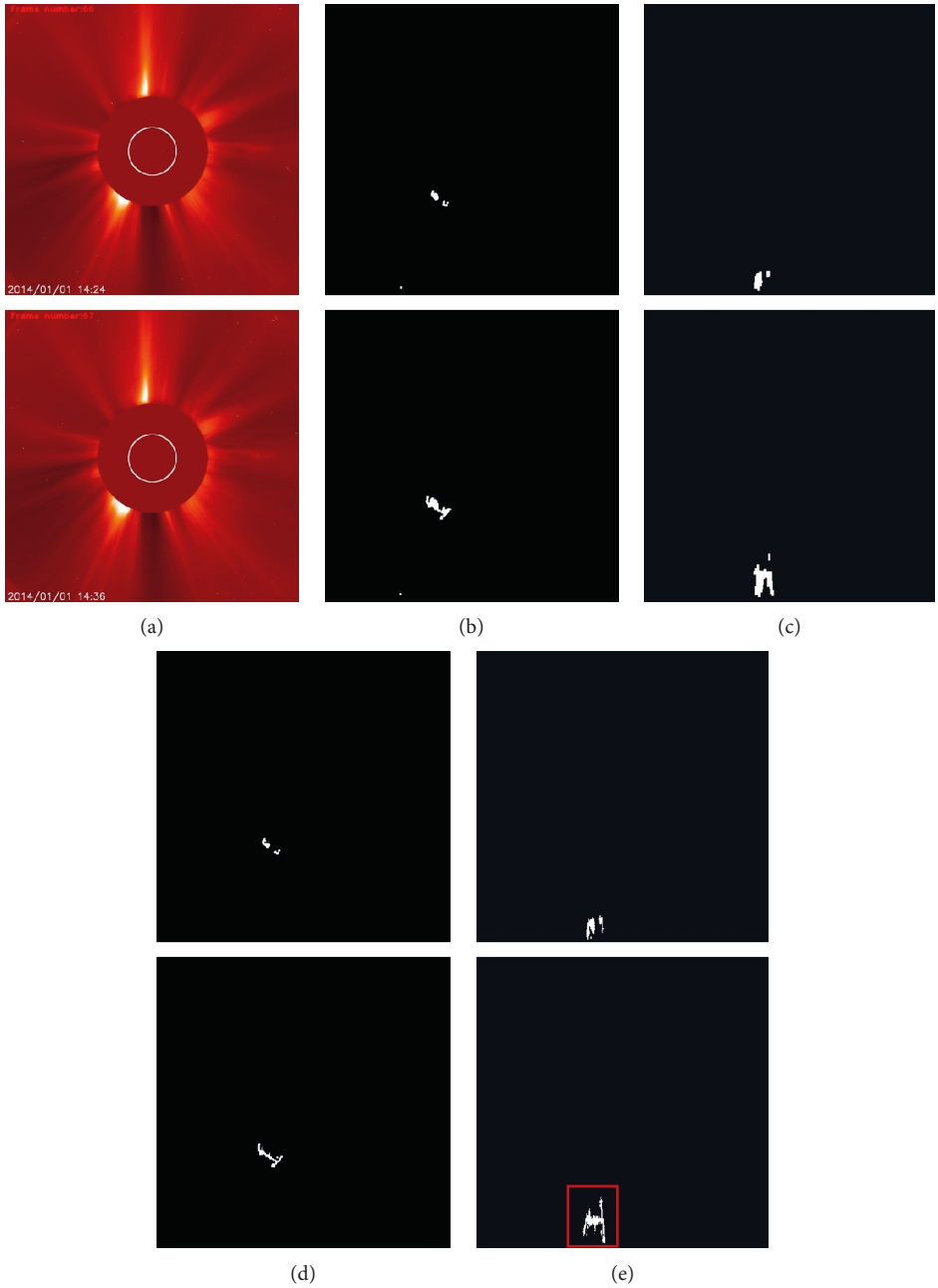


Figure 4: An example of CME detection process. (a) LASCO C2 pseudocolor coronagraph images; (b) the candidate CME area images; (c) the polar images of (b); (d) the increasing region images of the candidate CME areas; (e) the polar images of (d).

angle parameter d (refer to the CME list in CDAW in 2014, the minimum angle is 5 degrees, and we set $d = 4$).

5. Results and Validation

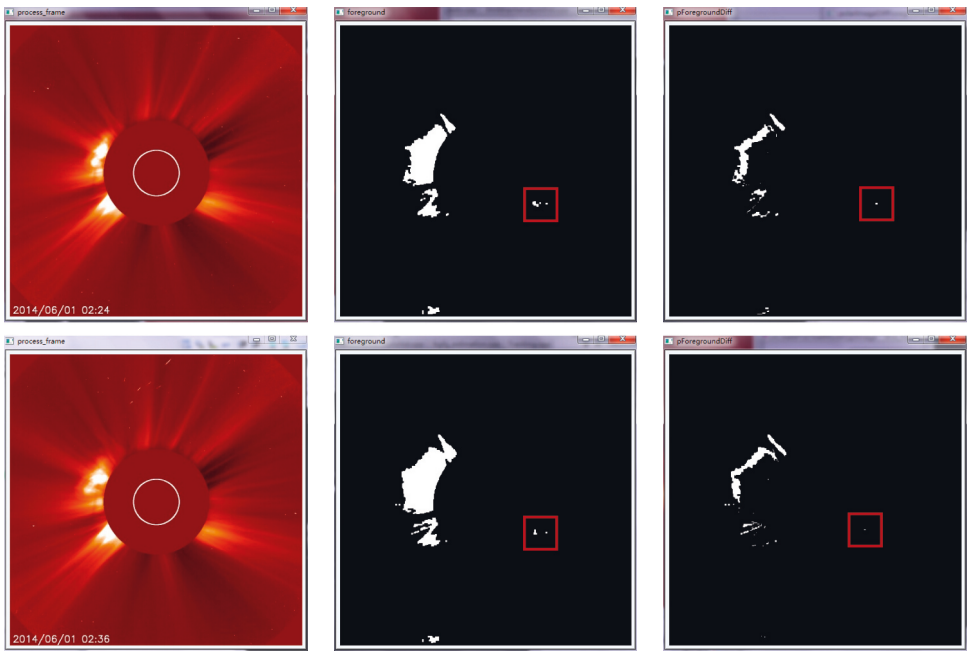
In this section, the visual examples and comparison on LASCO C2 pseudocolor coronagraph images are described, respectively.

5.1. Results. We present the results obtained by running the detection algorithms based on adaptive background learning technology.

An experimental process figure of the extraction candidate CME region is based on the scene modeling, we use data from the LASCO C2 pseudocolor coronagraph images, and $1024 * 1024$ image sequences are processed. Figure 3 is a CME candidate area segmentation process, which includes 6 frames (22 : 12 : 05, 22 : 24 : 05, 22 : 36 : 05, 23 : 12 : 10, 23 : 24 : 05, and 23 : 36 : 06 in 2014/03/04) processing result; column (a) is LASCO C2 pseudocolor coronagraph images; column (b) is the reference background images; column (c) is the difference images between the two sequential frames; column (d) is the difference images between the current image and reference background; column (e) is the final candidate

Table 1: Comparison the results of different CME detection methods.

Methods	Detected CME number	Detected CME number in the CDAW catalog	Accuracy rate (%)	False-negative rate (%)	Undetected CME number in the CDAW catalog
CDAW	259				
CORIMP	132	47	18.15	32.82	212
SEEDS	410	117	45.17	113.12	142
CACTus	188	85	32.82	39.77	174
Our method	283	189	72.97	36.29	70



(a) (b) (c)

Figure 5: A very poor CME event (appearance date-time (UT): 2014/06/01 02 : 24 : 05) detected by the adaptive background learning method. (a) LASCO C2 pseudocolor coronagraph images; (b) the candidate CME area images; (c) the increasing region images of the candidate CME areas.

CME area images; column (f) is the changing region images of the candidate CME areas.

Figure 4 is an example of CME detection process, which includes 2 frames (14 : 12 and 14 : 24 in 2014/01/01) processing result; column (a) is original coronal images; column (b) is candidate CME area images; column (c) is polar images of the candidate CME areas; column (d) is the increasing region images of the candidate CME areas; column (e) is the polar images of the increasing regions. The red box area in the last image is the detected CME area.

5.2. Validation and Comparison. Without loss of generality, we have chosen a full month of pseudocolor coronagraph image sequences observed by LASCO C2 in June 2014 as a test dataset for comparison. The manual CDAW list is used as a reference, and we compared the results of the adaptive background learning method with CORIMP, CACTus, and SEEDS catalog to verify the effectiveness of our proposed algorithm. The main comparisons include accuracy rate, false-negative rate, and the number of undetected CME events. The accuracy rate is the ratio of the total number of the CME events which were both detected by the automated method and recorded in the CDAW list to the total number of CME events in the CDAW catalog. The false-negative rate is the ratio of the total number of the CME events which the automated method did not detect but were recorded in the CDAW catalog to the total number of CME events in the CDAW catalog.

In the comparison experiments, for each CME event recorded in the CDAW catalog, if other automated methods detected the CME event within the time range of this event and within the angular range of this event, it considers that the automated method detect a CDAW list CME event. The comparison of the detection results by the adaptive background detection algorithm we propose with the other automated algorithms is shown in Table 1.

For the processing datasets, as shown in Table 1, the method we propose has a higher accuracy rate than other methods, and the false-negative rate is only higher than CORIMP method and lower than the SEEDS and CACTus

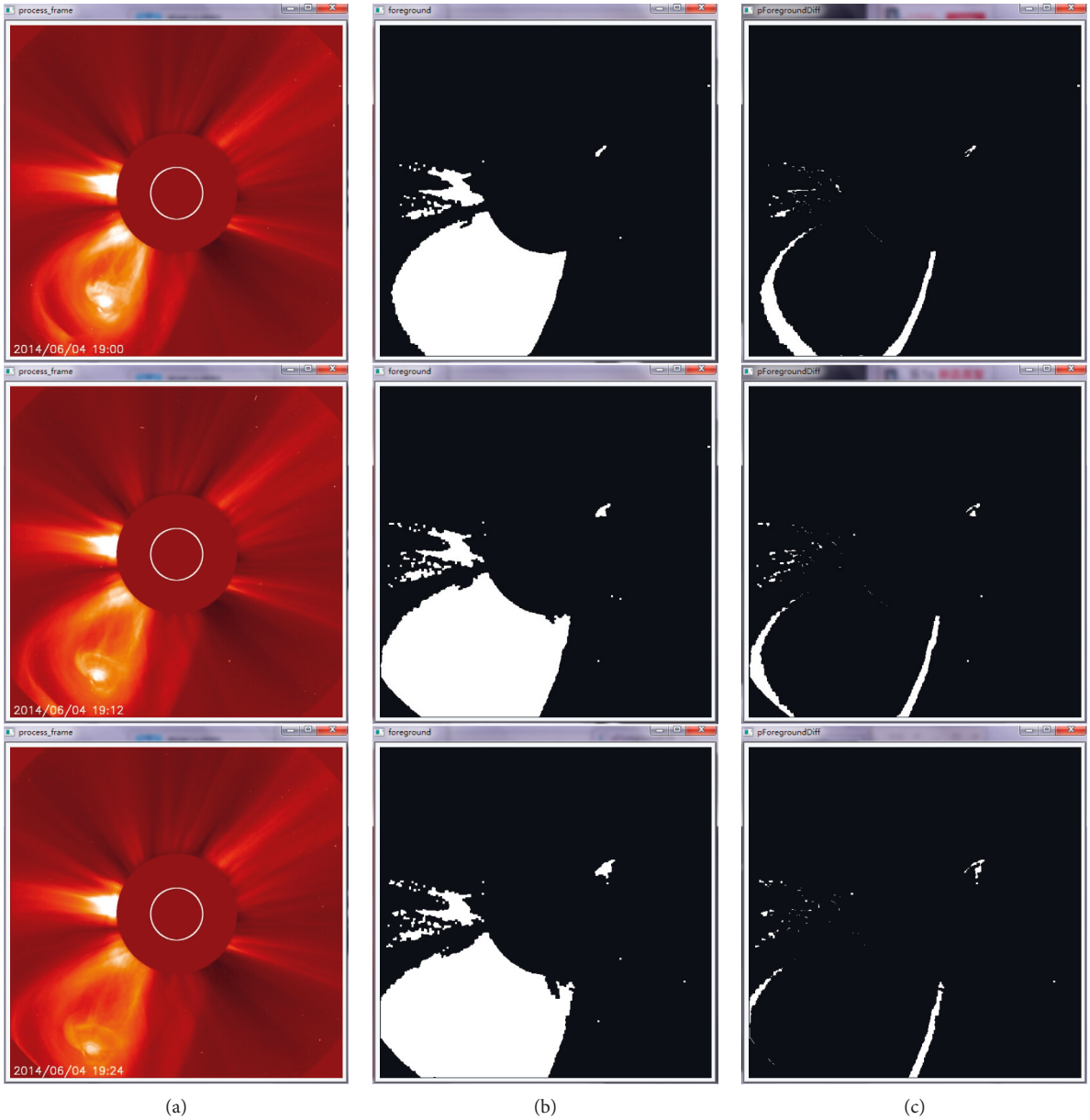


Figure 6: A poor CME event (appearance date-time (UT): 2014/06/04 19 : 00 : 05) detected by the adaptive background learning method. (a) LASCO C2 pseudocolor coronagraph images; (b) the candidate CME area images; (c) the increasing region images of the candidate CME areas.

methods. For the total detected CME number, our proposed method is higher than CORIMP and CACTus and is only lower than the SEEDS method. In terms of the undetected CME events in CDAW catalog number comparison, our method is the lowest.

In recent years, the CDAW catalog CME events are more finely recorded; especially, the very weak events in the Helmet streamers are recorded, and the number of the event recorded is also more and more. For example, CME event number recorded in 1996 is 206 and in 2014 it is 2477. Such changes make the automatic CME detection very difficult, so the novel detection method must detect the subtle changes in the coronal images. The automatically detecting CME method based on adaptive background learning can represent the dynamic scene very well, which is suitable for the event detection in dynamic scenes. Figure 5 is an example of the very weak CME event detection by our method, which occurred in the Helmet streamer area. This event was not detected by CORIMP, SEEDS, and CACTus methods and only recorded in CDAW catalog. Figure 5 shows two continuous coronagraph images, the candidate CME area images, the candidate CME change area images, and the very weak CME event area located in the red box. In the experiment, for the very weak CME events, our detection algorithm can only detect parts of the CME event, and this is the main reason to cause the misdetection. Figure 6 is a weak CME event detection process images, the weak CME event area is also located in

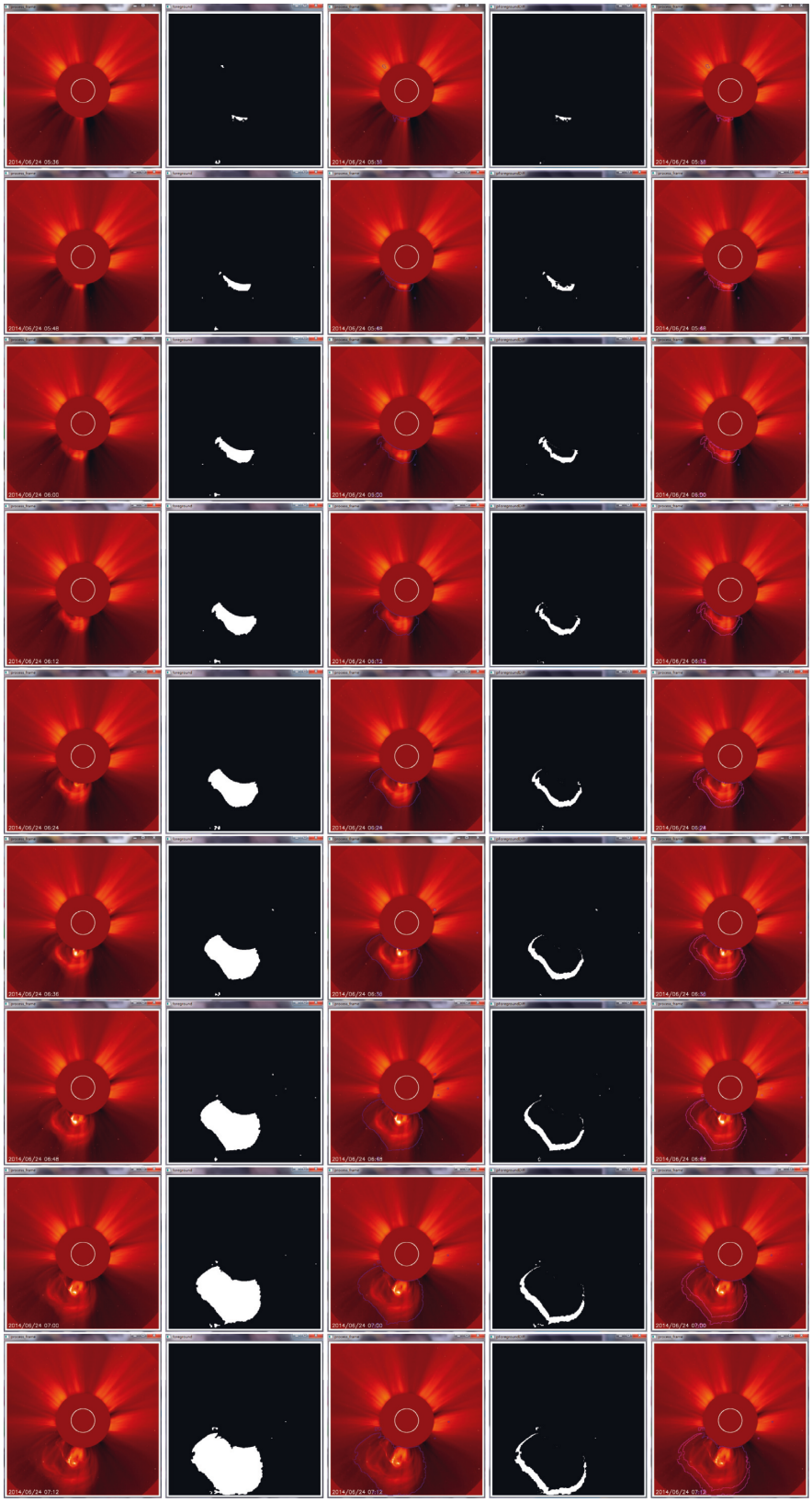
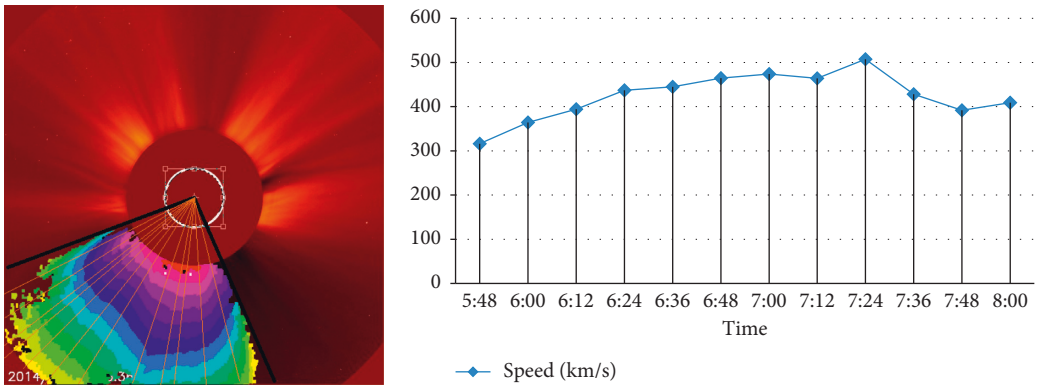


Figure 7: Morphological change description graph of a CME event (appearance date-time (UT): 2014/06/24 05 : 36 : 05) by using the intermediate images of our propose method.

Table 2: CME information comparison table on the event shown in Figure 7.

Methods	Comparison items				
	Central PA (deg)	Angular width (deg)	Linear (median) speed (km/s)	Min speed (km/s)	Max speed (km/s)
CDAW	158	177	633		
CORIMP	168	77	442		755
SEEDS	158	94	511		
CACTus	167	96	473	403	600
Our method	157	95	425	316	507



(a) (b)

Figure 8: Speed calculation sketch map and speed change curve chart.

the red box, the first column is the coronagraph images, the second column is the foreground detected by our method, and the third column is the change area of the foreground. This event was also not detected by CORIMP, SEEDS, and CACTus methods.

5.3. Computation of Information on CME. The information on CME events can be calculated conveniently by using the processed images. Figure 7 is a sequence of processed images of CME events. Our method detected the event's first C2 appearance date-time (UT): 2014/06/24 05 : 35 : 05, and the duration of this CME event is 2.6 hours, including 14 frames. In Figure 7, we show nine processed results of these frames. The first column is the coronal images; the second column is the detected candidate CME regions; the third column is the outline of the candidate CME regions which were marked by the blue curve; the fourth column is the changing areas of the candidate CME regions' images; the fifth column is the contour of the changing areas which were marked by the purple curve. We use the location information of the timestamp to filter the noise caused by the timestamp, so we can extract a more accurate CME area and ensure that the final calculated CME feature information is more accurate. The comparison of extracted information on this CME with other methods is shown in Table 2. During the calculation of our method, the speed of each frame can be calculated according to the change of each frame, the calculating schematic diagram, and the speed change curve shown in Figure 8.

In Table 2, the speed calculated by our method is the lowest; this is mainly due to the method did not use the frontier extreme point of the CME area in each frame to calculate the speed, but use the average value of the frontier sample points to calculate the speed. If the speed is calculated according to the extreme points of the frontier extreme point, the average speed of this CME event calculated by our method is 490 km/s, which is similar to the other automatic detection methods.

6. Discussion and Conclusion

In this paper, we have developed a new method that is capable of detecting, tracking, and calculating the information of CMEs in SOHO/LASCO C2 pseudocolor coronagraph images. The basic algorithm includes the following: (i) establishing and maintaining the background model of the coronal image sequences, (ii) detecting the candidate regions of CME based on Bayesian theorem, (iii) identifying the CME events, and (iv) calculating the information of CME events.

This novel method is based on adaptive background learning technology, and through the static and dynamic characteristics to model the background, this method can describe the complex background well, especially the dynamic changes in the background. So by using the proposed method to detect the CME in the superposition area with the Helmet streamers has more obvious advantage. At the same time, due to the background modeling learning, in this method, the information of multiframe images is counted. In this way, the influence on the results caused by the noise in the single-frame image can be suppressed and can enhance the robustness to CME detection. Our CME event identification method is based on the candidate CME

area. It uses the fact that the CME region always enlarged gradually; on the one hand, it can avoid the effect of the noise, and on the other hand, it can effectively track a complete CME event. Finally, through the detected region information on each frame, it is convenient and effective to extract the morphological and motion information of the CME event.

Automated methods such as CACTus, SEEDS, and CORIMP have a low detection rate of CMEs compared to CDAW catalogs made by human observers. This is mainly because the method of manual labeling in recent years has marked poor CME events, especially the poor events in the Helmet streamers. So new approaches are needed to detect subtle changes in the dynamic scenes, and the method we proposed has good performance in this aspect.

Similar to other automated methods, the biggest problem in the adaptive background learning method is the estimation of the various parameters and thresholds, such as quantized levels of the pixel information, learning update rate, and foreground detection threshold. For example, a small foreground detection threshold can reduce not only the false-negative rate, but also the accuracy rate. So the selection of these empirical values has a certain effect on the algorithm, and further investigation will be carried out in these areas. We are also planning to apply the method to corona images acquired by other devices.

Data Availability

The SOHO/LASCO data used to support the findings of this study are available from the SOHO/LASCO Instrument Homepage (http://lasco-www.nrl.navy.mil/).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 11603016 and 11873062), Key Scientific Research Foundation Project of Southwest Forestry University (Grant no. 111827), and Open Research Program of CAS Key Laboratory of Solar Activity, National Astronomical Observatories (KLSA201909). SOHO is a project of international cooperation between ESA and NASA. The SOHO/LASCO data used here are produced by a consortium of the Naval Research Laboratory (USA), Max-Planck-Institute für Aeronomie (Germany), Laboratoire d'Astronomie Spatiale (France), and the University of Birmingham (UK). The authors acknowledge the use of the CME catalog generated and maintained at the CDAW Data Center by NASA and the Catholic University of America in cooperation with the Naval Research Laboratory. The CACTus CME catalog is generated and maintained by the SIDC at the Royal Observatory of Belgium. The SEEDS CME catalog has been supported by NASA Living with a Star Program and NASA Applied Information Systems Research Program. The CORIMP CME catalog has been provided by Institute for Astronomy University of Hawaii.

References

[1] E. R. Christian, *Coronal Mass Ejections*, NASA/Goddard Space Flight Center, Maryland, USA, 2012.

[2] D. H. Hathaway, *Coronal Mass Ejections*, NASA/Marshall Space Flight Center, Maryland, USA, 2014.

[3] B. C. Low, "Coronal mass ejections," *Reviews of Geophysics*, vol. 25, no. 3, pp. 663–675, 2016.

[4] F. A. Cucinotta, "Space radiation risks for astronauts on multiple international space station missions," *PLoS One*, vol. 9, no. 4, Article ID e96099, 2014.

[5] R. Tousey, "The solar corona," in *Space Research XIII*, Akademic-Verlag, Berlin, Germany, 1973.

[6] G. E. Brueckner, R. A. Howard, M. J. Koomen et al., "The large angle spectroscopic coronagraph (LASCO)," *Solar Physics*, vol. 162, no. 1-2, pp. 357–402, 1995.

[7] R. A. Howard, J. D. Moses, and D. G. Socker, "Sun-earth connection coronal and heliospheric investigation (SECCHI)," in *Proceedings of the International Symposium on Optical Science and Technology*, San Diego, CA, USA, May 2000.

[8] N. Gopalswamy, S. Yashiro, G. Michalek et al., "The SOHO/LASCO CME catalog," *Earth, Moon, and Planets*, vol. 104, no. 1–4, pp. 295–313, 2009.

[9] E. Robbrecht and D. Berghmans, "Automated recognition of coronal mass ejections (CMEs) in near-real-time data," *Astronomy & Astrophysics*, vol. 425, no. 3, pp. 1097–1106, 2004.

[10] Y. Boursier, A. Llebaria, F. Goudail et al., "Automatic detection of coronal mass ejections on LASCO-C2 synoptic maps," *Proceedings of SPIE—The International Society for Optical Engineering*, vol. 5901, 2005.

[11] O. Olmedo, J. Zhang, K. Poland, and K. Borne, "Automatic detection and tracking of coronal mass ejections in coronagraph time series," *Solar Physics*, vol. 248, no. 2, pp. 485–499, 2008.

[12] C. A. Young and P. T. Gallagher, "Multiscale edge detection in the corona," *Solar Physics*, vol. 248, no. 2, pp. 457–469, 2008.

[13] N. A. Goussies, M. E. Mejail, J. Jacobo, and G. Stenborg, "Detection and tracking of coronal mass ejections based on supervised segmentation and level set," *Pattern Recognition Letters*, vol. 31, no. 6, pp. 496–501, 2010.

[14] J. P. Byrne, P. T. Gallagher, R. T. J. McAteer, and C. A. Young, "The kinematics of coronal mass ejections using multiscale methods," *Astronomy & Astrophysics*, vol. 495, no. 1, pp. 325–334, 2009.

[15] P. T. Gallagher, C. A. Young, J. P. Byrne, and R. T. J. McAteer, "Coronal mass ejection detection using wavelets, curvelets and ridgelets: applications for space weather monitoring," *Advances in Space Research*, vol. 47, no. 12, pp. 2118–2126, 2011.

[16] Z. Zhao-xian, W. Ya-li, and L. Jin-sheng, "A method to automatic detecting coronal mass ejections in coronagraph based on frequency spectrum analysis," in *Proceedings of the 2012 International Conference of Modern Computer Science and Applications*, pp. 223–227, Springer, Berlin, Heidelberg, June 2013.

[17] A. Bemporad, V. Andretta, M. Pancrazzi et al., "On-board CME detection algorithm for the solar orbiter-METIS coronagraph," *Proceedings of SPIE—The International Society for Optical Engineering*, vol. 9152, p. 91520K, 2014.

[18] L. Zhang, J. Yin, J. Lin et al., "Detection of coronal mass ejections using multiple features and space-time continuity," *Solar Physics*, vol. 292, no. 7, p. 91, 2017.

[19] R. Patel, K. Amareswari, V. Pant et al., "Onboard automated CME detection algorithm for the visible emission line coronagraph on ADITYA-L1," *Solar Physics*, vol. 293, no. 7, pp. 1–25, 2018.

[20] D. B. Dhuri, S. M. Hanasoge, and M. C. M. Cheung, "Machine learning reveals systematic accumulation of electric current in lead-up to solar flares," *Proceedings of the National Academy of Sciences*, vol. 116, no. 23, pp. 11141–11146, 2019.

[21] X. Huang, H. Wang, L. Xu, J. Liu, R. Li, and X. Dai, "Deep learning based solar flare forecasting model. I: results for line-of-sight magnetograms," *The Astrophysical Journal*, vol. 856, no. 1, p. 7, 2018.

[22] P. Wang, Y. Zhang, L. Feng et al., "A new automatic tool for CME detection and tracking with machine learning techniques," 2019, https://arxiv.org/abs/1907.08798.

[23] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[24] L. Li, W. Huang, I. Y. H. Gu et al., "Foreground object detection from videos containing complex background," in *Proceedings of the Eleventh ACM International Conference on Multimedia*, January 2003.

[25] Q. Mo, F. Dai, D. Liu, J. Qin, Z. Xie, and T. Li, "Development of private processes: a refinement approach," *IEEE Access*, vol. 7, pp. 31517–31534, 2019.

[26] O. Olmedo, *A study of the initiation process of coronal mass ejections and the tool for their auto-detection*, Ph.D. Thesis, *College of Science*, pp. 137–149, 2011.

Hindawi
Advances in Astronomy
Volume 2019, Article ID 2397536, 11 pages
https://doi.org/10.1155/2019/2397536

Research Article

Extracting Filaments Based on Morphology Components Analysis from Radio Astronomical Images

M. Zhu,[1] W. Liu,[1] B. Y. Wang,[1] M. F. Zhang,[2] W. W. Tian,[3] X. C. Yu,[1] T. H. Liang,[1] D. Wu,[2] D. Hu,[1] and F. Q. Duan[1]

[1] *College of Information Science and Technology, Beijing Normal University, Beijing, China*
[2] *Key Laboratory of Optical Astronomy, National Astronomical Observatory of China, Beijing, China*
[3] *The University of Chinese Academy of Sciences, Beijing, China*

Correspondence should be addressed to X. C. Yu; yuxianchuan@163.com

Received 17 October 2018; Accepted 3 March 2019; Published 2 June 2019

Guest Editor: Junhui Fan

Filaments are a type of wide-existing astronomical structure. It is a challenge to separate filaments from radio astronomical images, because their radiation is usually weak. What is more, filaments often mix with bright objects, e.g., stars, which makes it difficult to separate them. In order to extract filaments, A. Men'shchikov proposed a method "getfilaments" to find filaments automatically. However, the algorithm removed tiny structures by counting connected pixels number simply. Removing tiny structures based on local information might remove some part of the filaments because filaments in radio astronomical image are usually weak. In order to solve this problem, we applied morphology components analysis (MCA) to process each singe spatial scale image and proposed a filaments extraction algorithm based on MCA. MCA uses a dictionary whose elements can be wavelet translation function, curvelet translation function, or ridgelet translation function to decompose images. Different selection of elements in the dictionary can get different morphology components of the spatial scale image. By using MCA, we can get line structure, gauss sources, and other structures in spatial scale images and exclude the components that are not related to filaments. Experimental results showed that our proposed method based on MCA is effective in extracting filaments from real radio astronomical images, and images processed by our method have higher peak signal-to-noise ratio (PSNR).

1. Introduction

A substantial part of interstellar medium exists in the form of a fascinating web of omnipresent filamentary structures [1], called filaments. The astronomical filament is first discovered in the Milky Way. Along with the development of telescopes, various filaments come into sight. Among them the filaments in star-forming regions are the most fascinating, many magnetohydrodynamic (MHD) simulations have shown that giant molecular clouds (GMCs) primarily evolve into filaments before they collapse to form stars [2, 3]. Recent observations also confirm these simulations [4, 5]. Since the formation of massive stellar objects is still unclear, further research on filaments is essential. Filaments in Galactic and cosmological fields are also important. Studies have argued that low mass galaxies got their gas through "cold accretion", which is often directed along filaments [6, 7]. Some researchers paid more attention to the large-scale filaments of the universe [8], which may give clues to better understand the slightly nonuniform cosmic microwave background (CMB) and the birth of the first generation of stars. In addition, filaments have been observed in other objects, such as supernova remnants (SNR) [9] and protoplanetary disk [10].

The fact that many filaments are fuzzy in images causes difficulty to distinguish them from background and surrounding objects. Schneider et al. [11] investigated spatial and density structure of the Rosette molecular cloud, by applying a curvelet analysis, a filament-tracing algorithm (DisPerSE), and probability density functions (PDFs) on Herschel column density maps. Hennebelle et al. [12] showed a method based on adaptive mesh refinement magneto hydrodynamic simulations, which treat self-consistently cooling and self-gravity. Tugay [13] proposed a layer smoothing method,

which described cellular large-scale structure of the universe (LSS) as a grid of clusters with density larger than a limited value, to detect extragalactic filaments. Men'shchikov [14] proposed a multi-scale filaments extraction method named getfilaments, which decomposed a simulated astronomical image containing filaments into spatial images at different scales to prevent interaction influence of different spatial scale structures. The getfilaments works well in simulated images and has been used to identify filaments for real astronomical images, e.g., the far-infrared images of Musca cloud observed with Herschel [15]. However, getfilaments might exclude some tiny structure of filaments in astronomy images, because it removes tiny structures just by counting connected pixels number, and filaments in astronomy images are usually weak.

In this paper, we develop an improved method based on morphology components analysis (MCA) and getfilaments. MCA is able to decompose the image into morphological components based on different features from the perspective of mathematical morphology and is often used in image restoration, separation, and decomposition [16–19]. The basic idea of MCA decomposition algorithm is to choose two dictionaries: smooth dictionary and texture dictionary, to represent morphology components [20]. We can design different dictionaries to represent different sparse components in the image. Smooth dictionary produces the decomposed smooth component which carries the geometric and piecewise smooth information of the image, and texture dictionary produces the decomposed texture component which carries the marginal and edge information.

The paper is structured as follows. The improved method named filament extraction algorithm based on MCA is described in Section 2. Section 3 is devoted to discussing experimental results of our method and comparing our method with the getfilaments method by employing data from GALFA-HI of Arecibo.

2. Filament Extraction Algorithm Based on MCA

2.1. MCA Model. MCA was proposed by Starck et al. [17]. MCA is a kind of decomposition algorithm based on signal sparsity and morphological diversity. MCA assumes that signals are linear combinations of several morphological components, and each morphological component can be sparsely represented on its own dictionary.

We assume that image $\mathbf{x}$ comprises M different morphological components: $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_M$. We design different dictionaries $\mathbf{D}_i$ for different morphological components $\mathbf{x}_i$ and assume all components mix together linearly. The image $\mathbf{x}$ as an one-dimensional vector of length M can then be represented as follows:

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}, \tag{1}$$

where the matrix $\mathbf{D} = [D_1, \ldots, D_P] \in R^{M\times P}$(typically, $M \ll P$) is a dictionary. $\boldsymbol{\alpha} \in R^P$ is the vector of sparse coefficients. The equivalent constrained optimization problem is as follows:

$$\begin{aligned} &\left\{\boldsymbol{\alpha}_1^{\text{opt}}, \boldsymbol{\alpha}_2^{\text{opt}}, \ldots, \boldsymbol{\alpha}_P^{\text{opt}}\right\} \\ &\quad = \arg\min_{\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_P\}} \sum_{k=1}^{P} \left\|\boldsymbol{\alpha}_k\right\|_1, \\ &\text{subject to}: \quad \mathbf{x} = \sum_{k=1}^{P} \mathbf{D}_k \boldsymbol{\alpha}_k. \end{aligned} \tag{2}$$

However, this model does not take into account factors that may lead to the failure of the image decomposition, such as noise. When noise exists in the image $\mathbf{x}$, the vector $\{\boldsymbol{\alpha}_1^{\text{opt}}, \boldsymbol{\alpha}_2^{\text{opt}}, \ldots, \boldsymbol{\alpha}_P^{\text{opt}}\}$ might be not sparse since noise cannot be sparsely represented. For this kind of noise, we put the noise in the error item to achieve the sparse decomposition of the image $\mathbf{x}$. The constraint in (2) is modified as follows:

$$\begin{aligned} &\left\{\boldsymbol{\alpha}_1^{\text{opt}}, \boldsymbol{\alpha}_2^{\text{opt}}, \ldots, \boldsymbol{\alpha}_P^{\text{opt}}\right\} \\ &\quad = \arg\min_{\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_P\}} \sum_{k=1}^{P} \left\|\boldsymbol{\alpha}_k\right\|_1, \\ &\text{subject to}: \quad \left\|\mathbf{x} - \sum_{k=1}^{P} \mathbf{D}_k \boldsymbol{\alpha}_k\right\|_1 \le \epsilon. \end{aligned} \tag{3}$$

where ϵ represents the noise level in the image $\mathbf{x}$.

2.2. MCA Decompostion Algorithm. In this paper, we focus on the image decomposition into two components: cartoon layer and texture layer. Cartoon layer contains cartoon and piecewise smooth information, and texture layer may contain other texture information, marginal information, and noises [21, 22]. Studies [23, 24] have shown that noises exist in both cartoon and texture layer. In other words, the smooth part not only contains the majority of the useful information, but also contains a small part of the noise. If we set the same threshold of noise variance for the whole image, rather than calculating the threshold for each part of the image, some useful information might be removed. We therefore introduce the MCA decomposition algorithm to process an image into smooth (cartoon) layer and texture layer.

We assume that matrix $\mathbf{D}_t$ is the dictionary matrix of the texture layer and that $\mathbf{D}_c$ is the dictionary matrix of the cartoon layer. A solution for the decomposition could be obtained by relaxing the constraint in (3) to become an approximate one:

$$\begin{aligned} \left\{\boldsymbol{\alpha}_t^{\text{opt}}, \boldsymbol{\alpha}_c^{\text{opt}}\right\} &= \arg\min_{\{\boldsymbol{\alpha}_t,\boldsymbol{\alpha}_c\}} \left\|\boldsymbol{\alpha}_t\right\|_1 + \left\|\boldsymbol{\alpha}_c\right\|_1 \\ &\quad + \lambda \left\|\mathbf{x} - \mathbf{D}_t\boldsymbol{\alpha}_t - \mathbf{D}_c\boldsymbol{\alpha}_c\right\|_2^2, \end{aligned} \tag{4}$$

where λ is a Lagrange operators. Define $\mathbf{x}_t = \mathbf{D}_t\boldsymbol{\alpha}_t$ and $\mathbf{x}_c = \mathbf{D}_c\boldsymbol{\alpha}_c$. Given $\mathbf{x}_t$, we can recover $\boldsymbol{\alpha}_t$ as $\boldsymbol{\alpha}_t = \mathbf{D}_t^+\mathbf{x}_t$, where $\mathbf{D}_t^+$ is the Moore-Penrose pseudoinverse of $\mathbf{D}_t$. In order to get

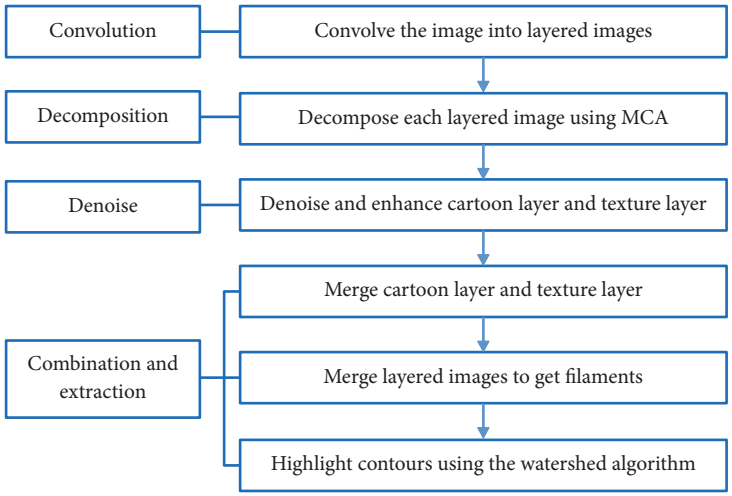


Figure 1: Flow of the filament extraction algorithm.

piecewise smooth component, add a TV (Total Variation) penalty [25] to fit the smooth layer. TV is used to damp ringing artifacts near edges and oscillating. Put these back into (4), and, thus, we obtain the following:

$$\left\{\mathbf{x}_t^{\text{opt}}, \mathbf{x}_c^{\text{opt}}\right\} = \arg\min_{\{\mathbf{x}_t, \mathbf{x}_c\}} \left\|\mathbf{D}_t^+ \mathbf{x}_t\right\|_1 + \left\|\mathbf{D}_c^+ \mathbf{x}_c\right\|_1 + \lambda \left\|\mathbf{x} - \mathbf{x}_t - \mathbf{x}_c\right\|_2^2 + \gamma \left\|TV\left(\mathbf{x}_c\right)\right\|. \tag{5}$$

γ is the TV regularization parameter, in multiscale method, a lower γ is able to remove the artefacts caused by curvelet. $TV(\mathbf{x}_c)$ is a measure of the amount of oscillations in the cartoon layer. Penalizing with TV, the cartoon layer is closer to the piecewise smooth image. However, TV suffers from the so-called staircase effect that impacts the quality of images reconstruction. The adaptive TV [26] and the higher order derivative [27] are solutions to reduce the staircase effect.

Then we discuss the choice of dictionaries for the cartoon and the texture layer. Appropriate dictionaries are very important for sparse representations over the image. Generally, the choice of dictionaries depends on experiences. The structure in the dictionary is more matched with the image easier to form a sparse representation. The commonly used dictionary of MCA includes wavelet transform, ridgelet transform, curvelet transform, discrete cosine transform (DCT), and so on. The two dictionaries used in this paper are described as follows.

First, we choose curvelet as the dictionary for cartoon layer. The curvelet transform based on the multiscale ridgelet transform was proposed by Cand & Donoho [28]. It first decomposes the image into a set of wavelet bands and then analyzes each band with the ridgelet transform at different scale levels. The curvelet transform performs well at the detection of anisotropic structures, smooth curves, and edges of different lengths [29].

We next choose the local DCT as the dictionary for texture layer. DCT is a variant of the discrete Fourier transform (DFT). It uses a symmetric signal extension to replace complex analysis with real numbers and is appropriate for the sparse representation of the texture and periodic part of the image.

2.3. Filament Extraction Algorithm Based on MCA. The shape of the filaments can be obtained by applying the above filament extraction algorithm to radio astronomical images. We finally use the watershed algorithm to highlight filaments.

The watershed algorithm [30, 31] is based on mathematical morphology. The watershed algorithm segments an image into nonoverlapping regions and gets a pixel width and continuous boundary for the purpose of extracting and identifying a specific area [32, 33]. A grayscale image can be viewed as a topographic surface. A high grayscale value of a pixel denotes a peak or hill while a low grayscale denotes a valley. Each local minimum of pixels and the affected region are called a catchment basin, and the boundary of catchment basins forms the watershed. By filling each isolated valley (local minimum) with differently colored water (labels), the region of influence of each local minimum gradually expands outwards. The adjacent regions then converge, and the boundaries that form the watershed appear.

The whole filament extraction algorithm (https://github.com/MiWBY/MCA) can be roughly divided into four steps (as shown in Figure 1).

(1) Convolution. First, using a Gaussian filter, we convolve the original images into a series of layered images. Different full widths at half maximum can be set for different image layers:

$$X_j = G_{j-1} * X - G_j * X \quad \left(j = 1, 2, \ldots, N_s\right). \tag{6}$$

where X is the original image, X_j is the jth subimage after convolution, G_{j-1} and G_j are different Gaussian beams for different spatial components, $*$ is the convolution operation, and N_s is the number of the layers.

In this process, structures at different scales in the astronomical images can be separated into different layers (subimages), and each layer contains similar scales, which

make the input sources become simpler in the later denoising and extraction process.

(2) Decompostion. We apply the MCA algorithm to each layered image so that each layered image is decomposed into a cartoon layer and a texture layer. Here we use the curvelet as the dictionary for the cartoon layer and local DCT as the dictionary for the texture layer as described in Section 2.2. The cartoon layer contains most of filaments and low-frequency noise, and the texture layer contains sources, high-frequency noise, and small part of filaments.

Starck et al. [17] proposed the MCA decomposition algorithm based on BCR algorithm (Block Coordinate Relaxation). The algorithm is given as follows. Input: The subimage X_j after convolution, which is described as the input image $\mathbf{x}$ here, dictionary $\mathbf{D}_c$ of the cartoon layer, dictionary $\mathbf{D}_t$ of the texture layer, number of iterations $L_{\max}$, and the threshold $\delta = \lambda \cdot L_{\max}$.

Output: Cartoon layer $\mathbf{x}_c$ and texture layer $\mathbf{x}_t$.

(1) Initialize $L_{\max}$, and $\lambda = k * \epsilon$ (typically, k = 3), where ϵ is the value of noise level. Then the threshold $\delta = \lambda \cdot L_{\max}$.

(2) For $j = 1 : L_{\max}$

For $k = 1 : P$

(i) Update $\mathbf{x}_c$ assuming $\mathbf{x}_t$ is fixed:

(a) Caculate the residual $\mathbf{r} = \mathbf{x} - \mathbf{x}_c - \mathbf{x}_t$.
(b) Calculate $\boldsymbol{\alpha}_c = \mathbf{D}_c^+(x_c + \mathbf{r})$.
(c) Soft thresholding the coefficient $\boldsymbol{\alpha}_c$ with the δ threshold and obtain $\widehat{\boldsymbol{\alpha}}_c$.
(d) Reconstruct $\mathbf{x}_c$ by $\mathbf{x}_c = \mathbf{D}_c\widehat{\boldsymbol{\alpha}}_c$.

(ii) Update $\mathbf{x}_t$ with the above method

Apply the TV correction by $\mathbf{x}_c = \mathbf{x}_c - \mu\gamma(TV(\mathbf{x}_c)/\mathbf{x}_c)$, where μ is the minimum parameter, and is chosen by a line-decreasing the overall penalty function, or as a fixed step-size of moderate value that guarantees convergence.

(3) Update the threshold by $\gamma = \gamma - \lambda$.

If $\gamma > \lambda$, return step 2. Else, finish.

(3) Denoise. First, we denoise each layer using the iterative cleaning algorithm proposed by Men'shchikov et al. [34]. The cleaning algorithm employs a global intensity threshold for single-scale images, as the larger-scale background has been effectively filtered out by the spatial decomposition. This iterative algorithm automatically finds a cut-off level that separates the signal of important sources from the noise and background at each scale. Next, we enhance details for both the cartoon layer and texture layer.

(4) Combination and Extraction. To get the extracted filaments, we first merge the cartoon layer and textual layer for each layered image. Because filaments are irregular, and structures of filaments exist in both cartoon layer and texture layer, it is not appropriate to use just one component to represent the filaments. For example, if we just use cartoon layer to represent filaments, filaments may lose some texture. Thus, we merge the cartoon layer and textual layer to represent filaments better. Next, the layered subimages are added together to produce the filaments. Finally, we apply the watershed algorithm to highlight the contours of filaments.

By applying MCA to decompose a real image, new features (components) can be obtained. This leads to better image separability. Furthermore, the smooth components have a better signal-to-noise ratio than the original image.

3. Extraction Results

3.1. Results for a Simulated Image. Before applying our method to real radio astronomical images, we simulated an image that is composed of a straight filament with 37″ size of FWHM, a string of sources with 24″ size of FWHM, a simple background with 4000″ size of FWHM, and a moderate-level noise with noise level=1.05 to test the improved algorithm (Figure 2(a)). The simulation method is the same as that mentioned in Men'shchikov et al. [14]. In the simulated image, there is only one spatial component, while our method assumes there are many spatial components, which is similar to real astronomical images. In other words, even if there is only one spatial component, our method will also treat it as many components.

We first extract filaments using MCA method without convolution and denoising (Figure 2). In Figure 2(c), texture layer (especially in the area marked by red box) still contains part of filaments structures, which means just using cartoon layer to represent filaments is insufficient. So it is necessary to combine cartoon layer and texture layer. However, noises and sources also exist in texture layer. If the two layers are combined directly, the reconstructed filaments contains noises and sources (Figure 2(d)), so denoising is necessary before combination.

Next, extracted results obtained using our improved method are shown in Figure 3. Compared to Figure 2(d), the reconstructed filaments in Figure 3(e) contain less noise. The edge of the filament is unreal as the result of decomposition.

3.2. Results for Astronomical Images

3.2.1. Decomposition and Denosing Results. Aiming at real radio astronomical images, we compare the extraction results of our method with those of the getfilaments method. We employ data from GALFA-HI of Arecibo as example images. The equatorial coordinates of the objects are (12.00h, +10.35°), and the object name is 'GALFA-HI RA+DEC Tile 004.00+02.35'. The data cube contains 2048 images at different velocity (with respect to the local standard stationary system). Here we select the 715th image from the 2048 original images as the experimental image (Figure 4). In the experimental images, filaments describe significantly elongated structures. After convolution of the 715th image, we obtain 99 layers (subimages) at different scales (Figure 5) and select the 40th subimage as a comparison example (Figure 5(b)). In order to display the image properly and improve visual

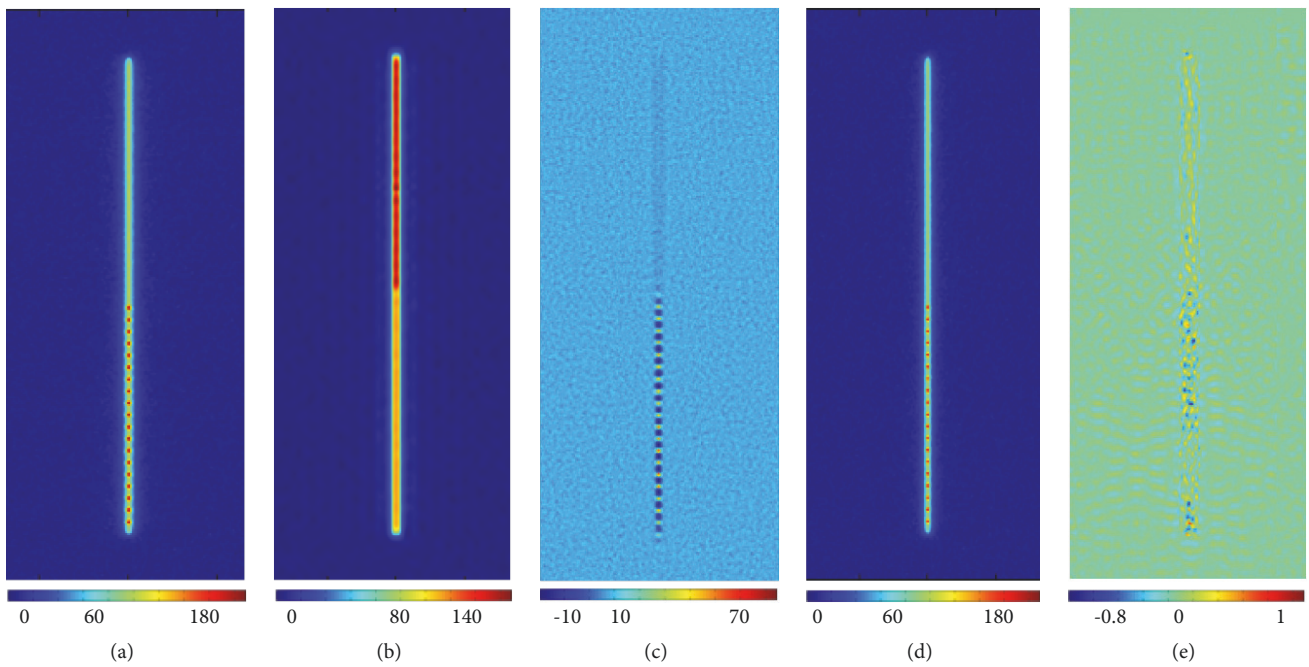


Figure 2: Results of the simulated images obtained using MCA without convolution and denoising. (a) Original simulated image. (b) Cartoon layer. (c) Texture layer. (d) Reconstructed filaments without denoising. (e) Residuals.

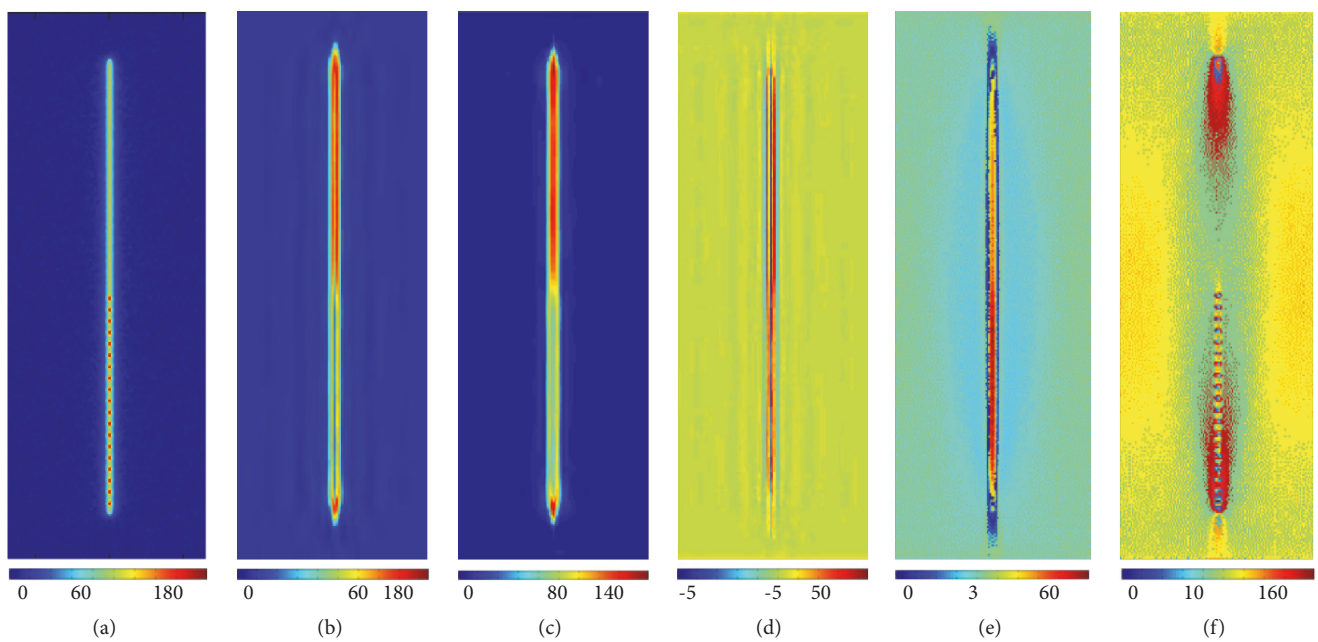


Figure 3: Extraction results for the simulated image obtained using our method. (a) Original simulated image. (b) The 40th subimage after convolution. (c) Cartoon layer of the 40th subimage after decomposition and denoising. (d) Texture layer of 40th subimage after decomposition and denoising. (e) Reconstructed filaments. (f) Residuals.

contrast between getfilaments and our method, we mark the image with different colors according to the intensity (Unit: MJy/sr) in the image.

First, we apply the MCA algorithm to process image layers before applying the iterative cleaning algorithm. As described in Section 2.2, we choose curvelet as the cartoon layer dictionary and LDCT as the texture layer dictionary for MCA. We decompose each layered image to get the cartoon layer and texture layer (Figure 6). The cartoon layer contains smooth parts of the image and retains most of the low-frequency information of the filament in the layered image. The frequency of the texture layer is higher. The texture

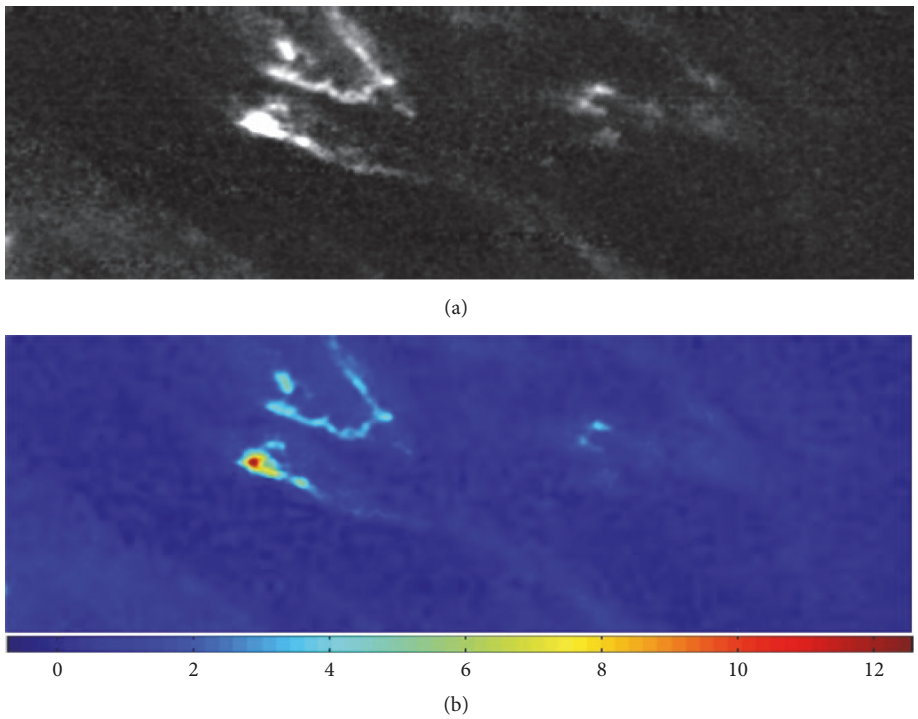


Figure 4: The 715th original image from GALFA-HI. (a) Original image for experiments. (b) Colored image for better visual contrast.

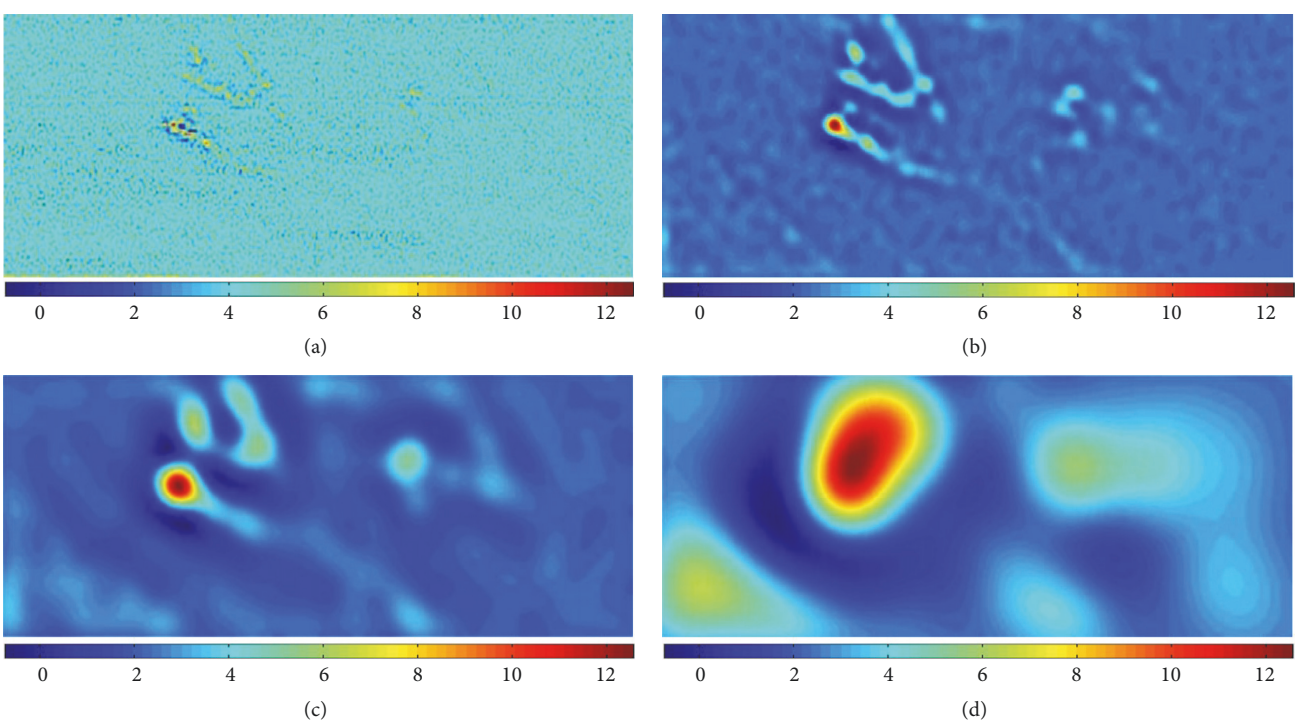


Figure 5: Images at different scales after convolution of the 715th image. Choose the 40th subimage as comparison example. (a) The 1st subimage. (b) The 40th subimage. (c) The 60th subimage. (d) The 80th subimage.

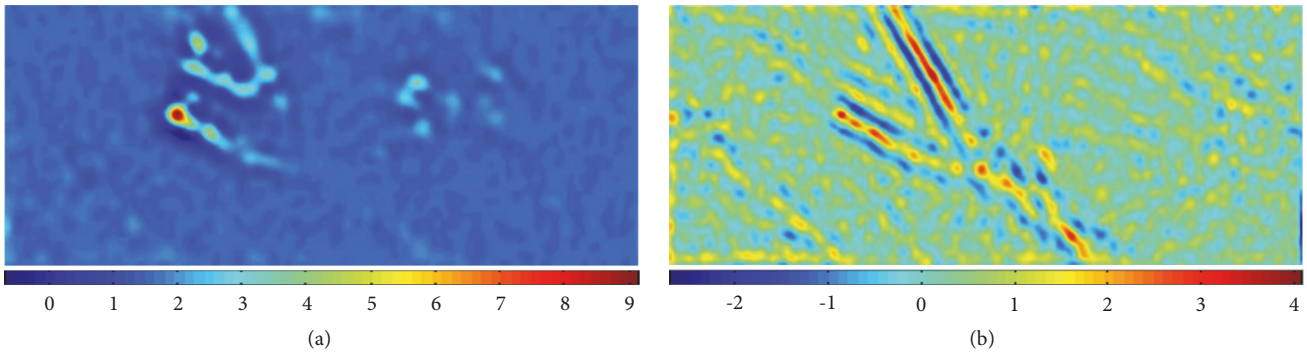


Figure 6: Decomposition results of the 40th subimage obtained using MCA. (a) The cartoon layer obtained using the MCA algorithm. (b) The texture layer obtained using the MCA algorithm.

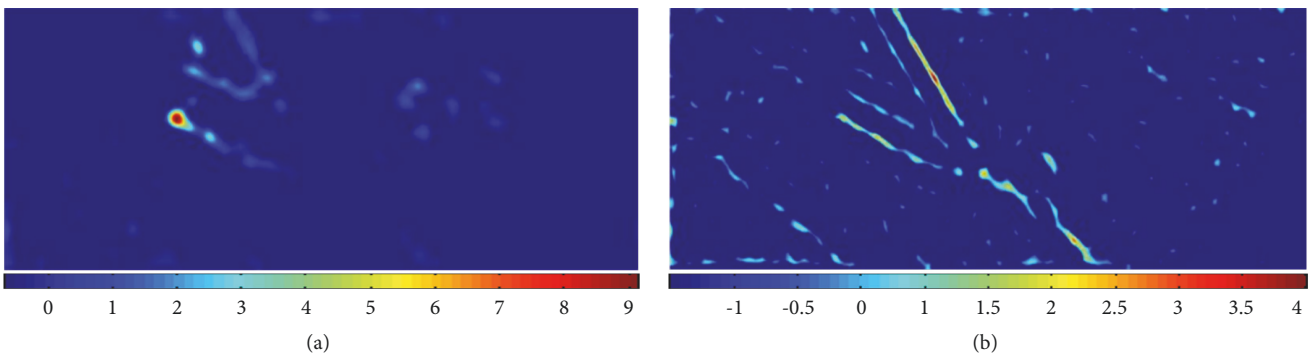


Figure 7: Denoising results for the cartoon layer and texture layer obtained using the iterative cleaning algorithm. (a) Denoising results for the cartoon layer of the 40th subimage. (b) Denoising results for the texture layer of the 40th subimage.

layer contains more edge information that is difficult to be distinguished visually in the layered image. The texture layer is also part of the filament. The texture layer also contains part of the filament. It shows some artefacts that might be caused by DCT in texture layer; these artefacts can be removed after denoising.

We then set different reasonable threshold for cartoon layer and texture layer, respectively, and apply the iterative cleaning algorithm to the cartoon layer and texture layer to remove noise (as shown in Figure 7). Compared to Figure 6(b), Figure 7(b) almost contains no artefacts.

Next, the cartoon layer and texture layer are fused according to the intensity ratio (e.g., the information of the texture layer is expanded by a factor of 5). Small structures can then be retained and interference information is removed. Processing results after fusion are shown in Figure 8(c). To allow comparison with our method, we also use the iterative cleaning algorithm directly to process the 40th subimage; the results are shown in Figure 8(b).

As seen in Figure 8(b), most of the noises in the 40th subimage are removed by cleaning. However, getfilaments method sets only one noise threshold, and the values less than the threshold are cleared. This might clear weak information of the filament. As shown in the red box of Figure 8(b), the weak part of the filament is directly removed. Figure 8(c) is the denoised image after MCA decomposition. Structures of the filament are more complete than those in panel b. Setting different threshold of noise variance for the two parts can avoid the removal of useful information, especially in the area marked by red box. Our method not only removes noise from the cartoon layer and texture layer but also strengthens details in the image synthesis process. Applying MCA to extract filament can retain much structural information of the filament.

3.2.2. Extraction Results. Finally, the layered subimages are added together to produce the filament. Figure 9 shows the extraction results obtained using the getfilaments algorithm. Figure 9(a) is the extracted filament done by getfilaments. Compared to the input filament in Figure 3(a), most of noises are cleaned and structures of the filament are also removed (marked in red boxes in Figures 9(b) and 9(c)). The extraction results obtained using our method are shown in Figure 10. Comparing with the getfilaments method, our method obtains the filament more complete, excludes noise, and retains more structural information, especially in the three areas marked by red boxes (Figures 10(b) and 10(c)).

3.3. Peak Signal-to-Noise Ratio. We compare the peak signal-to-noise ratio (PSNR) of images processed by the getfilaments

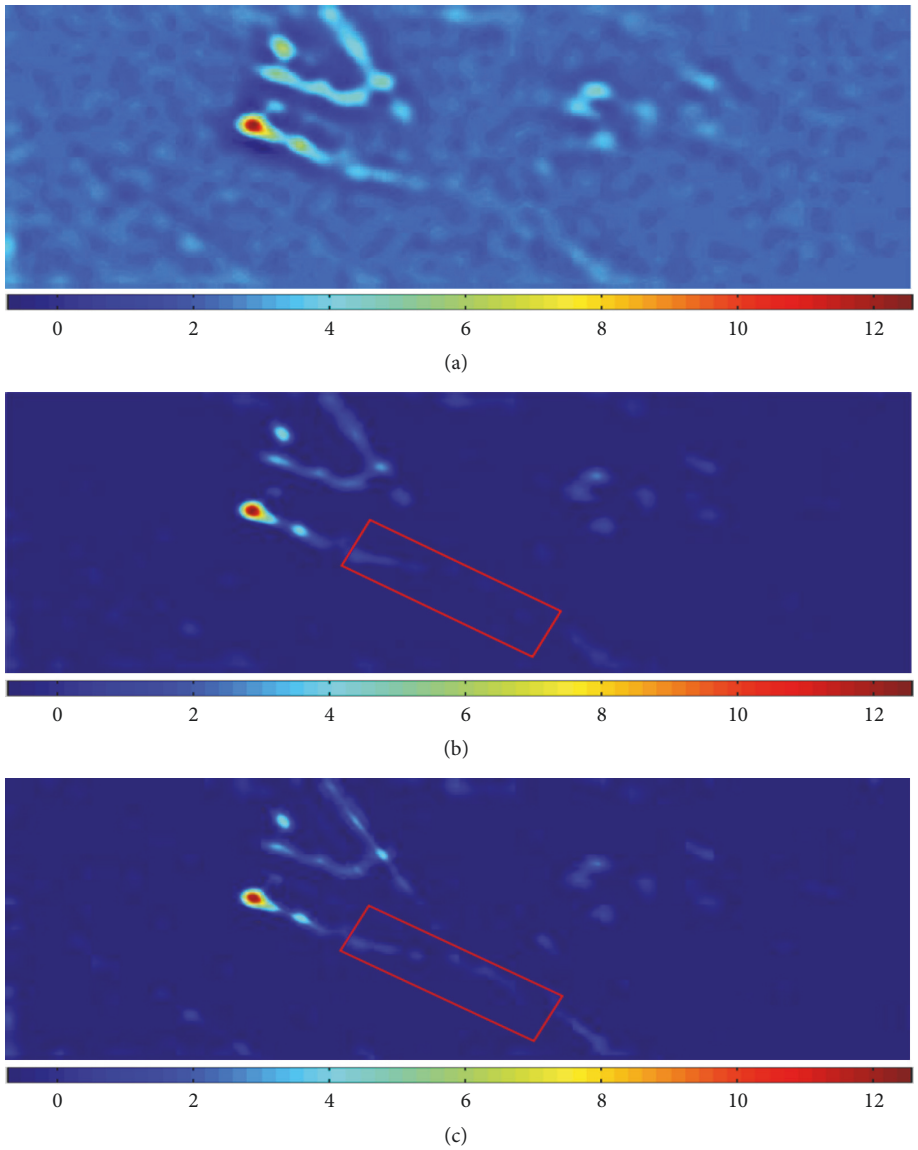


Figure 8: Denoising results of the 40th subimage. (a) The original 40th subimage. (b) The 40th subimage obtained using the iterative cleaning algorithm directly. (c) The denoised 40th subimage after MCA decomposition and fusion.

algorithm with that of our method. The PSNR is the objective criterion most widely used to evaluate image quality. An image has less noise when the PSNR is higher. The PSNR is defined as follows:

$$PSNR = 10 * \log_{10} \frac{(2^n - 1)^2}{MSE}. \quad (7)$$

where MSE is the mean square error (i.e., difference) between the original image and the image after noise is superimposed and n is 8 since pixels are represented using 8 bits per sample.

For different intensities of salt-and-pepper noise, we analyze the PSNRs of images processed by different methods. Table 1 shows that the PSNR of the images processed using our improved method is always higher than that of getfilaments method, which means that the images processed by MCA have less noise.

Table 1: PSNR comparison of images processed using different methods.

Noise intensity	Images processed by different methods		
	Original	Getfilaments	Our method
0.1	18.2545	18.3443	18.5836
0.15	15.4864	16.5203	16.8379
0.2	14.9552	15.3325	15.7445
0.3	13.2314	13.7914	13.8200
0.5	11.2925	11.3277	11.8972

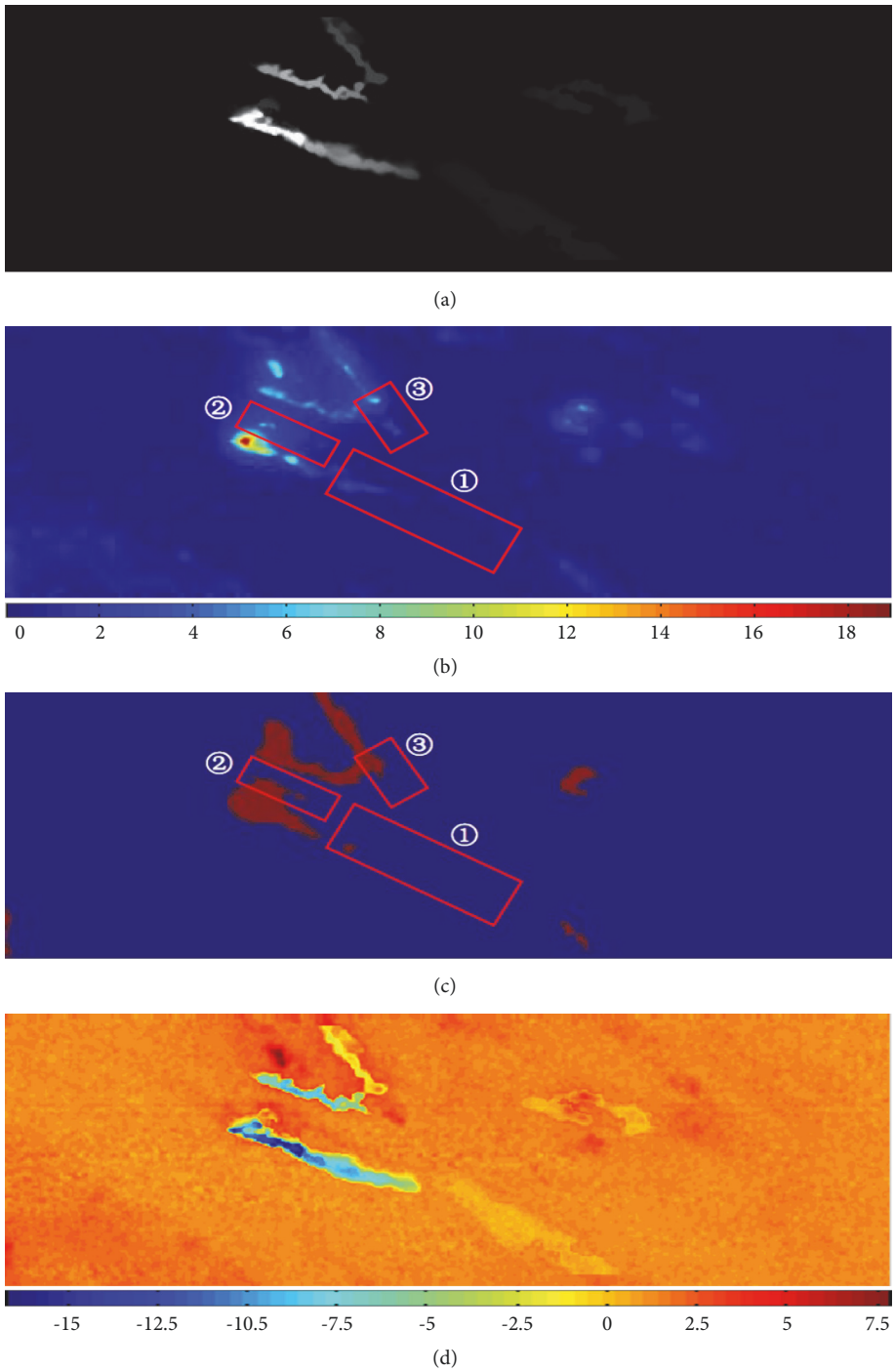


Figure 9: Extraction results using the getfilaments algorithm. (a) Extracted filament. Part of information is removed, including noises and structures. (b) Extracted filament with colors for contrast. Structures of the filament are cleaned in three marked places. (c) Contour extraction of the filament. (d) Residuals after the subtraction of the extracted filament from the original input image.

Data Availability

The data suffixed with fits used to support the findings of this study were supplied by National Astronomical Observatory of China under license and so cannot be made freely available.

Disclosure

F. Q. Duan present address is College of Information Science and Technology, Beijing Normal University, Beijing, China. The authors presented this work in 2016 Astronomical Data Analysis Systems and Software Conference.

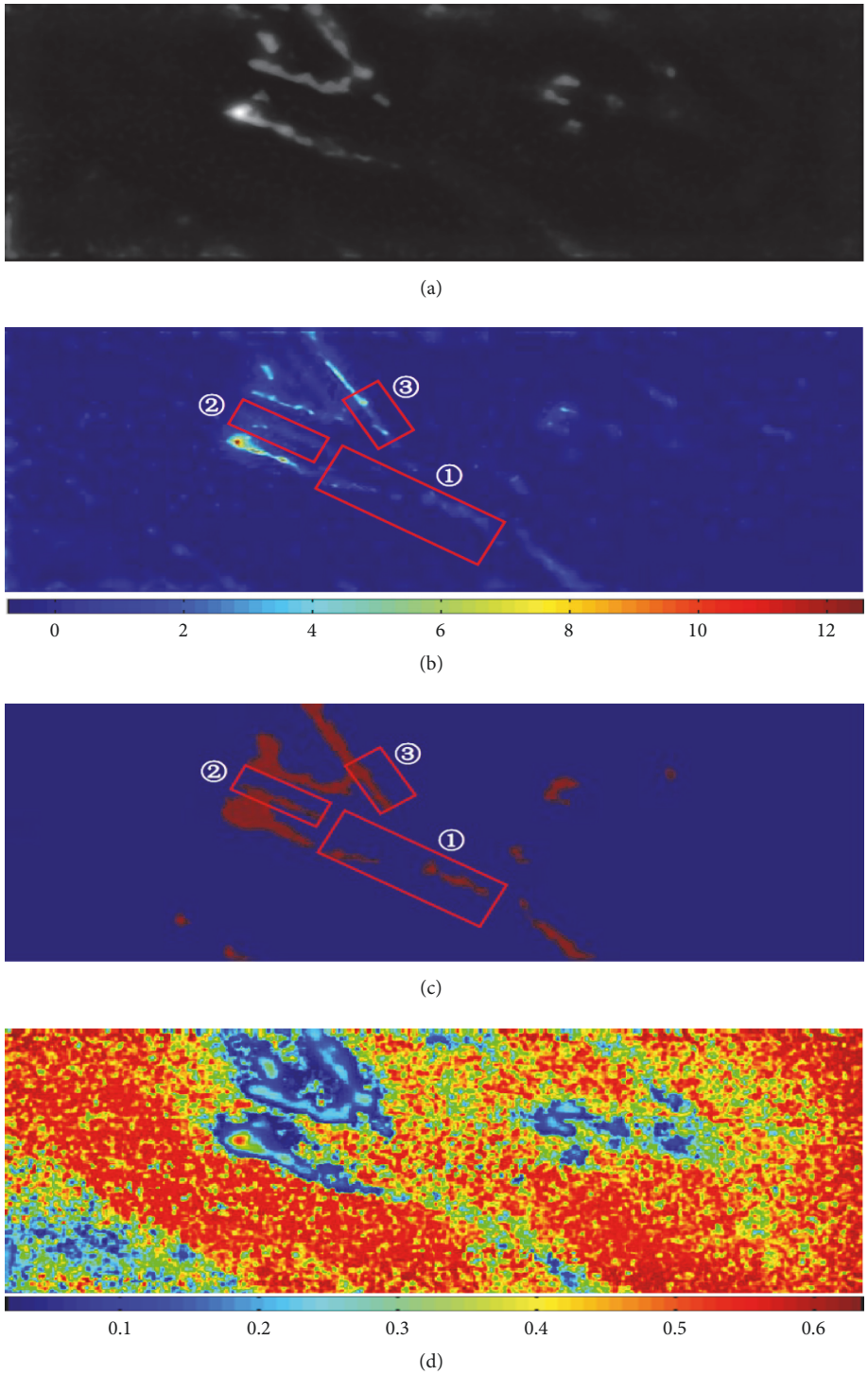


Figure 10: Extraction results obtained using our method. (a) Un-colored extracted filament. Noises are removed, and filament's structures are retained. (b) Colored extracted filament. Compared to the filament in Figure 9(b), filament's structures are more complete, especially in three marked places. (c) Contour extraction. (d) Residuals after the subtraction of the extracted filament from the original image.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study is supported by the National Natural Science Foundation of China (no. 41272359), Ministry of Land and

Resources for the Public Welfare Industry Research Projects (201511079-02), and Ph.D. Programs Foundation of Ministry of Education of China (20120003110032).

References

[1] A. Men'shchikov, P. André, P. Didelon et al., "Filamentary structures and compact objects in the aquila and polaris clouds observed by herschel," *Astronomy and Astrophysics*, vol. 518, article L103, 7 pages, 2010.

[2] G. C. Gómez and E. Vazquez-Semadeni, "Filaments in simulations of molecular cloud formation," *The Astrophysical Journal*, vol. 791, no. 2, article 6298, 2014.

[3] P. Padoan and Å. Nordlund, "The stellar initial mass function from turbulent fragmentation," *The Astrophysical Journal*, vol. 576, no. 2, pp. 870–879, 2002.

[4] P. André, A. Men'shchikov, S. Bontemps et al., "From filamentary clouds to prestellar cores to the stellar IMF: initial highlights from the herschel gould belt survey," *Astronomy & Astrophysics*, vol. 518, article L102, 7 pages, 2010.

[5] S. Molinari, B. Swinyard, J. Bally et al., "Clouds, filaments, and protostars: the herschel Hi-GAL milky way," *Astronomy and Astrophysics*, vol. 518, Article ID L100, 5 pages, 2010.

[6] D. Kereš, N. Katz, D. H. Weinberg, and R. Davé, "How do galaxies get their gas?" *Monthly Notices of the Royal Astronomical Society*, vol. 363, no. 1, pp. 2–28, 2005.

[7] D. Kereš, N. Katz, M. Fardal, R. Davé, and D. H. Weinberg, "Galaxies in a simulated ΛcDM universe - I. Cold mode and hot cores," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 1, pp. 160–179, 2009.

[8] S. F. Shandarin and Y. B. Zeldovich, "The large-scale structure of the universe: turbulence, intermittency, structures in a self-gravitating medium," *Reviews of Modern Physics*, vol. 61, no. 2, pp. 185–220, 1989.

[9] C. F. McKee and L. L. Cowie, "The interaction between the blast wave of a supernova remnant and interstellar clouds," *The Astrophysical Journal*, vol. 195, pp. 715–725, 1975.

[10] S. Casassus, G. van der Plas, M. Sebastian Perez et al., "Flows of gas through a protoplanetary gap," *Nature*, vol. 493, pp. 191–194, 2013.

[11] N. Schneider, T. Csengeri, M. Hennemann et al., "Cluster-formation in the Rosette molecular cloud at the junctions of filaments (Corrigendum)," *Astronomy and Astrophysics*, vol. 551, article C1, 1 page, 2013.

[12] P. Hennebelle, R. Banerjee, E. Vázquez-Semadeni, R. S. Klessen, and E. Audit, "From the warm magnetized atomic medium to molecular clouds," *Astronomy & Astrophysics*, vol. 486, no. 3, pp. L43–L46, 2008.

[13] A. V. Tugay, "Extragalactic filament detection with a layer smoothing method," *Physics*, vol. 1, 2014, https://arxiv.org/abs/1410.2971.

[14] A. Men'shchikov, "A multi-scale filament extraction method: getfilaments," *Astronomy and Astrophysics*, vol. 560, article A63, 15 pages, 2013.

[15] N. L. J. Cox, D. Arzoumanian, Ph. André et al., "Filamentary structure and magnetic field orientation in Musca," *Astronomy & Astrophysics*, vol. 590, Article A110, 8 pages, 2016.

[16] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.

[17] J. L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[18] S. Velasco-Forero and J. Angulo, "Classification of hyperspectral images by tensor modeling and additive morphological decomposition," *Pattern Recognition*, vol. 46, no. 2, pp. 566–577, 2013.

[19] L. Yan, X. Xiaohua, and L. Jian-Huang, "Face hallucination based on morphological component analysis," *Signal Processing*, vol. 93, no. 2, pp. 445–458, 2013.

[20] Z. Xue, J. Li, L. Cheng, and P. Du, "Spectral–spatial classification of hyperspectral data via morphological component analysis-based image separation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 70–84, 2015.

[21] D. Szolgay and T. Sziranyi, "Adaptive image decomposition into cartoon and texture parts optimized by the orthogonality criterion," *IEEE Transactions on Image Processing*, pp. 3405–3415, 2012.

[22] A. Buades, T. Le, J.-M. Morel, and L. Vese, "Fast Cartoon + Texture Image Filters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 19, no. 8, 2010.

[23] C. Yu, Q. Qiu, Y. Zhao, and X. Chen, "Satellite image classification using morphological component analysis of texture and cartoon layers," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1109–1113, 2013.

[24] X. Xiang, L. Jun, H. Xin, M. Dalla Mura, and P. Antonio, "Multiple morphological component analysis based decomposition for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 3083–3102, 2016.

[25] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[26] L. L. Jiang, H. Q. Yin, and X. C. Feng, "Adaptive variational models for image decomposition combining staircase reduction and texture extraction," *Journal of Systems Engineering and Electronics*, vol. 20, no. 2, pp. 254–259, 2009.

[27] T. F. Chan, S. Esedoglu, and F. E. Park, "Image decomposition combining staircase reduction and texture extraction," *Journal of Visual Communication and Image Representation*, vol. 18, no. 6, pp. 464–486, 2007.

[28] E. J. Cand and D. L. Donoho, "Curvelets, multiresolution representation, and scaling laws," 1, 2000.

[29] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[30] L. Vincent and P. Soille, "Morphological segmentation of binary patterns," *Pattern Recognition Letters*, 1991.

[31] N. Laurent and S. Michel, "Geodesic saliency of watershed contours and hierarchical segmentation," vol 18, 1163, 1996.

[32] F. Meyer, "Watersheds on weighted graphs," *Pattern Recognition Letters*, vol. 47, pp. 72–79, 2014.

[33] F. Malmberg and C. L. L. Hendriks, "An efficient algorithm for exact evaluation of stochastic watersheds," *Pattern Recognition Letters*, vol. 47, pp. 80–84, 2014.

[34] A. Men'shchikov, P. André, P. Didelon et al., "A multi-scale, multi-wavelength source extraction method: getsources," *Astronomy and Astrophysics*, vol. 542, article A81, 31 pages, 2012.

Hindawi
Advances in Astronomy
Volume 2019, Article ID 8087405, 5 pages
https://doi.org/10.1155/2019/8087405

Research Article

A New Image Restoration Method for MUSER

Wei Wang [1,2] **and Yihua Yan** [1,2,3]

[1] *National Astronomical Observatories, Chinese Academy of Sciences, China*
[2] *Key Laboratory of Solar Activity of Chinese Academy of Sciences, China*
[3] *University of Chinese Academy of Sciences, China*

Correspondence should be addressed to Wei Wang; wwang@nao.cas.cn

Received 18 January 2019; Accepted 8 April 2019; Published 2 May 2019

Guest Editor: Junhui Fan

Solar radio images in decimeter wave range consist of many complicated components including a disk component, some bright and weak compact sources, and many diffuse features. Complicated structures combining these various components maybe cause restoration failure when using conventional algorithms. Furthermore, the images at different frequencies band are pretty different. Therefore, restoration method for solar radio image is different from other radio sources. Some image restoration methods were applied and obtained good results on Nancay radioheliograph images and Nobeyama radioheliograph images, and some new methods were introduced into processing these complicated solar radio images in recent years. For a new radioheliograph with ultrawide frequency band, new image restoration method which can maximize function of telescope is demanded. Different images could be obtained from the same visibilities data by using different weighting functions in imaging processing. In this paper, a new restoration method for solar radio image was proposed. Two images with different weighting functions from the same data are combined in this method. This restoration method has applied to data processing of Mingantu spectral radioheliograph.

1. Introduction

Mingantu spectral radioheliograph (MUSER) is a solar dedicated radio interferometric array that can produce high temporal resolution, high spectral resolution, and high spatial resolution radio images at decimeter to centimeter wave range [1]. Observation in routine has been carried out for 3 years since 2016; a mass of observational data and some image results have been obtained. The solar radio images are very different from other radio astronomical images. This is because there are a lot of different kinds of structures in solar radio images, such as a solar disk component; some bright compact sources like flares; some faint compact sources like radio emissions associated with X-ray bright points and X-ray jets; many diffuse features like active regions, depressions associated with dark filaments, and plateau-shaped enhancements in polar and coronal-hole regions [2]. Such complicated structures could cause restoration failure and increased computational time when using conventional algorithms. Using standard CLEAN algorithm and its evolutional algorithm cannot obtain good results for MUSER images with ultrawide frequency band.

Some methods developed for other radioheliographs in the world give us some experiences to solve this problem. A new improved multiscale CLEAN algorithm was applied to Nancay radioheliograph (NRH) data. This algorithm restores solar image with a wide spectrum of spatial scales by using 5 or 6 uv-filters including a central disk and 4 or 5 rings, and it is less time-consuming than the ordinary CLEAN procedure [3]. About Nobeyama radioheliograph (NoRH), there are 3 methods to process image; the first of the three used standard CLEAN algorithms, optimized to flares using dirty models, which mean the dirty images of model Gaussian sources with various sizes [4]. The Steer algorithm was the second method, which is better for diffuse radio sources [2]. The third method is to attain the expected resolution in partial images and introduces two-step CLEANs [5].

According to our knowledge of solar physics, we know that solar radio image at different frequency band corresponds to positions in different height from solar surface.

That means solar radio images at low frequency band are much different than ones at higher frequency band. The frequency of MUSER is 0.4GHz to 15 GHz, which is much wider than NRH and NoRH. It is hard to process these data at wide frequency band with the same method. Therefore, methods for NoRH and NRH could not be applied directly to process MUSER data. Therefore, a new image restoration method is proposed to process MUSER data in wide frequency band in this paper.

2. Restoration Method

Usually the weighting function is specified in many Fourier transform imaging programs to fine tune the beam shape [6]. The beam shape can be controlled or be extended by the weighting function. The beam shape means spatial resolution. The image with different spatial resolution can be processed from the same visibilities by using different weighting functions. The main idea of this restoration method is combination of two images which applied different weighting functions and different clean procedures.

2.1. The Weighting Functions. The points in u-v coverage are always insufficient for an interferometric array, especially for solar observation in snapshot mode. The radio image of the sun is changing with time, it is impossible to combine the visibilities with longer integration time to improve quality of image. The integration time for solar observation is only a couple of milliseconds. The distribution of points in u-v plane is not uniform. The density weighting function can be used to compensate for the clumping of data in u-v plane by weighting by the reciprocal of the local density [6]. There are two choices for the weighting function in a general way, natural weighting and uniform weighting.

For natural weighting, all of points in u-v plane were treated equally. For solar observation, a large member of short baselines are necessary for achieving the solar disk and large-scale structure well. The number of MUSER antennas within 100 meters is about 40 percent. Therefore, natural weighting emphasizes the data from the short baselines. The shape of beam intends to be extended, and a relative wide beam size will occur in the image with natural weighting. But the advantage of extended beam size is to get much better signal-to-noise ratio for detecting the weak sources, such as dark filaments and coronal hole.

For uniform weighting, points within denser areas in u-v plane have lower weights than others. The density distribution of central area is dense and the surrounding is sparse gradually. That means the weight of data from short baseline is lower than from long baseline. It makes against observing large-scale structure of solar image. Compared to natural weighting, the beam size with uniform weighting is less than with natural weighting and is better for obtaining high spatial resolution for the compact sources in the sun, such as flare and X-ray jets.

Sometimes, these two weighting functions are insufficient especially in the snapshot mode of observing. For super uniform weighting, robust weighting was developed to be response to these problems [6]. In this image restoration method, only natural weighting and uniform weighting have been taken into account.

2.2. The Deconvolution. The visibilities output from correlator of an interferometry can be used to produce an image of distribution of sky brightness. But the quality of image is insufficient due to the loss of visibilities caused by limited number of antennas. So, the deconvolution is necessary for radio images observed by interferometric array. The CLEAN algorithm and some derivative algorithm are the predominant algorithm used for deconvolution of radio images. In this restoration method, standard CLEAN algorithm and Steer algorithm are combined for MUSER image.

The CLEAN algorithm was devised by J. Hogbom in 1974 [7]; it intends to create a model of the true "sky-brightness distribution" from the "dirty image" which is the observed image using the "dirty beam". The model is called the "clean image" and is created using the "clean beam," which represents an ideal PSF. It is typically a Gaussian function of the same FWHM as the central component of the dirty beam [8]. The CLEAN algorithm can get better results for point sources or point-like sources. But the extended sources sometimes could not be recovered because of lacking the extremely short or zero spacings. For solar observation, solar disk component always existed in images. So standard CLEAN procedure is insufficient to process the data of MUSER.

The Steer algorithm is proposed by Steer et al. in 1984 [9]; it was applied to the data processing in daily solar images in the automatic data processing routine of the Nobeyama radioheliograph [2]. The Steer algorithm deals with radio sources in a group rather than individually. Components brighter than a certain fraction of the peak brightness in the dirty image are regarded as a group, called a contour-trim; the contour-trim is convolved with the dirty beam to create a dirty contour-trim, which is extracted from the dirty image [9]. The Steer algorithm needs less computational time than standard CLEAN algorithm due to group subtraction other than point subtraction.

2.3. Combination of Weighting and Deconvolution. According to analysis of different weighting functions and deconvolution algorithm, uniform weighting and standard CLEAN algorithm intended to process point sources and point-like sources, and natural weighting and Steer algorithm intended to deal with extended sources. There are various structures including a disk, compact sources, and diffuse sources in solar radio images as mentioned in the Introduction of this paper. Therefore, the main idea is combination of different weighting functions and deconvolution algorithms; a new synthetic restoration method is proposed and proceeds as follows.

(1) Remove partial solar disk component

(2) Obtain the first dirty image from the data with uniform weighting

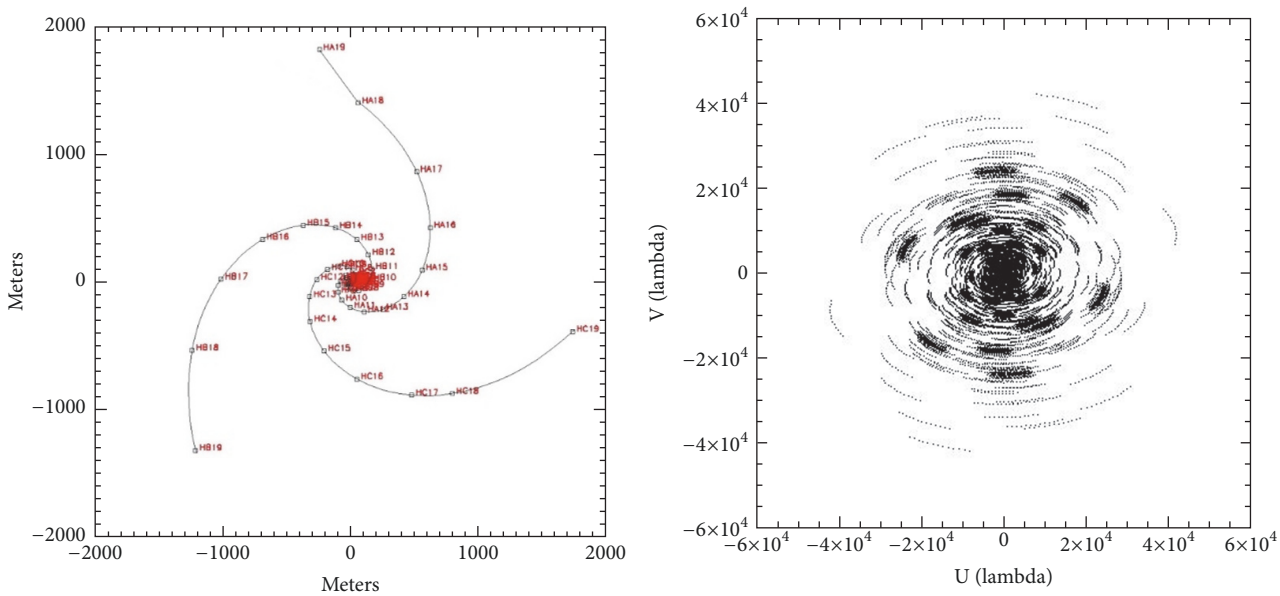


Figure 1: (Left) positions of antennas and (right) u-v coverage.

(3) A Hogbom CLEAN algorithm is performed on the first dirty image; get a clean image

(4) Obtain the second dirty image from the data with natural weighting

(5) A Steer algorithm is performed on the second dirty image; get another clean image

(6) Add the two clean images together; add the residuals of these dirty images to the clean image

(7) Add back the solar disk component and get the final image

A solar disk component created by convolution of a model of solar disk and the dirty beam was subtracted from the visibilities data before applying this method and added back in clean images finally. It must be noted that the component subtracted is not all solar disk and is just part of solar disk. The reason of this is more easily detecting the weak sources in the solar disk than the solar disk removed completely. Therefore, some bright and faint compact sources, many diffuse structures, and partial disk were left for restoration. Besides that the peak values of the two clean images should be scaled to the same before adding together.

3. Result of MUSER Observation

This new restoration method was performed on observational data of MUSER. The data on July 5, 2016, was processed with this restoration method. The frequency of this observation is 4.2 GHz. The sun is very quiet without any bursts from flux density observed by Nobeyama radio polarimeters and radio spectrum of MUSER. To improve quality of image, the 2-hour data with 10-minute cadence were merged together. The integration time of each point in u-v plane is 60 milliseconds. The antenna configuration and u-v coverage of this data are shown in Figure 1. The dirty image with natural weighting function and clean image using Steer algorithm are shown in Figure 2. The dirty image with uniform weighting function and clean image using Hogbom CLEAN algorithm are shown in Figure 3. The final image using this restoration method and image observed with NoRH are shown in Figure 4.

From the final image using this restoration method applied to MUSER observational data, the quality of this image is better than only using one deconvolution algorithm, similar result to NoRH. There are differences between images of MUSER and NoRH due to different observational frequency.

4. Discussion

Only natural weighting and uniform weighting are performed in this restoration method in this paper. Solar radio images at different frequency are different. Also, the solar image in a different phase of the solar activity cycle, such as solar minimum and maximum, is much different. Therefore, some other weighting functions, such as robust weighting, perhaps need be considered on the data with other frequency bands in the future. In addition, the cleaned image by different weighting function is added together with equal weighting. Adjusting different weights for different cleaned images makes this restoration method more general for processing different solar images.

5. Conclusion

MUSER is a solar dedicated interferometry which is called new generation solar radio telescope. It covered decimeter to centimeter wave range. Some new restoration methods or algorithms are needed to meet the requirement of the

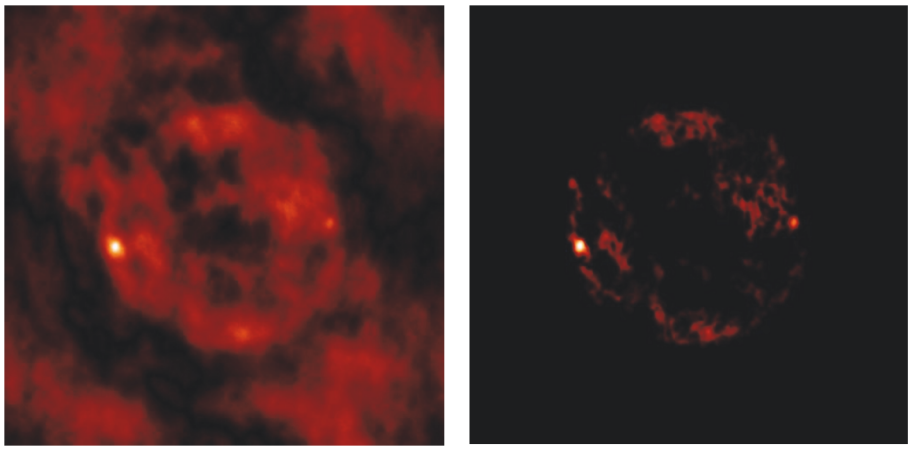


Figure 2: (Left) dirty image with natural weighting and (right) clean image using Steer algorithm.

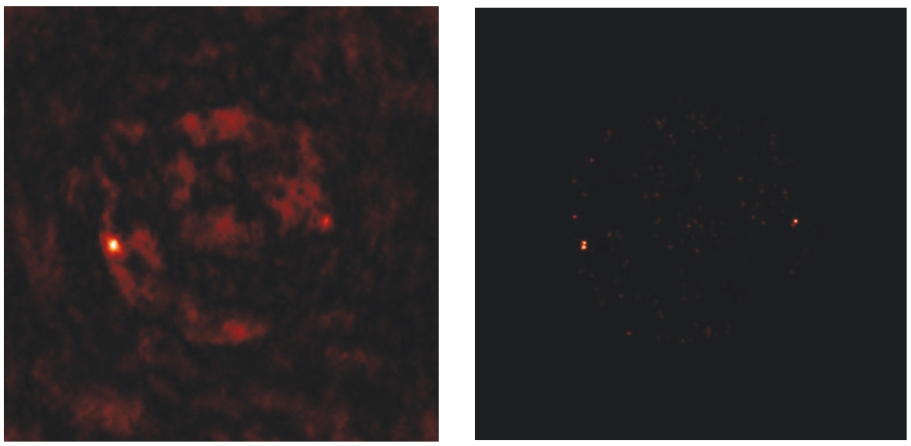


Figure 3: (Left) dirty image with uniform weighting and (right) clean image using Hogbom CLEAN.

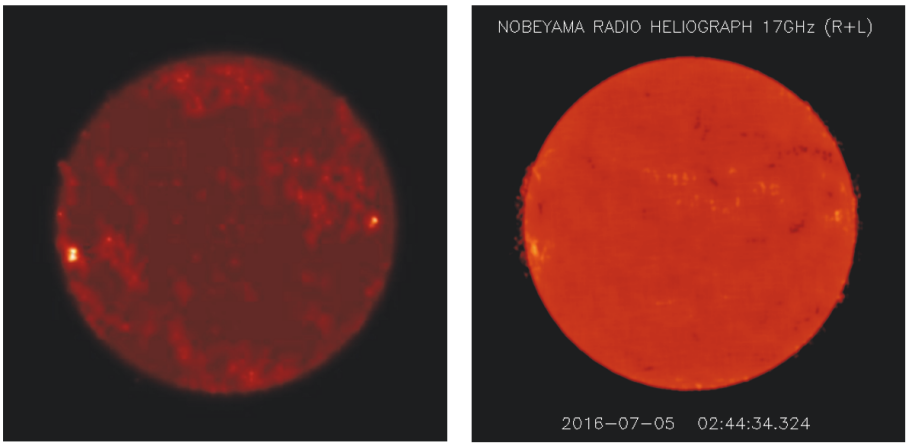


Figure 4: (Left) the final image of MUSER and (right) image of Nobeyama radioheliograph.

ultrawide frequency band and complicated structures of solar image. In this paper, a new restoration method is proposed and applied to the data of MUSER, obtaining some results.

Data Availability

The data of MUSER and processing manuscripts are open access for everyone concerned.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Prof. Dale Gary gratefully acknowledges comments and suggestions to improve the restoration method. This work is supported by National Natural Science Foundation of China, grants 11433006, 11790301, 11790305, and 11773043, and also supported by open research project of NAOC-Ali Astronomical Big Data Joint Research Center.

References

[1] W. Wang, Y. Yan, D. Liu et al., "Calibration and data processing for a chinese spectral radioheliograph in the decimeter wave range," *Publications of the Astronomical Society of Japan*, vol. 65, S18, 2013.

[2] H. Koshiishi, "Restoration of solar images by the Steer algorithm," *Astronomy & Astrophysics* , vol. 412, no. 3, pp. 893–896, 2003.

[3] C. Mercier, "High dynamic solar radio images by combining visibilities from the Giant Meterwave Radio Telescope (GMRT) and the Nancay Radioheliograph (NRH)," in *Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, vol. P75, 2006.

[4] Y. Hanaoka, "Processing of the Nobeyama Radioheliograph Data," in *Proceedings of Kofu Symposium*, pp. 35–43, Kofu, Japan, 1993.

[5] K. Fujiki, Ph.D. Dissertation, the Graduate University for Advanced Studies, 1998.

[6] R. Perley, F. Schwab, and A. Bridle, "Synthesis imaging in radio astronomy," *Astronomical Society of Pacific Conferences Series*, vol. Volume 6, 1999.

[7] J. A. Hogbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy & Astrophysics*, vol. 15, p. 417, 1974.

[8] U. J. Schwarz, "Mathematical-statistical Description of the Iterative Beam Removing Technique (Method CLEAN)," *Astronomy & Astrophysics*, vol. 65, p. 345, 1978.

[9] D. G. Steer, P. E. Dewdney, and M. R. Ito, "Enhancements to the deconvolution algorithm 'CLEAN'," *Astronomy & Astrophysics* , vol. 137, no. 2, pp. 159–165, 1984.

Hindawi
Advances in Astronomy
Volume 2019, Article ID 8741027, 12 pages
https://doi.org/10.1155/2019/8741027

Research Article

Design and Implementation of Xinjiang Astronomical Observatory Astronomical Data Transmission Visualization System

Xinchen Ye,[1] Hailong Zhang,[1,2] Yan Zhu,[1,3] Jie Wang,[1] Tohtonur Ergesh,[1] and Huijuan Li[4]

[1] *Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011, China*
[2] *Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, China*
[3] *University of Chinese Academy of Sciences, Beijing 100049, China*
[4] *Xinjiang University of Finance and Economics, Urumqi 830012, China*

Correspondence should be addressed to Hailong Zhang; zhanghailong@xao.ac.cn

Received 1 January 2019; Accepted 19 March 2019; Published 2 April 2019

Guest Editor: Junhui Fan

With the development of astronomical observation technology, astronomical devices produce more data than ever. Astronomical telescopes are usually far away from city, so the long-distance data transmission between telescope and data center faces great challenges. Visualization system of astronomical data transmission with four-layer structure was built to manage data transmission. This visualization system has a four-layer structure: hardware layer, system layer, middle layer, and visualization layer. System function includes automatic data transmission, log recording of transmission process, and display of the transmission status in dynamic web pages. Besides, the middle layer contains an alarm subsystem that can automatically send system exceptions to administrator. We also design corresponding mechanisms to ensure the high stability of the system and to control the data transmission when the network is unstable through adaptive algorithms. In test, this visualization system can run stably for a long time in unmanned manner. This system also provides a solution for the astronomical observation bases to automatically transmit data to the data center.

1. Introduction

With the development of astronomical observational technology, the data quality of telescope receiving equipment is improved. In the meantime, the volume of data generated by telescopes is increasing exponentially [1, 2]. For example, the world's largest fully steerable radio telescope GBT (Robert C. Byrd Green Bank Telescope)[3] generates more than 1.4PB data per year (http://data.xao.ac.cn/static/GBTArchiveProcess.pdf). The world's largest radio telescope FAST (five-hundred-meter aperture spherical radio telescope)[4]'s 19-beam receiver[5] produces 8 bit$\times 10^4 \times 2 \times 4 \times 19$ data per second; more than 10PB of data will be stored per year. After SKA (Square Kilometre Array) [6, 7] is built, it is going to produce 1PB data per day [8].

Due to the specificity of astronomical observations, observatory sites are usually far from data centers. Data needs to be transmitted over a leased line from observatory site to data center owing to the instability in transmission in outdoor data lines over a long distance. Astronomical data transmission requires a complete management system [9] which meets the following conditions: complete logging, user-friendly visual interface for administrators to control data transfer process, high stability to guarantee system running in unattended state for a long time, and automatic sending of alert email to administrator when data transfer process fails.

NGAS (Next Generation Archive System) [10, 11] is the most commonly used archiving software in the field of radio

astronomy. NGAS is for astronomical data archiving, processing, searching, and synchronization. Nowadays, NGAS is used in data archiving of multiple telescopes. MWA (Murchison Widefield Array) [12] is a precursor for SKA witch uses NGAS to synchronize data from Massachusetts Institute of Technology and Victoria University of Wellington. ALMA (Atacama Large Millimeter/submillimeter Array) [13] also uses NGAS for data synchronization [14, 15].

NGAS is already a relatively complete astronomical data archiving system. However, as NGAS is a software developed more than 10 years ago, there are also some problems [16].

(1) NGAS uses HTTP-based methods to transmit data. It was uncertain whether the existing NGAS architecture scales up to cope with a larger amount of data.

(2) Sometimes the dataflow may saturate the transmission bandwidth, and NGAS lacks an effective mechanism to solve this problem.

(3) Users cannot intuitively understand the status of data transmission through NGAS.

This paper designs and develops an astronomical data visualization transmission system, which is based on the actual requirement of data transmission of Xinjiang Astronomical Observatory (XAO) of Chinese Academy of Sciences (CAS). This system contains functions including astronomical data transmission control, log recording during transmission, autoalarming, and visual interface. It is able to efficiently help administrators to control data transmission and it can run steadily for a long time unattended. The total transmission will be recorded in detail for later troubleshooting. The visual interface can show situation of data transmission intuitively. The adopted modular developing technique will make it easier in later transplanting to central controlling systems or large screen display.

2. System Architecture Design

XAO's Nanshan 26m Radio Telescope (NSRT) [17] is about 100km away from the data center of XAO; observation data need to be sent to the data center through a dedicated line every day. At present, there is no systematic management system for data transmission. The 110-meter radio telescope [18] to be built by XAO, Qitai, Xinjiang, will be the world's largest fully steerable radio telescope, and its data transmission line will exceed 200 km. Its data transmission process will be displayed in the large screen system in the future.

The system architecture was designed based on the actual needs of XAO. Astronomical data visualization transmission system adopts a four-layer architecture. The four layers are hardware layer, system layer, middle layer, and visualization layer. The system architecture diagram is shown in Figure 1.

(1) The hardware layer provides a hardware environment for data transfer. The system design and development described in this paper are based on a test hardware environment.

(2) The system layer includes a log subsystem and a data transfer subsystem. The log subsystem is used to record the log of transfer processes and provides management program for administrator. The core of the data transfer subsystem is the rsync transport framework. The data transfer encapsulates the shell commands to call the rsync command.

(3) The middle layer is mainly composed of control programs. These programs are responsible for controlling the subsystems of the system layer and managing the log files and database. The middle layer is also responsible for receiving instructions from and transferring data to the visualization layer. When the transmission process is abnormal, the alarm program will automatically respond and send an alert message to the administrator.

(4) The visualization layer is developed based on web technology and the data transmission situation is visually displayed by visual charts. The system administrator can quickly grasp the data transmission situation information and quickly solve the problem.

The four-layer architecture adopted by the system meets the construction needs. In the development process, there are some problems in the original architecture design. In this paper, we modify the original architecture to get rid of these problems. The layered architecture design is convenient for development and management of this system. Problems in the system testing can be layered. At the same time, such layered architecture is easy for system reuse or porting in future.

3. System Function Realization

3.1. Hardware Layer Test Environment. We used three servers for building the hardware environment. The servers are interconnected through a Gigabit switch. The servers for data sending and receiving are both HP P4300 G2 data server with 2 Intel E5520 CPU, 20 GB RAM, and 6.4TB hard disk. Control server uses the DELL PowerEdge R710 with 2 Intel Xeon 5680 CPU, 32GB RAM, and 3.6TB hard disk.

Because the control server load is low, it is recommended to configure the control program on the nondedicated server to reduce the cost of equipment and energy in a real environment.

3.2. System Layer

3.2.1. Log Subsystem. The log subsystem includes log collection, log storage, log management, and management program. The log subsystem is an independent development module and it has a complete set of data processing flow. So it can be split and used separately. The log subsystem structure diagram is shown in Figure 2.

The log content stored in the database is designed mainly for the convenience of visualization layer invoking; it contains 6 database tables.

(1) File table (files): it is used to record the specific information of each file.

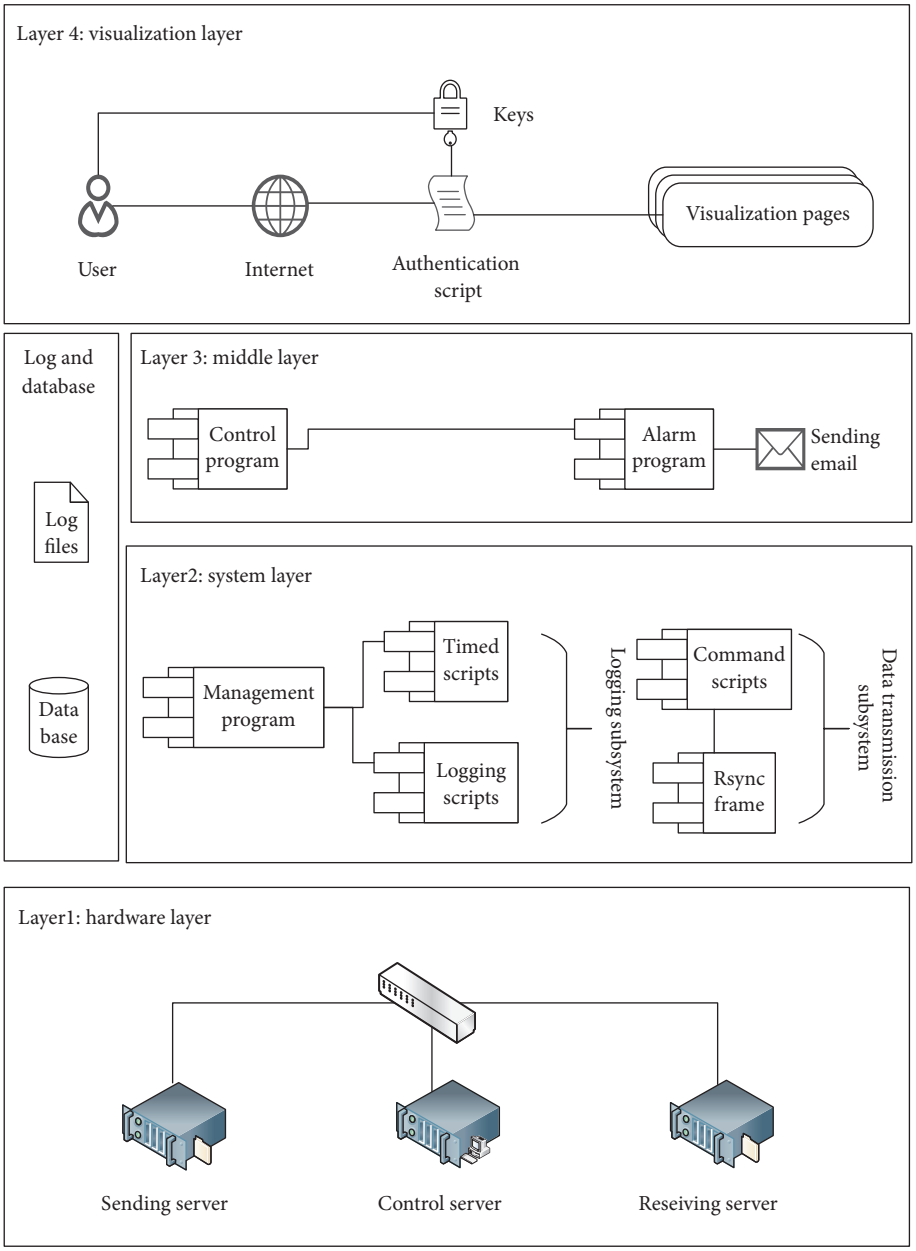


Figure 1: System architecture diagram.

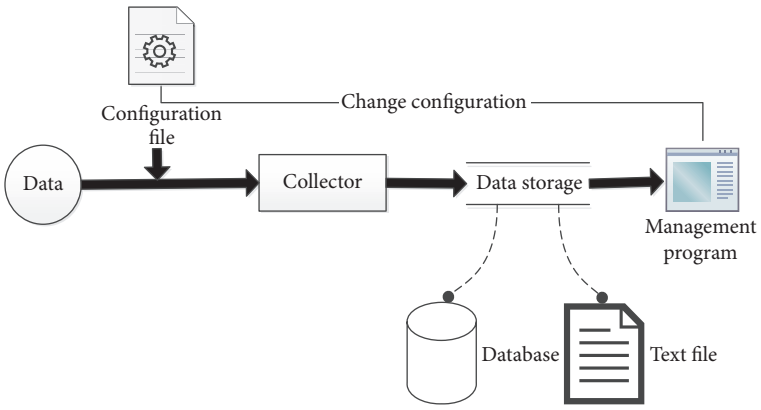


Figure 2: Log subsystem structure diagram.

(2) Astronomical data table (data): it is for recording data storage information.

(3) Folder table (folder): it is for recording information of subfolders of root directory.

(4) Daily data delta table (dayData): it is for recording information of daily data increments.

(5) Daily folder data delta table (dayFolderdata): it is for recording daily data increments of subfolders in root directory.

(6) Scripts monitoring table (proc_tatus): it is used to record the running of the scripts.

The information about specific field is shown in Table 1.

In the traditional log management system, the administrator's manipulation on the log file is usually performed in the form of command line, which is not convenient and not intuitive. Log query and control management interface developed in Qt creator [19] using Qt language facilitate the management of log system. Its functional structure diagram is shown in Figure 3. Through the management interface, log retrieval within a specified time range and various log queries can be realized; one-click backup of log files for a specified date range (3 months, half a year, one year) is also enabled. The log retrieval interface is shown in Figure 4.

3.2.2. Data Transmission Subsystem. The core of the logging subsystem is the remote synchronization tool rsync (Remote Sync). rsync is a mature mirror backup tool for Linux. It is used as a basic framework in a variety of data synchronization software [20]. Its main features are as follows.

(1) Rsync can mirror the entire directory and file system, and its transfer process can maintain the permissions, time, soft connection, and other information of the original file.

(2) Rsync supports incremental backup which can compress and decompress data in real time during transmission, so the transmission rate is faster. Besides, rsync can run on low bandwidth and high latency communication lines [21].

Rsync uses scp and ssh for data transmission. It will establish a virtual pipeline during transmission to ensure the security of data transmission. The rsync authentication process is shown in Figure 5.

The installation and configuration of rsync are more complicated. In addition to installation of the xinetd and rysnc packages, it also needs to set configuration files and synchronization folder permissions and configure the system firewall. We have packed up the rsync installation packages and the required configuration files for easy installation and use. rsync requires a manual authentication password during the transfer process. Expect tool is used to automate the authentication process; expect is a tool built on tcl to automate processes that require interaction.

Shell scripts are used to make the server use rsync automatically to synchronize the data in the specified folder. Some of the rsync statements are encapsulated in shell scripts, such as running, logging, and transferring. Transfer control can all be performed at the visual layer without having to operate on the command line. Specific package commands are shown in Table 2.

3.3. Middle Layer

3.3.1. Control Program. The control program is responsible for ensuring the normal running of scripts program, receiving commands from visual layer, sending commands to the system layer, and providing filtered log information to the visual layer. The control program is mainly composed of a set of shell scripts. A triangle daemon script architecture is designed to ensure stable running of unattended visual transmission system, as shown in Figure 6.

Two daemon scripts are used to monitor the core control scripts which also monitor each other. Under this architecture, system runs normally unless both daemon scripts and core control scripts are suspended at the same time. Except such situation, any script will be restarted when it fails. In the test, data transfer was sometimes suspended because of the rsync tool exception. A new monitor was developed to check the status of the rsync tool automatically. The rsync tool will be restarted if the monitor program finds an exception. In the last 1000 hours of testing, there was no manual intervention during system running.

Unstable transmission is likely to occur during long-distance data transmission. In this paper, the VSAN algorithm is designed in the control program to prevent the rsync from being repeatedly restarted when the network is not good and to ensure the stable operation of the system when the transmission quality is poor. The core idea of the VSAN algorithm is to transmit data normally when the network is unobstructed. When there are multiple small amounts of data, unified transmission will be sent after accumulating enough amounts of data. When the network delay is too high, the data transmission period will be extended. The flowchart of the VSAN algorithm is shown in Figure 7. Vn is the amount of data to be transmitted, Sd is the standard deviation of the transmission rate in 10 minutes, Ad is the average transmission rate in 10 minutes, and Nd is the transmission delay.

In the control program, the control interface can be used to start, shut down, and restart the system. It can also configure the system log storage directory, the size of a single log file, the log polling mode, and the system scan interval. The control interface is shown in Figure 8.

3.3.2. Alarm Program. The data transmission process will encounter various abnormal conditions. The alarm program periodically analyzes the log files to discover abnormal conditions in the system. Then the alarm program will automatically generate an exception report file and send an email to administrator for timely processing. The alarm program works by periodically analyzing specific log fields on the receiving server and the sending server to determine whether an exception has occurred and automatically writing the exception code value to the specified file on port 80. The control server periodically obtains the code value through the

Table 1: Database table.

Table name	Fieldname	Field type	Field description
files	File_name	char	File name
	File_size	int	File size
	storeDBtime	timestamp	File storage time
	File_time_last	timestamp	Last time the file was modified
	Md5value	char	Checksum
data	data_no	int	Order number
	data_volume	double	Total amount of data
	data_number	int	Total number of files
	data_time	timestamp	The time of the data is written to the database
	data_time_last	timestamp	The time of the data was last modified
	data_add	char	Alternate field
folder	auto_no	int	Order number
	folder_no	int	Folder number
	foldet_cycle_no	int	The time of the folder has been accessed
	folder_name	Char	Folder name
	folder_volume	Double	The size of the folder
	folder_number	Int	Total number of files in the folder
	folder_time	Timestamp	The time of the data was written to the database
	folder_time_last	timestamp	The time of the folder was last modified
	folder_add	char	Alternate field
dayData	auto_no	int	Order number
	dayDate	date	Date of the day
	data_volume	double	Total historical data
	data_number	int	Total number of historical files
	incre_volume	double	Data increase on the day
	incre_number	int	Increase in the number of files on the day
	dayData_add	char	Alternate field
dayFolderData	auto_no	int	Order number
	dayDate	date	Date of the day
	folder_name	char	Folder name
	folder_volume	double	The size of the folder
	folder_number	int	Total number of files in the folder
	incre_volume	double	Data increase of the folder on the day
	incre_number	int	Increase in the number of the folder's files on the day
	dayFolderData_add	char	Alternate field
proc_status	auto_no	int	Order number
	writeData	tinyint	Script writeData
	control	tinyint	Script control
	oneDayStat	tinyint	Script oneDayStat
	rsync	tinyint	Script rsync
	update_time	timestamp	Last updated

"heartbeat" method and sends an email of the corresponding content to the system administrator according to the different code values. The exceptions and exception codes are shown in Table 3.

We assume that the control server is usually located in the data center and it is rare for the network to be abnormal. In addition, the data center usually has its own network situation alarm system. So we use a separate alarm system architecture. The alert is not sent directly from the sending server or the receiving server.

3.4. Visualization Layer. The visualization layer provides web pages developed by HTML5, PHP, JavaScript, jQuery, and other web technologies to display astronomical data transmission. The visualization layer extracts the content from the middle layer and displays it in the web pages in the

Table 2: Package rsync command table.

Packaged command option	Package command function
-a, −−archive	Recursive mode, which means recursively transferring files and keeping all file attributes
-r, −−recursive	Treat subdirectories in recursive mode
-p, −−perms	Keep file permissions
-t, −−times	Keep file time information
-g, −−group	Keep file group information
-z, −−compress	Compress when transferring files
−−progress	Display the transmission process during transmission
-v, −−verbose	Detailed output mode
-q, −−quiet	Streamlined output mode
-b, −−backup	Back up legacy files in the target directory when there are changes
−−ignore-existing	Ignore files that already exist on the receiving server; only those newly created files are backed up

Table 3: Exception code correspondence table.

Exception codes	Exception
110	Everything is normal
111	rsync startup exception
777	Control module exception
555	Network delay is abnormal
444	Storage space is running out

form of charts. The adaptive layout of the web pages enables users to access the web pages through an authorized account using a computer or mobile browser. Here is a test web page http://210.73.36.12/qttas/for-test.php, which is open to all to browse parts of the visualization layer. The following part of this section will show the visualization web pages of data transmission between the Qitai Observatory and the Xinjiang Observatory headquarter.

The web pages are divided into five parts. The first part shows the running state of the scripts obtained from the table 'proce_status' in the database on the sending and receiving servers. Any script that is not running will be intuitively displayed in this part, as shown in Figure 9.

The second part is shown in Figure 10. This part displays the amount of data through the column charts. The column charts can display the volume of data that has been transmitted on the current day and the past 7 days. The second part calls the 'dayData' and 'data' database tables.

The third part is shown in Figure 11. This part shows the storage status of the sending server and the receiving server through pie charts. The third part can help administrators to determine whether data storage needs to be expanded. When the free space of the data server is lower than the threshold, the third part will be displayed in red.

The fourth part is shown in Figure 12. This part displays the data storage in the past 56 days in the form of color blocks. The darker the color block, the more the data produced in this day and the lighter, the less. In order to ensure that the color block display is natural and can truly reflect the amount of data, we first get the data volume sorting in the past 56 days by the bubble sorting method. The maximum data volume is Vmax, the minimum data volume is Vmin, the data volume interval is Vdi=Vmax-Vmin, and the daily data volume is Vday; the color value percentage of daily data is

$$\text{Cper} = \frac{(\text{Vday} - \text{Vmin})}{\text{Vdi}} \tag{1}$$

The fifth part displays the amount of the transferred data and the stored data per minute in the past 2 hours by the broken line charts. As shown in Figure 13, we simulate the real time in which the storage bandwidth is greater than the transmission bandwidth. The broken line charts can visually display the fluctuation of the data transmission rate, and the administrator can judge whether the data link is unblocked by these charts.

In addition to these five sections, the pages also display the servers and link status and the information of the data being transmitted in text. When the system fails, the alarm information will be displayed on the visualization pages.

The advanced query page is shown in Figure 14. The advanced query page requires an advanced authentication command to access. The advanced query page supports detailed information query about data storage for specifying day on data servers, results of file MD5 validation display, and even keyword retrieval.

4. Summary

This paper completes the construction and development of the astronomical data visualization system and provides a complete set of management system for transferring data from the astronomical observation site to the data center. We have completed the four-layer system architecture design based on the analysis of the advantages and disadvantages of the existing astronomical data transmission system and the actual needs of the Xinjiang Astronomical Observatory. During the development process, we have fixed the deficiencies in the original design, and the system was stable during the last 1000 hours of testing. This paper provides a feasible astronomical data transmission scheme, which assists administrators in managing the transmission process through the log system and the visual interface. As a newly developed system, the astronomical data visualization

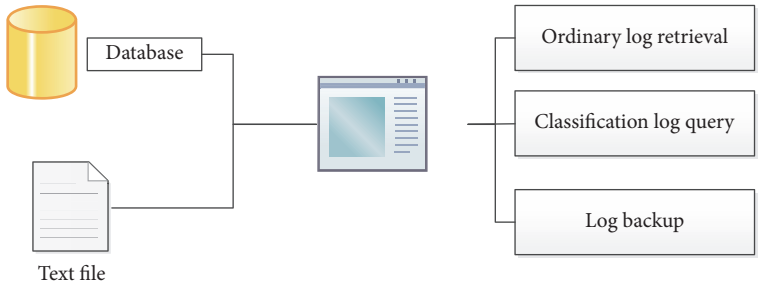


Figure 3: Log subsystem management interface functional structure.

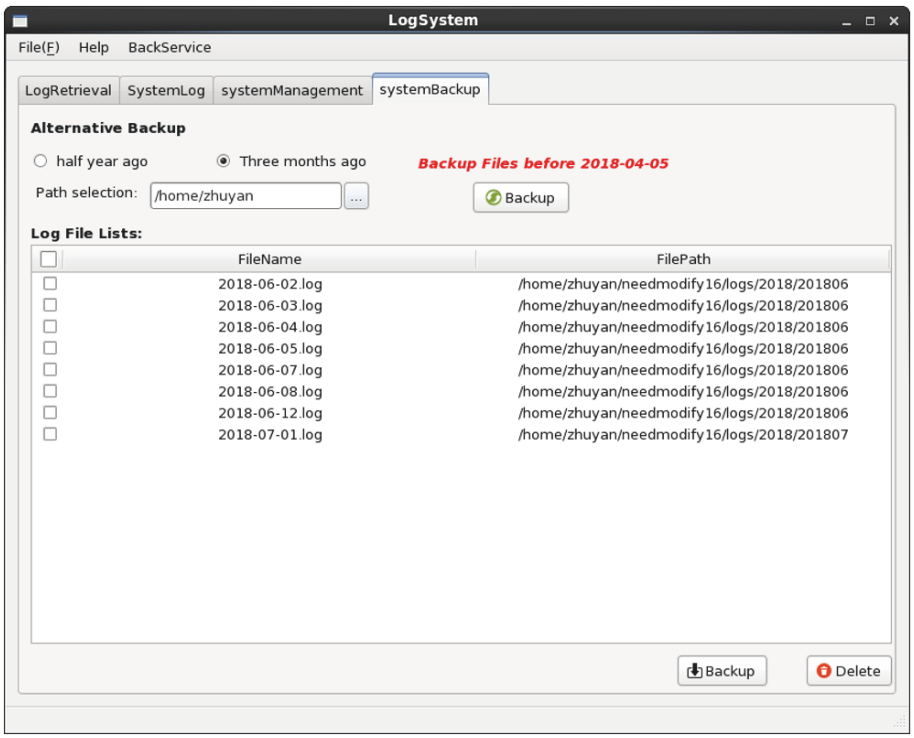


Figure 4: Log retrieval interface.

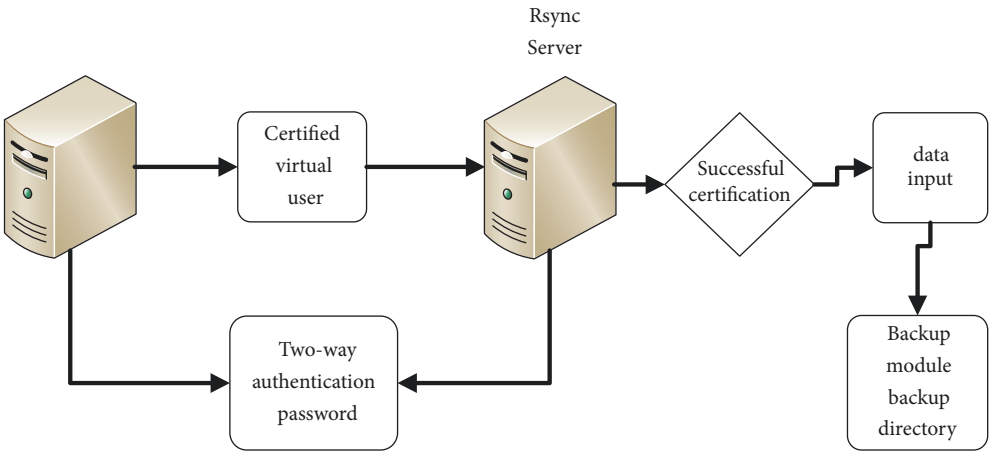


Figure 5: Rsync authentication process.

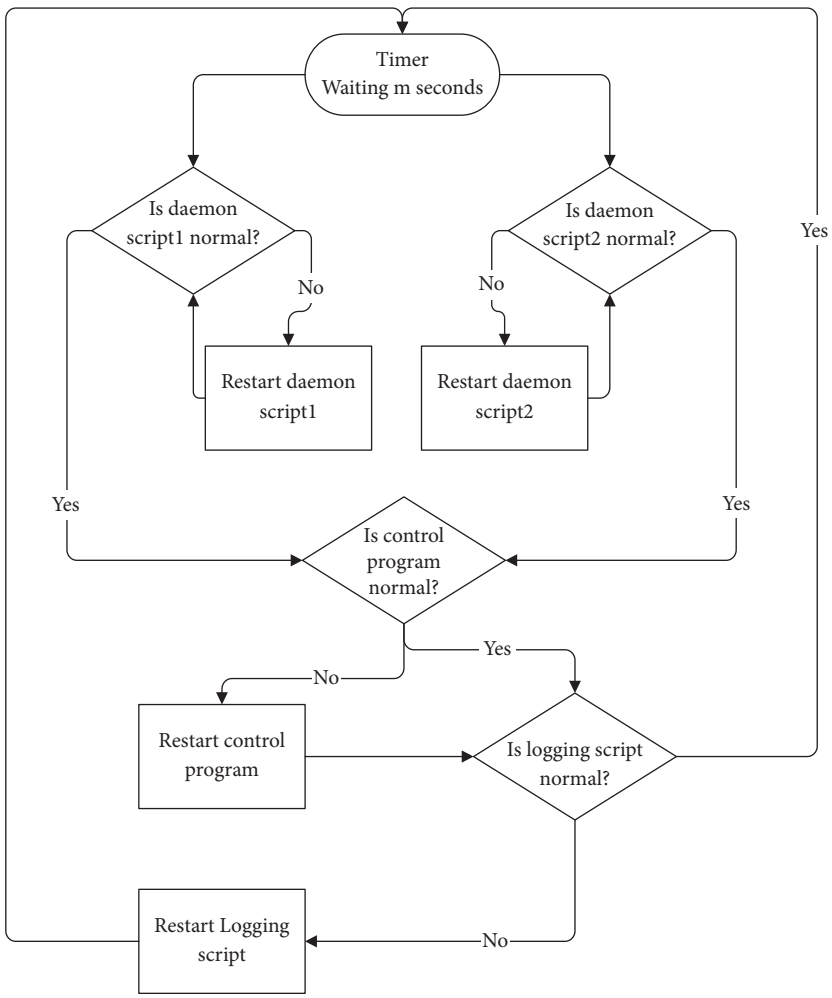


Figure 6: Daemon scripts running flowchart.

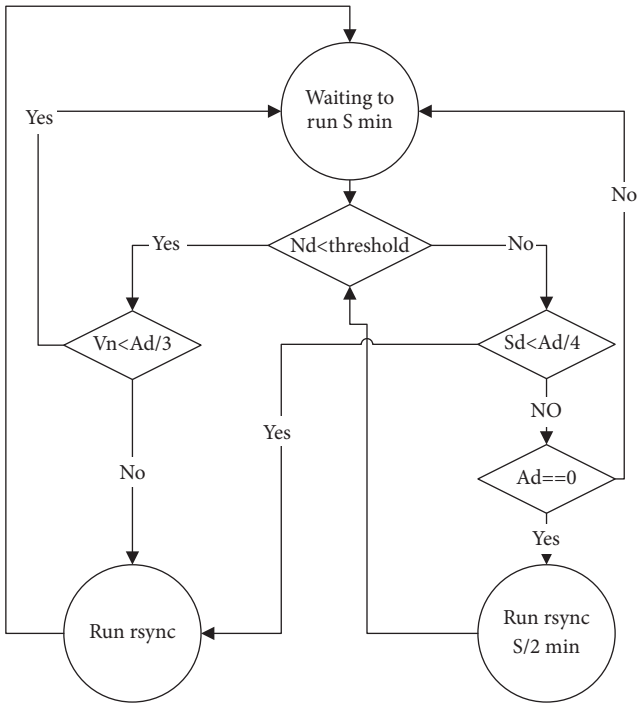


Figure 7: VSAN algorithm flowchart.

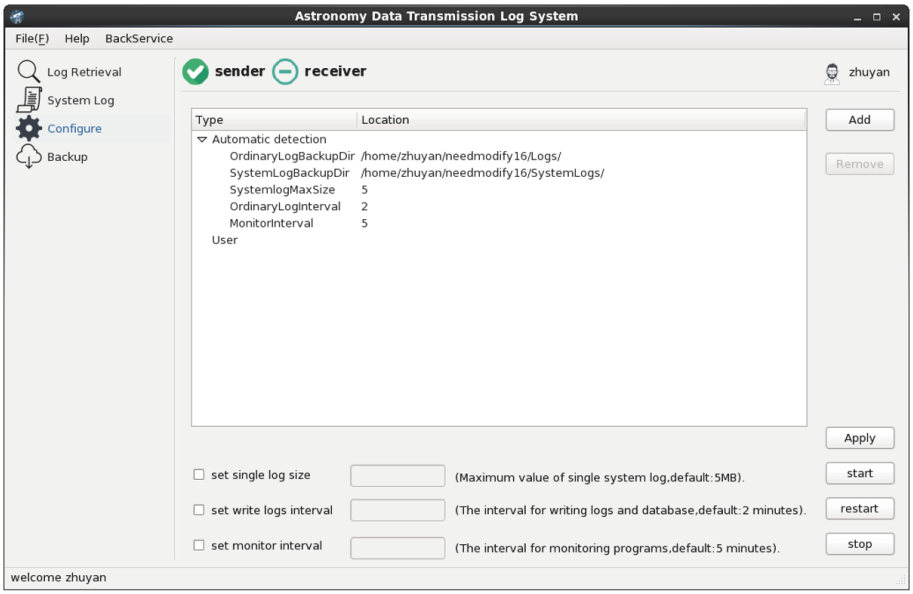


Figure 8: Control interface.

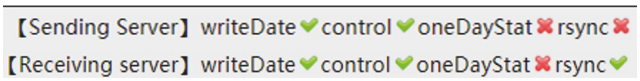


Figure 9: The first part of visualization web pages.

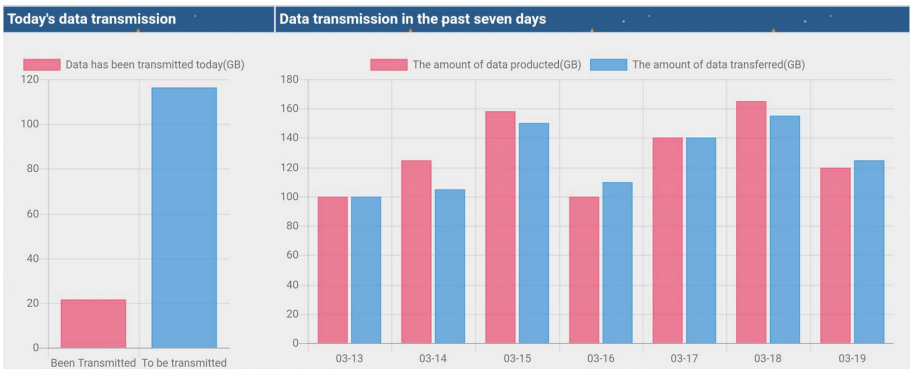


Figure 10: The second part of visualization web pages.

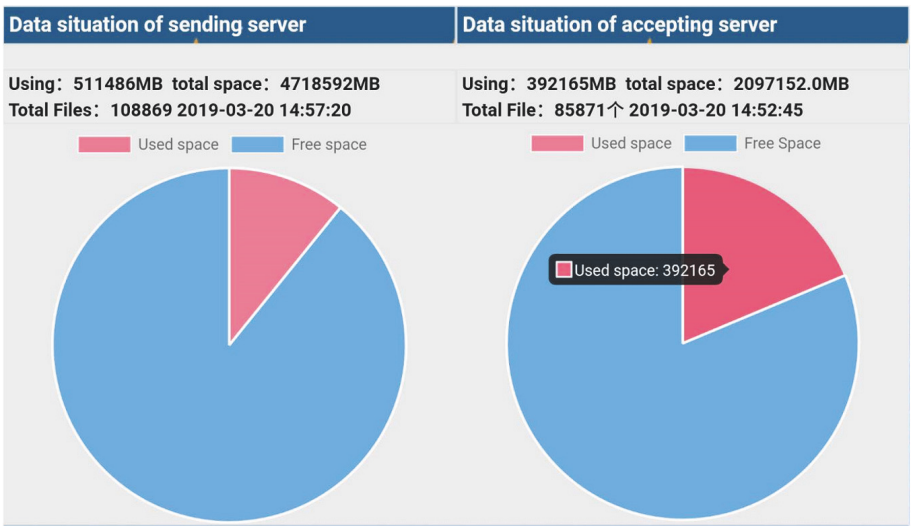


Figure 11: The third part of visualization web pages.

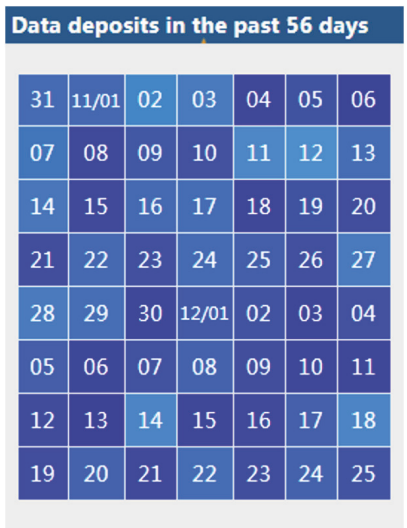


Figure 12: The fourth part of visualization web pages.

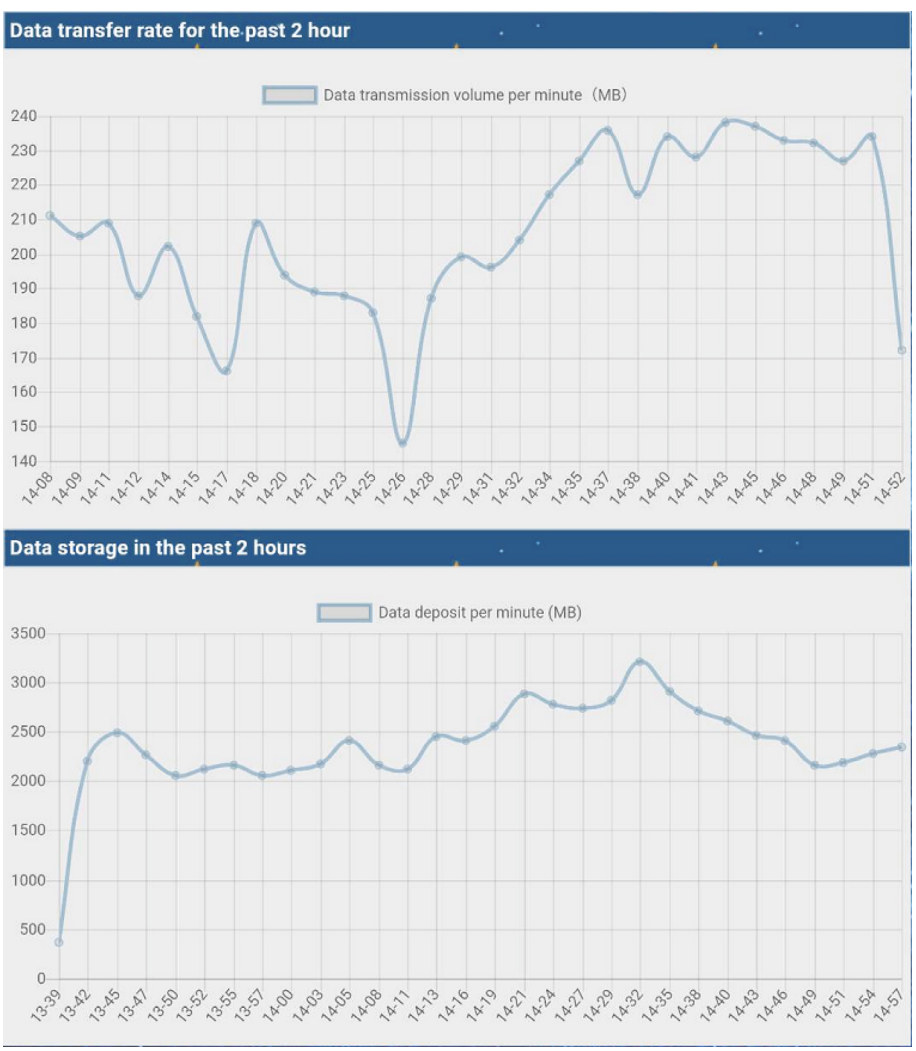


Figure 13: The fifth part of visualization web pages.

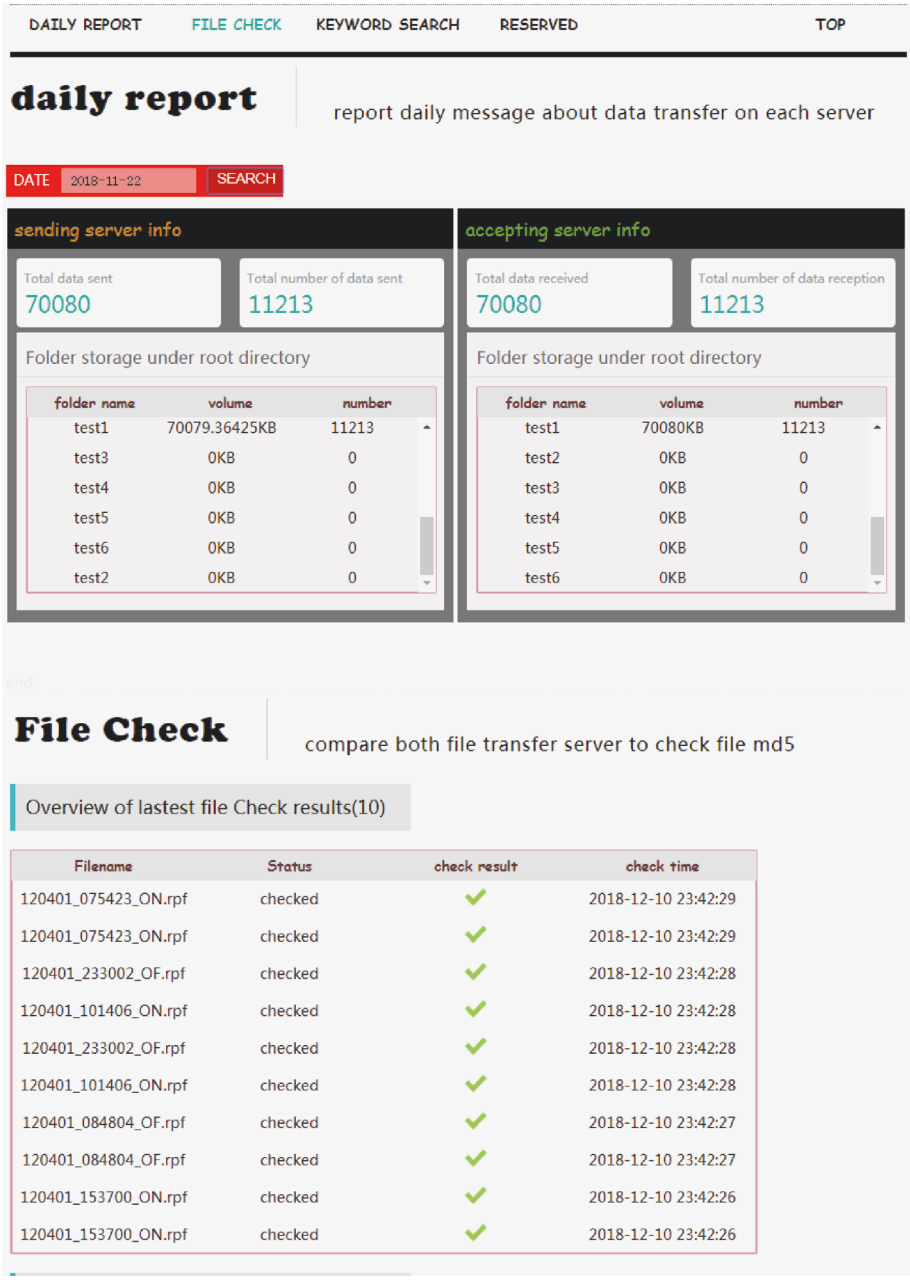


Figure 14: The advanced query page.

transmission system is still insufficient and will be further improved in the future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the support of the National Natural Science Fund of China (11873082, U1531125, 11803080, and 11503075), National Key Basic Research Program of China, 973 Program 2015CB857100, National Key Basic Research and Development Program 2018YFA0404704, and Youth Innovation Promotion Association CAS.

References

[1] P. Rosen, B. Wang, A. Seth et al., "Using contour trees in the analysis and visualization of radio astronomy data cubes," 2017, https://arxiv.org/abs/1704.04561.

[2] Z. D. Stephens, S. Lee Y, F. Faghri et al., "Big data: astronomical or genomical?" *PLoS Biology*, vol. 13, no. 7, Article ID e1002195, 2015.

[3] R. Prestage, K. Constantikes, T. Hunter et al., "The green bank telescope," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1382–1390, 2009.

[4] R. Nan, D. Li, C. Jin et al., "The five-hundred-meter aperture spherical radio telescope (FAST) project," *International Journal of Modern Physics D*, vol. 20, no. 6, pp. 989–1024, 2011.

[5] D. Li, P. Wang, L. Qian et al., "FAST in space: considerations for a multibeam, multipurpose survey using china's 500-m aperture spherical radio telescope (FAST)," *IEEE Microwave Magazine*, vol. 19, no. 3, pp. 112–119, 2018.

[6] D. H. Schaubert, A. O. Boryssenko, A. Van Ardenne et al., "The Square Kilometer Array (SKA) antenna," in *Proceedings of the 6th IEEE Phased Array Systems and Technology Symposium, Array 2003*, pp. 351–358, USA, October 2003.

[7] C. Carilli and S. Rawlings, "Motivation, key science projects, standards and assumptions," *New Astronomy Reviews*, vol. 48, pp. 979–984, 2004.

[8] G. Rhee, *Looking Ahead in Wonder: Telescopes at the Cosmic Frontier*, Springer, New York, NY, USA, 2013.

[9] E. Dovgan, C. Knapic, M. Sponza, and R. Smareglia, "A new archival infrastructure for highly-structured astronomical data," *Experimental Astronomy*, vol. 45, no. 1, pp. 41–55, 2018.

[10] A. Wicenec, J. Knudstrup, and S. Johnston, "ESO's next generation archive system," *Messenger*, vol. 129, pp. 27–31, 2002.

[11] A. Wicenec and J. Knudstrup, "ESO's next generation archive system in full operation," *The Messenger*, vol. 129, pp. 27–31, 2007.

[12] S. J. Tingay, R. Goeke, J. D. Bowman et al., "The Murchison widefield array: the square kilometre array precursor at low radio frequencies," *Publications of the Astronomical Society of Australia*, vol. 30, 2013.

[13] A. Wootten and A. Thompson, "The atacama large millimeter/submillimeter array," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1463–1471, 2009.

[14] A. Wicenec, S. Farrow, S. Gaudet, N. Hill, H. Meuss, and A. Stirling, *Astronomical Data Analysis Software and Systems (ADASS) XIII*, vol. 314, 2004.

[15] A. Manning, A. Wicenec, A. Checcucci, and J. A. Gonzalez Villalba, *Astronomical Data Analysis Software and Systems XXI*, P. Ballester, D. Egret, and N. P. F. Lorente, Eds., vol. 461, Astronomical Society of the Pacific Conference Series, 2012.

[16] C. Wu, A. Wicenec, D. Pallot, and A. Checcucci, "Optimising NGAS for the MWA archive," *Experimental Astronomy*, vol. 36, no. 3, pp. 679–694, 2013.

[17] Q. Xu, L. Yi, L. Li, M. Chen, N. Wang, and Y. Li, "A rapid feed switching mechanism design for NSRT," in *Ground-Based and Airborne Telescopes VII*, vol. 10700, International Society for Optics and Photonics, 2018.

[18] N. Wang, *Scientia Sinica Physica, Mechanica & Astronomica*, vol. 44, 2014.

[19] R. Rischpater, "Application development with Qt creator," 2014.

[20] Y. J. So, *SyncML Data Sync System and Data Exchanging Method for Data Exchange between Clients: U.S. Patent 7,917,653*, 2011.

[21] D. Rasch and R. C. Burns, "In-place rsync: file synchronization for mobile and wireless devices," in *Proceedings of the USENIX Annual Technical Conference, FREENIX Track*, p. 100, 2003.

Hindawi
Advances in Astronomy
Volume 2019, Article ID 9538061, 12 pages
https://doi.org/10.1155/2019/9538061

Research Article

Rain Attenuation Study at Ku-Band over Earth-Space Path in South Korea

Sujan Shrestha and Dong-You Choi

Department of Information and Communications Engineering, Chosun University, Republic of Korea

Correspondence should be addressed to Dong-You Choi; dychoi@chosun.ac.kr

Received 17 October 2018; Accepted 29 January 2019; Published 3 March 2019

Guest Editor: Feng Wang

Rain attenuation measurement techniques are studied with appropriate prediction of rain attenuation at Ku-band for Koreasat 6. This is accomplished by the establishment of experimental setup in Mokdong at 12.25 GHz link. The databases are analyzed for three years, 2013 till 2015. During observation period, rainfall rate of 50 mm/hr is obtained which is measured by OTT Parsivel showing the signal attenuated by 10.7 dB for 0.01% of the occurrence. Comparison with the measured data demonstrates that the proposed technique provides sufficiently accurate estimation for Ku-band signal attenuation in site specifically whose effectiveness is performed through the statistical analysis against the established rain attenuation models. The proposed technique is judged through the error matrices where relative error margins of 52.82, 4.11, and 23.64% are obtained for 0.1%, 0.01%, and 0.001% of the occurrence.

1. Introduction

Absorption and scattering by rain at frequencies above 10 GHz can result in the reduction of transmitted signal amplitude, which in turn reduce the reliability, availability, and performance of the communication link [1]. Ample spectrum is achieved through the satellite communication, but atmospheric propagation affects the service period at higher frequency such as Ku (14/12 GHz) bands [2]. This band of frequencies is suitable for direct to home (DTH) multimedia services, but significant attenuation is resulted due to rain during the transmission period [3]. The systems have been designed to operate at an acceptable performance level by providing adequate power margins on the uplink and the downlink segments [4]. Due to the lack of direct measurement of rain attenuation, modelling and prediction approaches are used for the estimation of attenuation [5]. Empirical method is suitable as compared to the physical methodology [6]. The variation of attenuation is dependent on specific attenuation, frequency, and polarization [7]. ITU-R P. 618-13 [8] shows reasonable good estimation for operating frequency till 55 GHz [4]. The analysis for 1-minute instance for rain rate is analyzed in local environment [9–12]. Additionally, the studies are carried out for Yong-in Station [13, 14]. Furthermore, rain induced attenuation is studied for beacon signals at Ka-band frequency in [15].

The technique for predicting the rain attenuation of Ku-band satellite signal during rain events has been presented for Mokdong Station. In addition, the established method of study of signal attenuation due to rain is analyzed [16–23]. The rain attenuation has been predicted from the synthetic storm technique and examination of storm speed effect, which is based on one year satellite beacon signal, rain intensity and storm speed measurement in Malaysia [24]. The site diversity techniques are studied in South East Asia [25]. The statistical variation has been assessed from the propagation experiment that was carried out in Madrid, Spain [26]. A channel model has been developed to predict time series of rain attenuation and the obtained rain rate and attenuation statistics at tropical and temperate regions are compared [27]. Similarly, the diurnal and monthly variation have been studied for Ka-band communication link [28]. The renowned methods for the estimation of signal attenuation due to rain have been analyzed [6, 8, 29–34]. However, the need for better prediction model is necessary to be studied as per the measurement performed in local environmental condition

which signifies the need for the proposed work. The rain rate value for 0.01% of occurrence is a crucial parameter in the analysis of slant path attenuation [35]. The short descriptions of various techniques are stated in Section 2. Section 3 highlights the equipment building and the measurement system. The outcomes are detailed in Section 4. Section 5 gives conclusion.

2. Studied Models

Three methods are crucial for the better estimation of signal attenuation due to rain which can be categorized as the evaluation of specific attenuation [38], appropriate height of rain drops [39], and calculation for over all-time percentages. The relationship between specific attenuation, $_R$ (dB/km) and rain rate R (mm/h) is given as

$$\gamma_R = kR^a \tag{1}$$

Two parameters for 12.25 GHz under circular polarization of the present location are obtained as k=0.024205 and α=1.151616. The ITU-R P. 839-4 [39] gives the required rain height estimation.

2.1. ITU-R P. 618-13. The ITU-R rain attenuation model [8] is the most widely accepted model by the international propagation community. The input parameters needed for the model are point rainfall rate of the location in 0.01% of an average year (mm/hr), height above sea level of the earth station (km), elevation angle (θ), the latitude of the earth station (φ), and frequency (GHz). The effective path length (L_E) can be obtained using (2a), where $_{0.01}$ is the vertical adjustment factor and L_R is calculated from the horizontal projection. The signal attenuations due to rain are analyzed from specific attenuation as

$$L_E = L_R \ \nu_{0.01} \text{ km} \tag{2a}$$

$$A_{0.01} = \gamma_R \ L_E \text{ dB} \tag{2b}$$

Similarly, (2c) can be used to find the attenuation for other instances of time percentage [8].

If $p \geq 1\%$ or $|\varphi| \geq 36^\circ$, $\beta = 0$.

If $p < 1\%$ and $|\varphi| < 36^\circ$ and $\theta \geq 25^\circ$, $\beta = -0.005(|\varphi| - 36)$.

Otherwise, $|\varphi| < 36^\circ$ and $\theta < 25^\circ$: $\beta = -0.005(|\varphi| - 36) + 1.8 - 4.25 \sin\theta$.

$$A_p = A_{0.01} \left(\frac{p}{0.01}\right)^{-(0.655+0.033\ln(p)-0.045\ln(A_{0.01})-\beta(1-p)\sin\theta)} \text{ dB} \tag{2c}$$

The model has been applied against the measurement result for earth-space communication at 12.25 GHz under circular polarization.

2.2. Unified Method. Full rainfall rate distributions are used to analyze the signal attenuation due to rain. The numerical coefficients are obtained by multiple nonlinear regressions, obtained from available ITU-R databanks [40]. Further description can be found in [29].

2.3. Dissanayake, Allnutt, and Haidara (DAH) Model. It is based on the log-normal distribution of rain rate and rain attenuation. Even though this method is similar to the ITU-R P. 618-12 [8] in which rain heights vary, in DAH model it is fixed to 5 km [6].

2.4. Simple Attenuation Model (SAM). It was developed to utilize algebraic equations for specific attenuation coefficients, isotherm height, and path profile on earth-space communication links operating in the range from 10 to 35 GHz. This method consists of the relationship between specific attenuation and rain rate, statistics of the point rainfall intensity and spatial distribution of rainfall [30].

2.5. Crane Global (CG) Model. The model depends on the 0°C isotherm height, H_0, and excessive precipitation events [31].

2.6. Ramachandran and Kumar (RK) Model. It studies the statistics that is based on elevation angle less than 60° [32].

2.7. Gracia Lopez (GL) Model. It features the suitable method for rain attenuation prediction to be used in satellite radio link which is tested over 77 satellite links placed in Europe, the US, Japan, and Australia. The coefficients details necessary for the calculation of rain attenuation are mentioned in [33].

2.8. Karasawa Model. This model was accepted by the European Space Agency as a suitable model which was designed to enhance the prediction performance at lower probability exceedance levels [34].

2.9. Proposed Approach. The proposed technique is based on the behavior of signal as it is attenuated towards the direction of communication path between satellite and ground station at 0.01% of occurrence. The two quantities are related as

$$A_{0.01} = kR^{\alpha}{}_{0.01} * L_{eff} \text{ dB} \tag{3}$$

The coefficient values are obtained from ITU-R P. 838-3 [38]. The quantitative relation between the attenuation due to rain and effective path length, L_{eff}, value at 0.01% of the occurrence is given as

$$L_{eff} = \frac{A_{0.01}}{kR_{0.01}^{\alpha}} \tag{4a}$$

This approach is studied for the experimental values of three years, from 2013 to 2015. Additionally, rain rate and beacon signal level data [14] in Young-in Station are analyzed together with Mokdong-13 na gil station database. The empirical notations are determined from the curve fitting technique as given in (3). The regression analysis relates, L_{eff}, to rain rate value as

$$L_{eff} = 14.42 * R_{0.01}^{-0.2782} \tag{4b}$$

Finally, substituting (4b) to (3), we get

$$A_{0.01} = 14.42 * k * R_{0.01}^{\alpha-0.2782} \tag{4c}$$

Table 1: Link description for 12.25 GHz [36].

Type	Descriptions	Specification
System Location	Location	37.5447°N, 126.8833°E
	Elevation angle	45°
	Degree of tilt	197.5°
	Low-lying (km)	0.055
Receiver Antenna	Prototype	Off-set parabolic
	Operating range	10.95 ~ 31
	Polarization	Circular
	Gain	55 dB ± 2 dB
	Dynamic Range (dB)	18.3
Optical Disdrometer	Type	OTT Parsivel
	Measuring area	54 cm^2
	Particle size	0.2 to 25 mm
	Velocity drop	0.2 to 20 m/s
	Rain Intensity	Up to 1200 mm/hr
	Temperature range	-40°C to +70°C

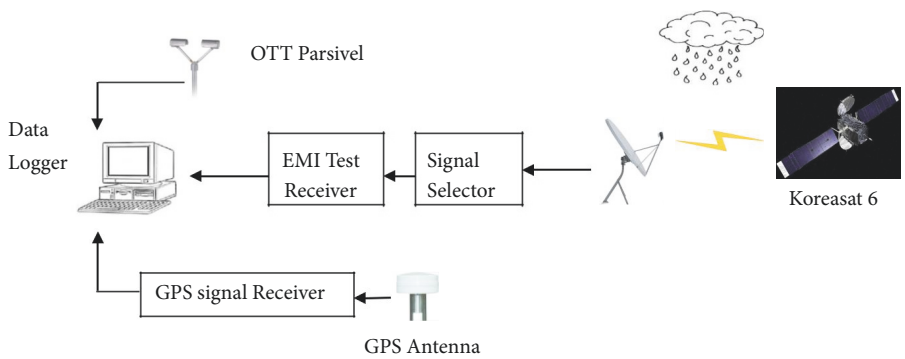


Figure 1: System diagram of experimental setup [36].

Due to the insufficiency of experimental rain attenuation and rain rate database from other locations of South Korea, the proposed method utilizes the extrapolation equation as stated in (2c). We have proposed the empirical relation between rain attenuation and rain rate at 0.01% of time. Furthermore, for other time percentages we have used extrapolation equation as stated by ITU-R P. 618-13 [8]. Additionally, the proposed approach differs with other models as per the empirical relationship obtained between rain attenuation and rain rate at 0.01% of time.

3. Experimental Setup

Rain attenuation of Ku-band beacon signal over an earth-space link has been measured by a receiving signal at frequency 12.25 GHz transmitted with circular polarization from satellite Koreasat 6, situated at 116°E at an elevation angle of 45° since 2013 [36]. The sampling interval of the data acquisition used in the present study is 10 second which is averaged over 1-minute distribution. The beacon signal strength was measured using a spectrum analyzer at 10 second interval while monitoring of the data was done on a personal computer. In addition, an optical disdrometer is used to measure the rain rates which operate simultaneously with the monitoring system of satellite beacon signal whose specification is given in Table 1. The beacon receiver measures the amplitude fluctuations in the received signal with enough accuracy and dynamic range. The radome prevents the antenna from wetting condition. The link has availability of 99.95% whose details are given in Table 1.

The received beacon signal levels are sampled for every 10 seconds and averaged over 1 minute. The rain rates are measured by OTT Parsivel with 99.95% of the validity. OTT Parsivel is the laser based optical disdrometer. The parallel rays between emitter and receiver help to measure the required precipitation characteristics with the change in signal output voltage. This determines the magnitude of drops particle size whereas the interval of original signal helps in determination of the velocity. This determines the required rain rate. The sensor built on the OTT Parsivel conveys messages to computer with the support of RS 485. Heating system obstructs the ice build-up process for every second. Further detail procedures are listed in [41, 42]. The system diagram of setup is shown in Figure 1.

Figure 1 shows the off-set antenna is faced towards Koreasat 6 satellites. The noise figure of 1.5 dB and bandwidth of

Figure 2: Variation of 12.25 GHz signal attenuation [36].

1000 MHz are maintained where the output is passed through the system that splits signal in 4 dB range which goes through spectrum analyzer. Finally, its output is stored in database. The excess attenuation over the clear weather values results in the calculation of required rain attenuation values which are measured simultaneously with the rain rates. The GPS antenna records the relative position through GPS signal receiver. Figure 2 shows the recorded signals.

Thus, the required attenuation is calculated as

$$\text{Attenuation (dB)} = \text{RSL}_{\text{clear sky}} - \text{RSL}_{\text{rainy}} \quad (5)$$

where RSL stands for receiver signal level. The procedure as depicted in [14] is used to determine the required rain rate and attenuation for 1-minute instance. The required rain rate and attenuation are calculated by arranging the data in the descending order. For instance, at 0.01% of the occurrence, 1-minute rain rate and attenuation values are taken in 158 ((3×365×24×60×0.01)/100) = 157.68 ≈ 158). In each year of measurement the values of 53 ((1×365×24×60×0.01)/100) = 52.56 ≈ 53) were taken. The annual values of 1-minute rain rate from 2013 till 2015 are shown in [15, 43]. Figure 3 shows the rain attenuation for 12.25 GHz frequency which is obtained on the yearly basis and when combined together from 2013 till 2015. The rain rate is greater in 2013 and 2014 which resulted in higher rain attenuation values of 10.8 dB and 12.3 dB, respectively, for 0.01% of the time. The combined data is considered for further analysis, which indicates the lower attenuation values at higher time percentages and higher values at lower time percentage. As seen, the attenuation values of 7.9, 10.7, and 19 dB are observed for 0.1%, 0.01%, and 0.001% of the occurrence, respectively. The higher attenuation for 1% of an average year might be due to the summation performed for 10 seconds measured data to be equivalent to 1-minute interval. Additionally, it might be due to the higher rain rate experienced for the time interval of 1% of an average year.

Furthermore, the plot of rainfall rate distribution for various integration times near location of Icheon is presented in [15] which shows the rain rate obtained from the experimental 1-minute rainfall amount database provided by Korea Meteorological Administration (KMA) from 2004 to 2013 [44]. The effect of rain attenuation on the terrestrial communication links has been studied over the 18, 38, and 75 GHz links [45–47]. This signifies the need for measurement result of RRA with longer experimental periods, which is still under progress.

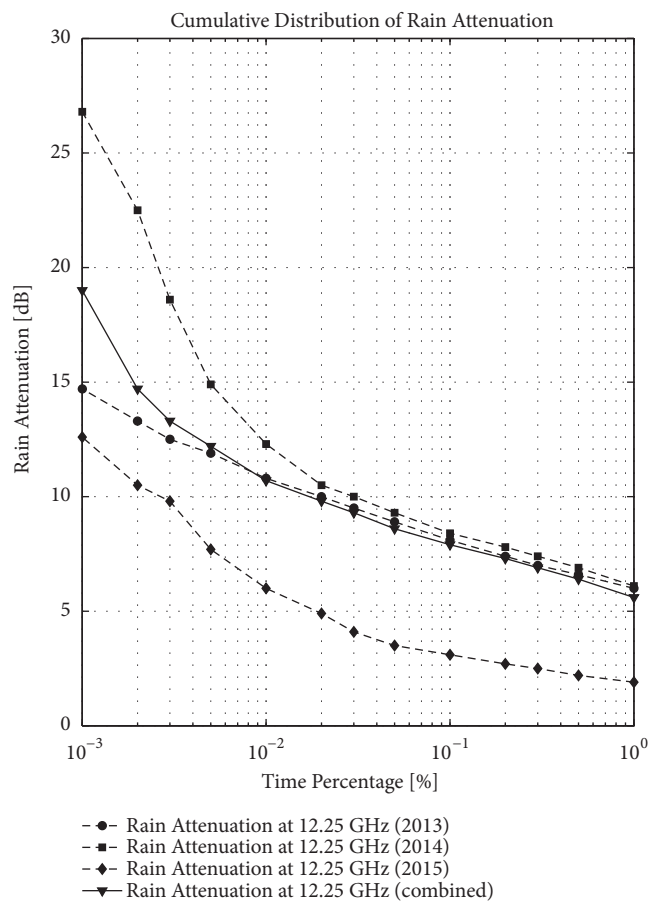


Figure 3: Distribution of Rain attenuation at Mokdong [36].

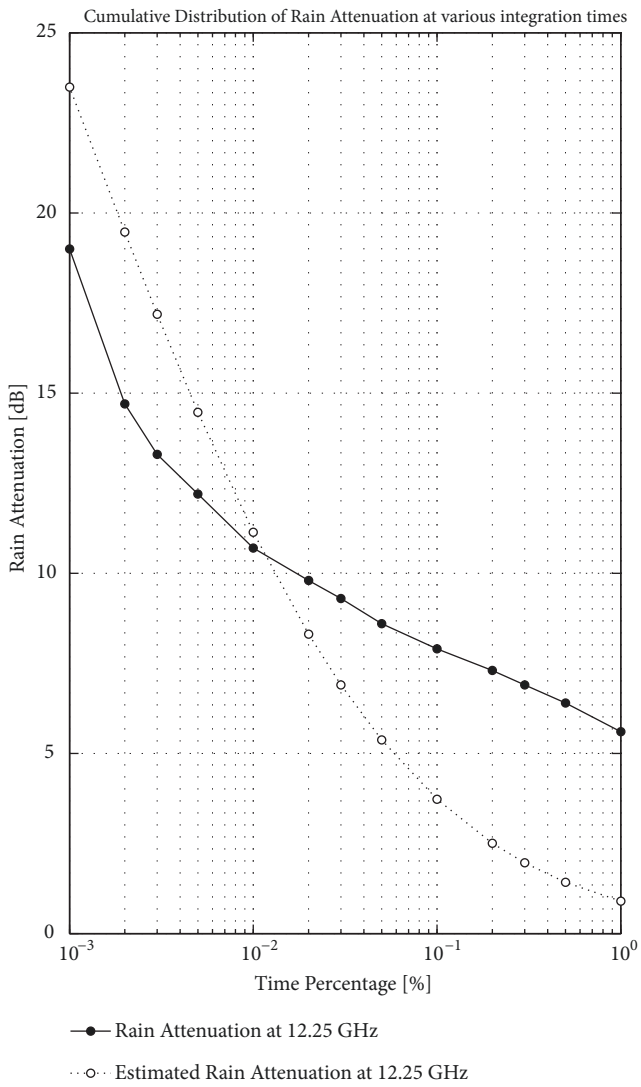


Figure 4: Variation of signals against ITU-R P. 618-13 for various integration times [37].

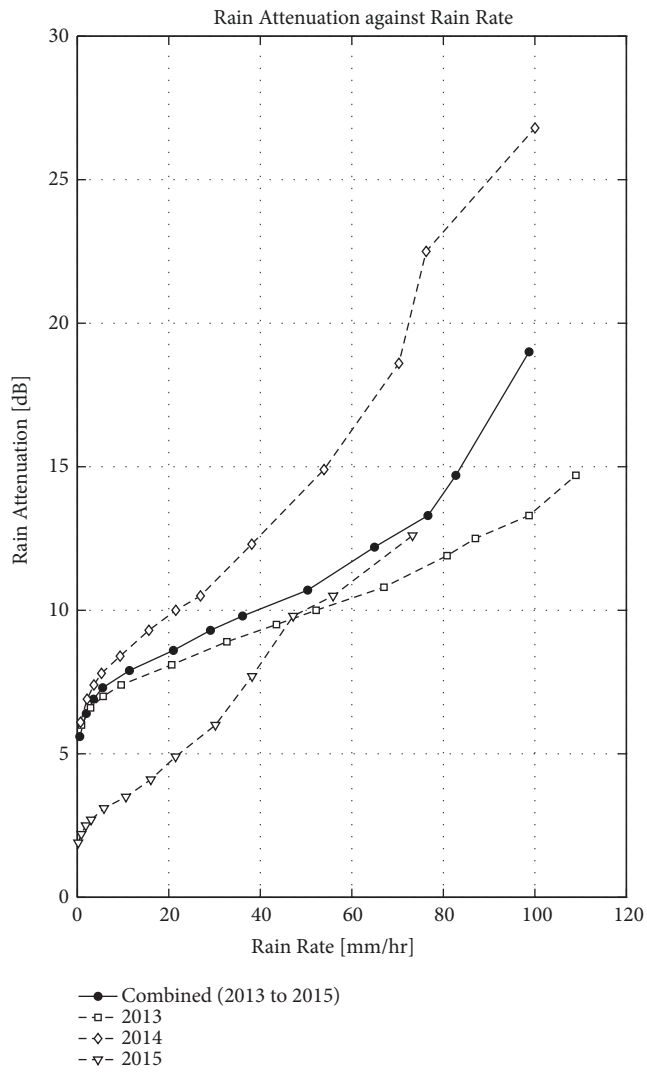


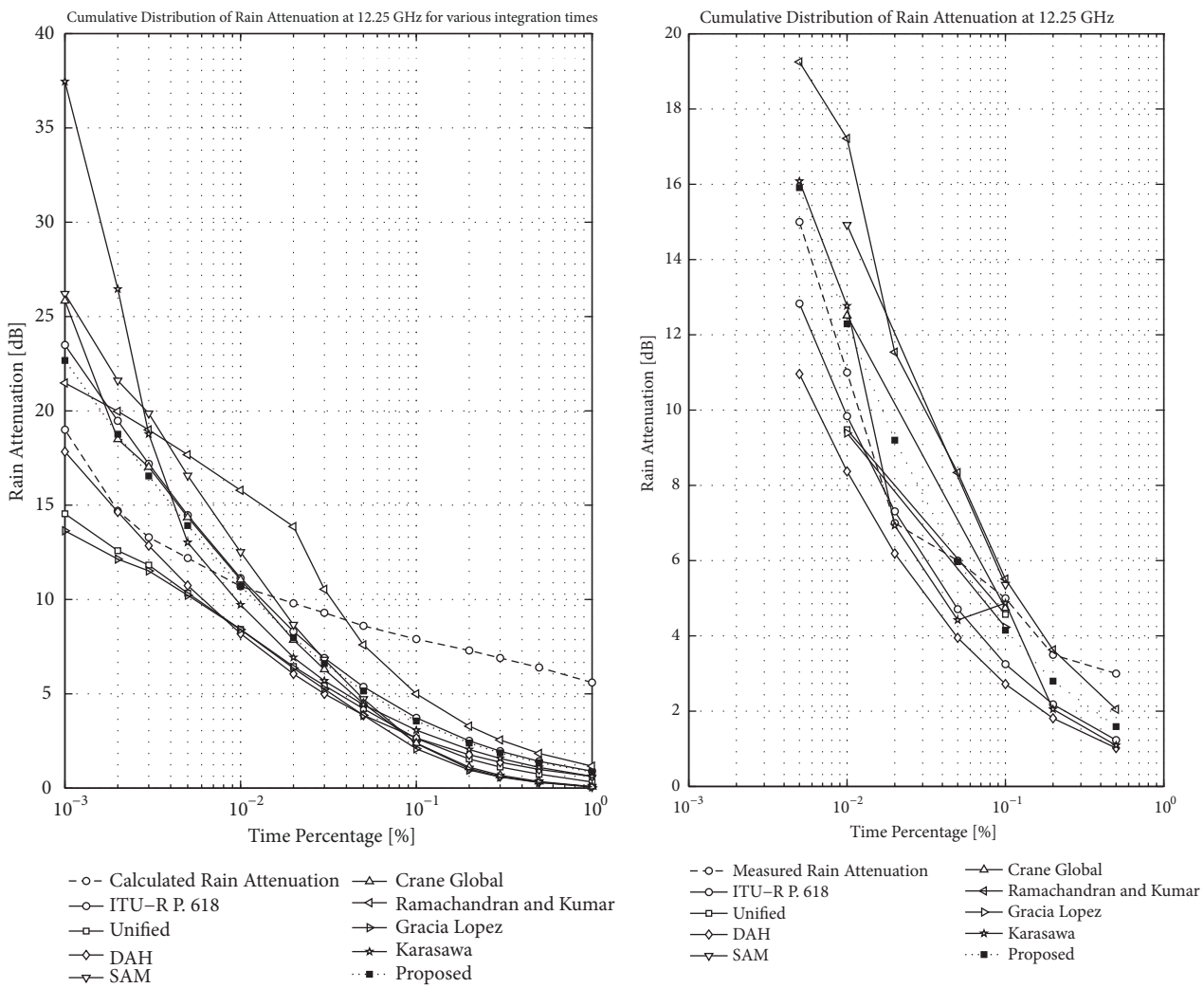
Figure 5: Rain attenuation versus rain rate [37].

4. Result and Discussions

Distributions of rain attenuation at 12.25 GHz are shown in Figure 4. It shows that ITU-R P. 618-13 gives close value at 0.01% of the time but deviates in lower and higher time period. However, difference in prediction is lower as compared to higher time period. As observed, calculated 1-minute rain attenuation values are 7.9, 10.7, and 19 dB and ITU-R P. 618-12 estimates 3.73, 11.14, and 23.49 dB for 0.1%, 0.01%, and 0.001% of the occurrence, respectively. This might be due to the low spatial resolution of the matrices. ITU-R P. 618-13 gives the reasonable estimation at 0.01% of the time, but for, other time percentages, this method results in poor estimation. This highlights the need for suitable 1-minute rain attenuation model for Ku-band satellite communication link. Furthermore, the relation between rain attenuation and rain rate is shown in Figure 5 for three years of measurement along with the combined values. This figure indicates that there is positive correlation between the rain rate and rain attenuation. The manuscript illustrates the suitable approach along with the categorization of the applicable prediction methods over the earth-space path for Ku-band satellite communication.

The proposed method along with the rain attenuation models is studied against the measured result as shown in Figures 6(a) and 6(b) for 12.25 GHz satellite communication links for Mokdong and Yong-in Station, respectively.

In 12.25 GHz from Figure 6(a), ITU-R P. 618-13 and the proposed approach are predicted preferably, but the chances of overestimation and underestimation are greater for lower and higher time percentages, respectively. The calculated rain attenuation values are 7.9, 10.7, and 19 dB at 0.1%, 0.01%, and 0.001% of occurrence, whereas ITU-R P. 618-13 and the proposed approach predict 3.73, 11.14, 23.49 dB and 3.56, 10.7, 22.67 dB, respectively. CG and SAM methods give similar result which overestimate and underestimate the measured values at lower and higher time percentages, respectively.



(a) Prediction at 12.25 GHz for Mokdong Station [37].

(b) Prediction at 12.25 GHz for Yong-in Station [37].

Figure 6

Interestingly, CG generates close approximation against the measured rain attenuation when P < 0.01%. It predicts 2.39, 11.05, and 25.84 dB, respectively, at 0.1%, 0.01%, and 0.001% of the time whereas SAM predicts 2.37, 12.52, and 26.2 dB, respectively. DAH, Unified, and GL show underestimation of rain attenuation for all-time percentage, but the closer prediction is obtained by the application of DAH method in low time of occurrence. As seen, DAH, Unified, and GL show 2.66, 8.19, and 17.84 dB; 2.63, 8.42, and 14.55 dB; and 2.09, 8.41, and 13.64 dB, respectively at 0.1%, 0.01%, and 0.001% of the time. Karasawa shows preferable estimation at 0.01% of the time, but greater overestimation and underestimation are observed for lower and higher time percentages, respectively, as compared to other methods. Additionally, RK gives the better estimation for higher time percentage, but overestimation is noted on lower time percentage. Karasawa and RK estimate 3.07, 9.72, and 37.45 dB and 5, 15.79, and 21.48 dB, respectively, at 0.1%, 0.01%, and 0.001% of occurrence. Additionally, effectiveness of methods is analyzed from error analysis.

Prominent methods along with proposed technique are tested in Yong-in Station. Unfortunately, unavailability of sufficient database for all-time percentages has limited the significance of the study. However, data at 0.1 and 0.01% of the time for 2007 as noted from [14] are considered for the suitability of methods. As depicted in Figure 6(b), ITU-R P. 618-13, RK, and proposed approach give the reasonable estimate with P ≤ 0.01%. For instance, measured rain attenuated signal variations are 4, 11, and 15 dB at 0.1, 0.01, and 0.005% of the occurrence while ITU-R P.618-13, proposed approach, and RK estimate 3.25, 9.84, and 12.83 dB; 4.15, 12.29, and 15.91 dB; and 5.51, 17.22, and 19.25 dB, respectively. Due to the dependability on the experimental data of rain rate that is available for only 0.1 and 0.01% of the occurrence, there is the limitation on analyses of Unified, SAM, CG, and GL methods. The finding suggests that these models are unable to give better estimation but at 0.1% of occurrence, SAM, CG, and GL result in closer estimation for measured rain attenuation data. For example, the attenuation values of 4.57, 9.48 dB; 5.36, 14.92 dB; 4.74, 12.51 dB; and 4.23, 9.39 dB are obtained

from Unified, SAM, CG, and GL methods for 0.1 and 0.01% of occurrence. Thus, those models preference can be better categorized from the recent RRA data from Mokdong Station. DAH method shows underestimation whereas Karasawa gives the fair close prediction of measured rain attenuation. For instance, DAH and Karasawa result in 2.72, 8.37, and 10.96 dB and 4.87, 12.77, and 16.08 dB, respectively, for 0.1, 0.01, and 0.005% of the occurrence.

Rain attenuation prediction model for Earth-satellite link is determined for exceeding time percentages 0.001% to 1%. Hence, the percentage errors, ε (P), between measured Earth-satellite attenuation data ($A_{\%p,\text{measured}}$) in dB and the model's predictions ($A_{\%p,\text{predicted}}$) in dB are calculated as follows:

$$\varepsilon(P)_T = \frac{A_{\%p,predicted} - A_{\%p,measured}}{A_{\%p,measured}} \times 100\,[\%] \qquad (6)$$

Variance (STD) and quadratic mean (RMS) are calculated as [9]. Ratio of predicted to measured attenuation is calculated and the natural logarithm of these error ratios is used as a test variable [48]. Average (μ_v), variance (σ_v), and quadratic mean (ρ_v) of test variable are calculated to provide statistics for prediction procedure whose comparison is given in Tables 2, 3, and 4 for 12.25 GHz link for Mokdong and Yong-in Station, along with the evaluation procedures adopted for comparison of prediction methods as recommended by ITU-R P.311-15 [48].

As depicted from Table 2, in the case of 12.25 GHz in Mokdong Station, ITU-R P.618-13, Unified, DAH, SAM, CG, GL, Karasawa, and the proposed technique result in greater inappropriateness in 0.05% ≤ P ≤ 1% as supported by higher variance and quadratic mean. Conversely, RK shows low error for same time percentage. Hence, in greater occurrence probability when 0.05% ≤ P ≤ 1%, RK method is suitable. When 0.001% ≤ P < 0.01%, ITU-R P. 618-13, SAM, CG, RK, Karasawa, and the proposed approach result in greater unsuitability. Interestingly, overestimation is observed to be lesser for ITU-R P. 618-13 and proposed procedure as shown by decreased variance and quadratic mean.

In contrast, DAH, Unified, and GL models give underestimation against the calculated rain attenuation, but less underestimation is shown by DAH method which signifies its suitability in low time occurrence. In addition, at 0.01%, ITU-R P. 618-13 and the proposed approach seem to provide fairly more accurate results than the other models of interest. To illustrate, ITU-R P. 618-13 results in error percentage of 53%, 4%, and 24% while it is 67%, 21%, 23%; 66%, 23%, 6%; 70%, 17%, 38%; 70%, 3%, 36%; 37%, 48%, 13%; 74%, 21%, 28%; 61%, 9%, 97%; 55%, 0%, 19% for Unified, DAH, SAM, CG, RK, GL, Karasawa, and the proposed approach at 0.1%, 0.01%, and 0.001% of occurrence, respectively. ITU-R P.618-13 and the proposed technique result in fewer values of μ_v, σ_v, and ρ_v. In order to support the scientific argument for these finding, we have tested the validity of prominent rain attenuation models and proposed techniques in Yong-in Station. Unfortunately, due to the limited availability of the data sources, accurate nature of these models cannot be justified clearly. However, in this region too, similar nature is observed from the application of ITU-R P. 618-13 and the proposed technique. The tests have been carried out from the data sources as mentioned in [14] for 2007.

Table 3 shows the similar trends of statistical results. ITU-R P. 618-13 and the proposed technique result in reasonable estimation for lower period 0.005% ≤ P ≤ 0.05%, as supported by decreased σ_v and ρ_v. Conversely, DAH and Karasawa give higher error chances for the same percentage of time where higher values of σ_v and ρ_v are observed. Hence, ITU-R P. 618-13 and the proposed technique have better effectiveness in low occurrence period, particularly at 0.01% of occurrence. In addition, in upper occurrence period for 0.1% ≤ P ≤ 0.5%, RK gives better estimation which is assisted by decreased error percentage, σ_v and ρ_v as compared to ITU-R P. 618-13, DAH, Karasawa and proposed procedure. As example, ITU-R P. 618-13 shows error chances of 22%, 11%, and 14% whereas 0%, 0.12%, and 6% for proposed procedure in 0.05%, 0.01%, and 0.005% of occurrence, respectively. Similarly, DAH, RK, and Karasawa show 34%, 24%, and 27%; 39%, 57%, and 0.28%; and 26%, 16%, and 7% for same percentage exceedance. Conversely, in 0.1 % ≤ P ≤ 0.5 %, ITU-R P. 618-13, DAH, RK, Karasawa, and the proposed procedure give 59%, 35%; 66%, 46%; 32%, 10%; 63%, 3% and 47%, 17% for 0.5% and 0.1% of occurrence, respectively. DAH, RK, and Karasawa generate excessive μ_v. In contrast, ITU-R P. 618-13 and proposed method result in less μ_v, as accomplished by less σ_v and ρ_v Additionally, for fixed two percentages of time particularly, 0.1% and 0.01% of occurrence, the calculated error matrices are tabulated in Table 4 due to the limited data sources for rest of the models performance as they are dependent on rain rate values.

Table 4 shows the CG gives less errors percentage values as supported by lower values of σ_v and ρ_v. As tested through ITU-R P. 311-15 [38] approach, this model gives lower μ_v, σ_v, and ρ_v. In opposite, Unified, SAM, and GL result in higher error chances. For instance, CG, Unified, SAM, and GL show 5%, 14%; 9%, 14%; 7%, 36%; 15%, 15% for 0.1% and 0.01% of occurrence, respectively. CG shows lower μ_v value as compared to other models, but still there is the requirement of test for the suitability of model throughout the all-time percentage ranging from 1% to 0.001% of occurrence. Thus, this paper gives more priority to estimate better prediction approach for the analyses done from Mokdong Station database due to the most recent and well-arranged system.

5. Conclusions

Various models are compared to predict rain attenuation at Ku-band over an earth-space path. The propagation impairments for 12.25 GHz satellite communication links in Mokdong Station, along with database provided from Young-in station are studied. The rain rates at 0.01% of the time are 50.35 and 59 mm/hr for two experimental locations, namely, Mokdong and Yong-in Stations. The results show that at higher time percentage when 0.01% ≤ P ≤ 1%, ITU-R P.618-13, Ramachandran and Kumar, proposed approach shows the better estimation of rain attenuation. Conversely, when 0.001% ≤ P ≤ 0.1%, ITU-R P. 618-13, DAH, and the proposed approach show good estimation against the measured results. The predictive capability of

Table 2: Calculated error percentage [0.001% to 1%] for 12.25 GHz.

Procedures	Matrices	Occurrence													ITU-R P.311-15		
		1	0.5	0.3	0.2	0.1	0.05	0.03	0.02	0.01	0.005	0.003	0.002	0.001	μ_v	σ_v	ρ_v
ITU-R P. 618-13	ε(P)	-0.84	-0.78	-0.72	-0.66	-0.53	-0.38	-0.26	-0.15	0.04	0.19	0.29	0.32	0.24			
	STD	0.59	0.53	0.47	0.41	0.28	0.13	0.01	0.10	0.29	0.43	0.54	0.57	0.48	-0.45	0.64	0.78
	RMS	4.70	4.98	4.94	4.79	4.17	3.23	2.40	1.49	0.44	2.27	3.89	4.77	4.49			
Unified	ε(P)	-0.94	-0.88	-0.84	-0.79	-0.67	-0.51	-0.42	-0.34	-0.21	-0.15	-0.11	-0.14	-0.23			
	STD	0.46	0.40	0.36	0.31	0.19	0.03	0.06	0.14	0.27	0.33	0.37	0.34	0.25	-0.02	0.63	0.63
	RMS	5.27	5.66	5.78	5.76	5.27	4.39	3.86	3.35	2.28	1.86	1.47	2.11	4.45			
DAH	ε(P)	-0.89	-0.85	-0.80	-0.76	-0.66	-0.55	-0.46	-0.38	-0.23	-0.12	-0.03	0.00	-0.06			
	STD	0.44	0.40	0.35	0.31	0.22	0.10	0.02	0.06	0.21	0.33	0.41	0.44	0.39	-0.76	0.66	1.01
	RMS	4.98	5.41	5.52	5.53	5.24	4.73	4.30	3.75	2.51	1.46	0.45	0.06	1.16			
SAM	ε(P)	-0.99	-0.95	-0.91	-0.86	-0.70	-0.45	-0.27	-0.12	0.17	0.36	0.49	0.47	0.38			
	STD	0.73	0.69	0.65	0.60	0.44	0.19	0.01	0.14	0.43	0.62	0.75	0.73	0.64	-0.87	1.35	1.61
	RMS	5.53	6.09	6.28	6.27	5.54	3.88	2.49	1.14	1.82	4.37	6.57	6.91	7.20			
Crane Global	ε(P)	-0.99	-0.95	-0.90	-0.85	-0.70	-0.48	-0.32	-0.20	0.03	0.18	0.28	0.26	0.36			
	STD	0.66	0.62	0.57	0.52	0.37	0.15	0.01	0.13	0.36	0.51	0.61	0.59	0.69	-0.9	1.26	1.55
	RMS	5.52	6.05	6.22	6.19	5.51	4.11	3.01	1.94	0.35	2.16	3.72	3.80	6.84			
Ramachandran and Kumar	ε(P)	-0.79	-0.71	-0.63	-0.55	-0.37	-0.12	0.13	0.42	0.48	0.45	0.43	0.36	0.13			
	STD	0.73	0.65	0.57	0.49	0.31	0.06	0.19	0.47	0.54	0.51	0.49	0.42	0.19	-0.21	0.62	0.65
	RMS	4.43	4.56	4.34	4.01	2.90	1.00	1.25	4.07	5.09	5.48	5.69	5.28	2.48			
Gracia Lopez	ε(P)	-0.99	-0.95	-0.91	-0.87	-0.74	-0.55	-0.44	-0.35	-0.21	-0.16	-0.13	-0.17	-0.28			
	STD	0.47	0.43	0.39	0.35	0.22	0.03	0.08	0.17	0.31	0.36	0.39	0.35	0.24	-1.16	1.16	1.64
	RMS	5.53	6.10	6.31	6.34	5.81	4.74	4.05	3.44	2.30	1.98	1.78	2.56	5.36			
Karasawa	ε(P)	-0.89	-0.83	-0.77	-0.72	-0.61	-0.49	-0.39	-0.29	-0.09	0.07	0.41	0.80	0.97			
	STD	0.67	0.61	0.55	0.50	0.39	0.27	0.17	0.08	0.13	0.29	0.63	1.02	1.19	-0.53	0.81	0.97
	RMS	4.97	5.30	5.32	5.25	4.83	4.18	3.61	2.87	0.98	0.83	5.47	11.76	18.45			
Proposed	ε(P)	-0.85	-0.79	-0.73	-0.67	-0.55	-0.40	-0.29	-0.19	0.00	0.14	0.24	0.28	0.19			
	STD	0.57	0.51	0.45	0.39	0.27	0.12	0.01	0.09	0.28	0.42	0.52	0.55	0.47	-0.49	0.64	0.81
	RMS	4.74	5.05	5.03	4.91	4.34	3.45	2.69	1.83	0.00	1.72	3.25	4.06	3.67			

Table 3: Calculated error percentage [0.005% to 0.5%] for 12.25 GHz.

Procedures	Matrices	Occurrence							ITU-R P.311-15		
		0.5	0.2	0.1	0.05	0.02	0.01	0.005	μ_v	σ_v	ρ_v
ITU-R P. 618-13	ε(P)	-0.59	-0.38	-0.35	-0.22	0.04	-0.11	-0.14	-0.15	0.19	0.24
	STD	0.46	0.24	0.22	0.08	0.18	0.03	0.01			
	RMS	1.77	1.32	1.75	1.30	0.31	1.16	2.17			
DAH	ε(P)	-0.66	-0.48	-0.46	-0.34	-0.12	-0.24	-0.27	-0.23	0.22	0.32
	STD	0.46	0.28	0.26	0.14	0.08	0.04	0.07			
	RMS	3.94	2.85	5.21	4.19	0.66	6.94	16.30			
Ramachandran and Kumar	ε(P)	-0.32	0.04	0.10	0.39	1.16	0.57	0.28	0.12	0.23	0.26
	STD	0.49	0.13	0.07	0.22	0.99	0.39	0.11			
	RMS	0.96	0.13	0.51	2.34	8.14	6.22	4.25			
Karasawa	ε(P)	-0.63	-0.42	-0.03	-0.26	-0.01	0.16	0.07	-0.1	0.24	0.26
	STD	0.55	0.33	0.06	0.18	0.08	0.25	0.16			
	RMS	1.90	1.45	0.13	1.58	0.07	1.77	1.08			
Proposed	ε(P)	-0.47	-0.20	-0.17	0.00	0.31	0.12	0.06	-0.03	0.17	0.17
	STD	0.44	0.17	0.14	0.02	0.34	0.14	0.09			
	RMS	1.41	0.70	0.85	0.03	2.20	1.29	0.91			

Table 4: Calculated error percentage over 0.1% and 0.01% of the time.

Methods	Parameters	Time Percentage		ITU-R P.311-15		
		0.1	0.01	μ_v	σ_v	ρ_v
Unified	ε(P)	-0.09	-0.14	-0.11	0.04	0.12
	STD	0.03	0.03			
	RMS	0.43	1.52			
SAM	ε(P)	0.07	0.36	0.18	0.12	0.22
	STD	0.14	0.14			
	RMS	0.36	3.92			
Crane Global	ε(P)	-0.05	0.14	0.04	0.09	0.10
	STD	0.10	0.10			
	RMS	0.26	1.51			
Gracia Lopez	ε(P)	-0.15	-0.15	-0.15	0.01	0.15
	STD	0.00	0.00			
	RMS	0.15	0.15			

the model is judged through the relative error analyses, standard deviation, root mean square values, and ITU-R P. 311-15. Thus, the paper presents comparison of measured data with the existing rain attenuation prediction models and categorizes the best fitting models. Overall, ITU-R P. 618-13 shows better applicability for prediction of rain attenuation until the sufficient database from other locations become available. ITU-R P. 618-13 model shows better estimation and could be used until and unless sufficient dataset of rain attenuation measurement at Ku-band is made available from other location of South Korea. However, as per the rain attenuation and rain rate measurement obtained from two specific locations, the proposed method shows better estimation. This emphasizes the advantages of the proposed approach as compared to existing models.

The empirical result generated would be the helpful tool for system designers to determine the link margin at the specific site. However, additional tests and experimental data are necessary for better understanding of the present line of study for Ku-band satellite communication link. The contribution describes some preliminary steps aiming at devising appropriate methodology for prediction of rain attenuation affecting earth-space communication link.

Data Availability

It could be provided upon request from the reader on the approval from National Radio Research Agency (RRA).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We want to extend our thankfulness towards National Radio Research Agency (RRA) for providing and supporting us with the valuable database of satellite system. Authors also like to thank School of Engineering, Macquarie University, NSW, Australia, for providing the environment to further carry out the research work.

References

[1] R. K. Crane, *Electromagnetic Wave Propagation through Rain*, John Wiley and Sons Series, 1996.

[2] Z. B. Hasanuddin, K. Fujisaki, K. Ishida, and M. Tateiba, "Measurement of Ku-band rain attenuation using several VSATs in Kyushu Island, Japan," *IEEE Antennas and Wireless Propagation Letters*, vol. 1, no. 1, pp. 116–119, 2002.

[3] J. E. Allnutt and F. Haidara, "Ku-band diurnal fade characteristics and fade event duration data from three, two-year, Earth–space radiometric experiments in Equatorial Africa," *International Journal of Satellite Communications and Networking*, vol. 18, no. 3, pp. 161–183, 2000.

[4] L. J. Ippolito Jr., *Satellite Communications Systems Engineering: Atmospheric Effects, Satellite Link Design and System Performance*, vol. 6, John Wiley and Sons, 2008.

[5] L. J. Ippolito Jr., "Rain Attenuation Prediction Methods," in *Radiowave Propagation in Satellite Communications*, pp. 64–92, Springer, Netherlands, 1986.

[6] A. Dissanayake, J. Allnutt, and F. Haidara, "A prediction model that combines rain attenuation and other propagation impairments along earth-satellite paths," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 10, pp. 1546–1558, 1997.

[7] J. S. Mandeep and K. Tanaka, "Effect of atmospheric parameters on satellite link," *International Journal of Infrared and Millimeter Waves*, vol. 28, no. 10, pp. 789–795, 2007.

[8] ITU-R P. 618-13, "Propagation data and prediction methods required for the design of Earth-space telecommunication systems," 2017.

[9] S. Shrestha, J.-J. Park, and D.-Y. Choi, "Rain rate modeling of 1-min from various integration times in South Korea," *SpringerPlus*, vol. 5, no. 1, 2016.

[10] S. Shrestha, J.-J. Park, S. W. Kim, J. J. Kim, J. H. Jung, and D. Y. Choi, "1-minute rain rate derivation from various integration times in South Korea," in *Proceedings of the 1st International*

Conference on Next Generation Computing, Korean Institute of Next Generation Computing, Bangkok, Thailand, January 2016.

[11] S. Shrestha and D.-Y. Choi, "Proposed one-minute rain rate conversion method for microwave applications in South Korea," *Journal of Information and Communication Convergence Engineering*, vol. 14, no. 3, pp. 153–162, 2016.

[12] S. Shrestha and D.-Y. Choi, "Study of 1-min rain rate integration statistic in South Korea," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 155, pp. 1–11, 2017.

[13] D. Y. Choi, "Rain attenuation prediction model by using the 1-hour rain rate without 1-minute rain rate conversion," *Journal of Computer Science and Network Security*, vol. 6, pp. 130–133, 2006.

[14] D. Y. Choi, J. Y. Pyun, S. K. Noh, and S. W. Lee, "Comparison of measured rain attenuation in the 12.25 GHz band with predictions by the ITU-R model," *International Journal of Antennas and Propagation*, vol. 2012, Article ID 415398, 5 pages, 2012.

[15] S. Shrestha and D.-Y. Choi, "Study of rain attenuation in Ka band for satellite communication in South Korea," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 148, pp. 53–63, 2016.

[16] A. I. Yussuff and N. H. H. Khamis, "Rain attenuation prediction model for lagos at millimeter wave bands," *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 3, pp. 639–646, 2014.

[17] A. I. Yussuff and N. H. Khamis, "Rain attenuation modelling and mitigation in the tropics: brief review," *International Journal of Electrical and Computer Engineering*, vol. 2, no. 6, article 748, 2012.

[18] J. S. Mandeep, R. Nalinggam, and W. B. Ismail, "Analysis of rain attenuation models for South East Asia countries," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 32, no. 2, pp. 233–240, 2011.

[19] K. Chakravarty and A. Maitra, "Rain attenuation studies over an earth-space path at a tropical location," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 72, no. 1, pp. 135–138, 2010.

[20] J. S. Mandeep, "Slant path rain attenuation comparison of prediction models for satellite applications in Malaysia," *Journal of Geophysical Research: Atmospheres*, vol. 114, no. D17, 2009.

[21] J. S. Mandeep, "Comparison of rainfall models with Ku-band beacon measurement," *Acta Astronautica*, vol. 64, no. 2, pp. 264–271, 2009.

[22] J. S. Mandeep, S. I. S. Hassan, and K. Tanaka, "Rainfall measurements at Ku-band satellite link in Penang, Malaysia," *IET Microwaves, Antennas & Propagation*, vol. 2, no. 2, pp. 147–151, 2008.

[23] S. Mandeep and J. E. Allnutt, "Rain attenuation predictions at Ku-band in South East Asia countries," *Progress in Electromagnetics Research*, vol. 76, pp. 65–74, 2007.

[24] A. K. Lwas, M. R. Islam, M. H. Habaebi, S. J. Mandeep, A. F. Ismail, and A. Zyoud, "Effects of wind velocity on slant path rain-attenuation for satellite application in Malaysia," *Acta Astronautica*, vol. 117, pp. 402–407, 2015.

[25] F. A. Semire, R. Mohd-Mokhtar, W. Ismail, N. Mohamad, and J. S. Mandeep, "Evaluation of site diversity rain attenuation mitigation technique in South-East Asia," *Acta Astronautica*, vol. 96, pp. 303–312, 2014.

[26] J. M. Garcia-Rubia, J. M. Riera, P. Garcia-Del-Pino, G. A. Siles, and A. Benarroch, "Experimental assessment of slant-path rain attenuation variability in the Ka-band," *International Journal of Satellite Communications and Networking*, vol. 34, no. 2, pp. 155–170, 2016.

[27] D. Das and A. Maitra, "Rain attenuation prediction during rain events in different climatic regions," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 128, pp. 1–7, 2015.

[28] S. Shrestha and D.-Y. Choi, "Diurnal and monthly variations of rain rate and rain attenuation on Ka-band satellite communication in South Korea," *Progress in Electromagnetics Research B*, vol. 80, pp. 151–171, 2018.

[29] L. Mello and M. S. Pontes, "Unified method for the prediction of rain attenuation in satellite and terrestrial links," *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, vol. 11, no. 1, pp. 1–14, 2012.

[30] W. L. Stutzman and K. M. Yon, "Simple rain attenuation model for earth-space radio links operating at 10-35 GHz," *Radio Science*, vol. 21, no. 1, pp. 65–72, 1986.

[31] R. K. Crane, "Prediction of Attenuation by Rain," *IEEE Transactions on Communications*, vol. 28, no. 9, pp. 1717–1733, 1980.

[32] V. Ramachandran and V. Kumar, "Modified rain attenuation model for tropical regions for Ku-Band signals," *International Journal of Satellite Communications and Networking*, vol. 25, no. 1, pp. 53–67, 2007.

[33] J. A. Garcia-Lopez, J. M. Hernando, and J. M. Selga, "Simple rain attenuation prediction method for satellite radio links," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 3, pp. 444–448, 1988.

[34] Y. Karasawa, "Consideration on prediction methods or rain attenuation on earth-space paths," Japan, April 1989, CCIRIWP 5.2, pp. 65-72, 1989.

[35] J. X. Yeo, Y. H. Lee, and J. T. Ong, "Rain attenuation prediction model for satellite communications in tropical regions," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 11, pp. 5775–5781, 2014.

[36] National Radio Research Agency (RRA) 767, Bitgaram-ro, Najusi, Jeollanam-do 58217, Republic of Korea, http://www.rra.go.kr/en/index.jsp.

[37] The mathworks, Inc. Protected by U.S. and international patents, http://www.mathworks.com.

[38] ITU-R, "Specific attenuation model for rain for use in prediction methods," Recommendation P.838-3, ITU-R Recommendations, P Series, ITU, Geneva, Intemational Telecommunications Union, 2005.

[39] ITU-R, "Rain Height Model for Prediction Methods," Recommendation P.839-4, ITU-R Recommendations, P Series, Geneva, International Telecommunications Union, 2013.

[40] ITU-R Databank DBSG3, http://www.itu.int/pub/R-SOFT-SG3/en.

[41] http://www.ott.com.

[42] OTT, "Operating instructions: Present Weather Sensor Parsivel," 70.200.005.B.E 08-1008.

[43] S. Shrestha and D.-Y. Choi, "Characterization of rain specific attenuation and frequency scaling method for satellite communication in South Korea," *International Journal of Antennas and Propagation*, vol. 2017, Article ID 8694748, 16 pages, 2017.

[44] Korea Meteorological Administration (KMA), http://web.kma.go.kr/eng/index.jsp.

[45] S. Shrestha and D.-Y. Choi, "Rain attenuation over terrestrial microwave links in South Korea," *IET Microwaves, Antennas & Propagation*, vol. 11, no. 7, pp. 1031–1039, 2017.

[46] S. Shrestha, J.-J. Lee, S.-W. Kim, and D.-Y. Choi, "Rain attenuation over terrestrial microwave links at 18 GHz as compared with prediction by ITU-R model," *Journal of Information and*

Communication Convergence Engineering, vol. 15, no. 3, pp. 143–150, 2017.

[47] S. Shrestha and D.-Y. Choi, "Rain attenuation statistics over millimeter wave bands in South Korea," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 152, pp. 1–10, 2017.

[48] ITU-R P. 311-15, "Acquisition, presentation and analysis of data in studies of radiowave propagation," International Telecommunication Union, Geneva, Switzerland, 2013.

Hindawi
Advances in Astronomy
Volume 2019, Article ID 5712682, 6 pages
https://doi.org/10.1155/2019/5712682

Research Article

The XinJiang Astronomical Observatory NSRT Pulsar Data Archive

Hailong Zhang,[1,2] Markus Demleitner,[3] Jie Wang,[1] Na Wang,[1,2] Jun Nie,[1,2] and Jianping Yuan[1,2]

[1]*Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011, China*
[2]*Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, West Beijing Road, Nanjing, Jiangsu 210008, China*
[3]*Heidelberg University, Zentrum für Astronomie, Mönchhof St. 12-14, 69120 Heidelberg, Germany*

Correspondence should be addressed to Hailong Zhang; zhanghailong@xao.ac.cn

Received 31 October 2018; Accepted 30 December 2018; Published 23 January 2019

Guest Editor: Junhui Fan

The sensitivity of our pulsar observation system with typical integration time of 16 minutes has reached 0.4 mJy after we have upgraded the receiver cooling system of NanShan 25-m radio telescope (NSRT) at Xinjiang Astronomical Observatory (XAO) in 2002. About 280 pulsars were observed after the upgrade. The NanShan station pulsar data is transmitted to the data center at XAO headquarters via a 300Mb/s dedicated data link for long-term storage, and the remote backup is realized by synchronous data transmission between NanShan station and XAO headquarters. Metadata extraction, archiving, and releasing are completed by the data release servers which are located in XAO headquarters. At present, more than 84000 pulsar observation data files in the PSRFITS format have been released with a maximum file size of 1GB. XAO pulsar data online service provides cone search and multiobject constraint query. We have published our data based on Virtual Observatory (VO) standards. The online pulsar data service is registered in the International Virtual Observatory Alliance's Registry. Users can access the data through common VO tools.

1. Introduction

A Pulsar [1] is a highly magnetized rotating neutron star that emits a beam of electromagnetic radiation. It is the product of massive star evolution and has the characteristics of small volume, large density, high-speed rotation, and strong magnetic field. Pulsar observation research is helpful to reveal the mystery of the birth and evolution of the universe and has very important scientific significance to promote the development of astrophysics, particle physics, and spacecraft navigation. It has always been a hot spot in the field of astronomy and astrophysics. Pulsar observation and theoretical research have made a rapid progress after Jocelyn Bell discovered the first radio pulsar in 1967 [2]. By October 2018, 2658 pulsars have been confirmed (http://www.atnf.csiro.au/research/pulsar/psrcat/).

The NanShan 25-m Radio Telescope [3] (latitude $43^\circ 28' 47''$ N, longitude $87^\circ 10' 42''$ E) has a major upgrade in 2014, after that the antenna diameter is now 26 meters. NSRT is an important member of VLBI network in China. It has carried out regular pulsar timing observation since Jan 2010. NSRT has a Declination range from -40° to 90° and became an important base for monitoring pulsar timing behavior. How to store, archive, and release the ever-increasing pulsar data is a vital issue in operating the system and explore the data scientific value.

Pulsar data taken by the Parkes Radio Telescope (https://www.parkes.atnf.csiro.au/) (Parkes) as far back as the early 1990s is archived for long-term storage in the CSIRO's Data Access Portal (DAP) (https://data.csiro.au/dap/), in CSIRO's data center, Canberra, Australia. The CSIRO DAP allows querying and download of pulsar data by any user; however some data are subject to an embargo period. In this case, members of a science team can log in with their account to access their data during the embargo period. Parkes pulsar data archive [4] can be visualized and manipulated with Virtual Observatory (VO) Services. The NSRT pulsar data retrieval service provides similar functionality to Parkes.

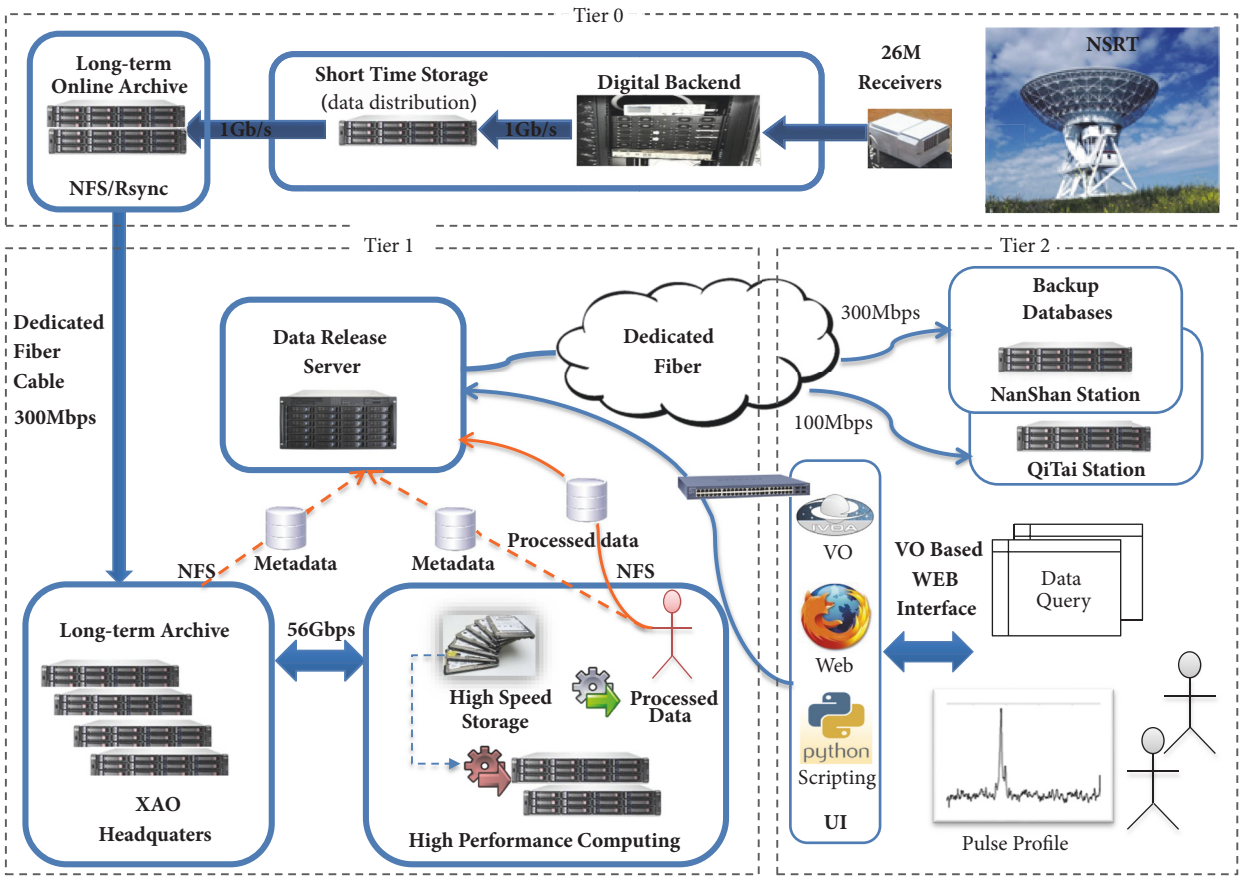


Figure 1: Overall design of NSRT pulsar data management system.

2. Pulsar Observing System

The pulsar de-dispersion observation system was set up at the end of 1999, which consists of a two-channel room-temperature receiver at L band, a down converter, a multichannel filterbank, and a digitizer. The receiver has a center frequency of 1540 MHz with bandwidth of 320 MHz, and the de-dispersion is implemented by a 2x128x2.5 MHz multichannel analog filterbank. Analogue signals from the filterbank are 1-bit digitized. After sampling, the raw data of each channel are folded at the apparent pulsar period to form subintegrations of 60 s which are stored in the "Timer" format [5]. The observation is started by 1-s pulse and the start time is checked using the 5-s pulse derived from the Observatory H-maser. A GPS time-transfer system is employed to align the Observatory clock with UTC. The pulsar timing program started since January 2000 and 74 strong pulsars [6] with flux density greater than 4mJy were detected.

In July 2002, an L band dual-polarization cryogenic receiver was used for pulsar observations. The noise temperature of the receiver was less than 10 K, the noise temperature of the system reached 20 K and 22 K, respectively, and the observation sensitivity reached 0.4 mJy [7] with integration time of 16 minutes. The sensitivity enables NSRT to observe approximately 280 pulsars.

In January 2010, the Digital FilterBank (DFB) system [8], which was developed by Australia Telescope National Facility (ATNF)(https://www.atnf.csiro.au/), was commissioned and put into operation. It has higher timing resolution which makes it detect 10 millisecond pulsars. The observation system has a calibration probe in the feed and receives orthogonal linear polarizations. A pulsed noise source is used to inject a linearly polarized calibration signal which is employed to calibrate the flux and polarization of pulsar. Absolute timing is provided by the Peripheral Component Interconnect (PCI) based Australia Telescope Distributed Clock (ATDC). The ATDC accepts a 5MHz maser signal which provides the base frequency reference for all of the DFB timing. The 1PPS signal, which is generated from the 5MHz input, is used to determine the absolute time. For each observation, the ephemeris from PSRCAT (http://www.atnf.csiro.au/people/pulsar/psrcat/) is used online for folding the signal with integration time of 30 s, where the total integration time is 4 to 16 minutes with a cadence of about 10 days. The DFB data format is PSRFITS (a file format customized for pulsar data storage, based on FITS format), and PSRCHIVE [9] software is used to analyze offline data, including the RFI mitigation, de-dispersion, fold, and ephemeris update.

3. Overall Design of Pulsar Data Management System

The overall design of NSRT pulsar data management system is shown in Figure 1. The design consists of three tiers, where

tier 0 is responsible for pulsar observation data acquisition and online archiving, tier 1 manages remote backup of original observation data and the metadata extraction, and tier 2 handles the publication and query facilities of the pulsar data.

Tier 0. NSRT data acquisition and online archiving were completed in NanShan Station. Tier 0 consists of telescope, receiver, digital backend, data temporary storage, and archiving. Pulsar data recording and preprocessing happen entirely within the telescope receiving system. The data capture program will perform the data preprocessing, de-dispersion, pulse period calculation, period folding, data storage, observation outline query, graphic output, antenna attitude control, and so on in real time. The typical sampling rate in the observation is 1 ms, and the time of each observation is determined by pulsar's flow intensity in the observation frequency band; generally an observation lasts from 4 to 16 minutes. All data will be preprocessed and written into a temporary storage server and long-term archived after validity check.

Tier 1. Pulsar observation data is synchronized between the storage system of NanShan station and the XAO headquarters through a dedicated 300Mpbs line. Currently, rsync (https://rsync.samba.org/) is used to perform incremental backups from NanShan station to XAO headquarters. All the metadata will be imported into the corresponding database to prepare for data release after extracting the metadata information from pulsar data. The data storage server, the Taurus high performance computing system, and the data release server use a common NFS server to implement the data transmission. For database data, there are backups in XAO headquarters and NanShan station. Users are allowed to log in to Taurus analysis system to download and process the data. A 56 Gbps InfiniBand switch connects the Taurus HPC and the long-term archive in XAO headquarters. All the data processed by users can be archived and released as needed.

Tier 2. XAO uses GAVO's DaCHS [10] server as the main data release framework to implement metadata operations and the data release based on VO standards. Users are allowed to access the online data retrieval platform through web browsers, standard VO tools, various scripts, etc. At present, the data release system supports cone search and multiconstraint target retrieval based on source name, observation date, observation ID, etc. Users can customize the limited amount of data output, output fields, sorting, and other contents. Currently the pulsar data retrieval service supports HTML (data is returned in your web browser; you can select additional columns in your output from an input field that pops down when you mouse over it), VOTable (https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaVOTable) (data is returned in IVOA's standard data format, the VOTable), FITS Table (the data is in the first extension; this contains much less meta information than a VOTable of the same data and thus should only be used if your backend tools do not understand VOTables), CSV (Comma Separated Values: this format carries almost no metadata as well, but it is understood by many database programs, spreadsheets, etc.), tar (users can download all matching items in a tar file), and other data output formats.

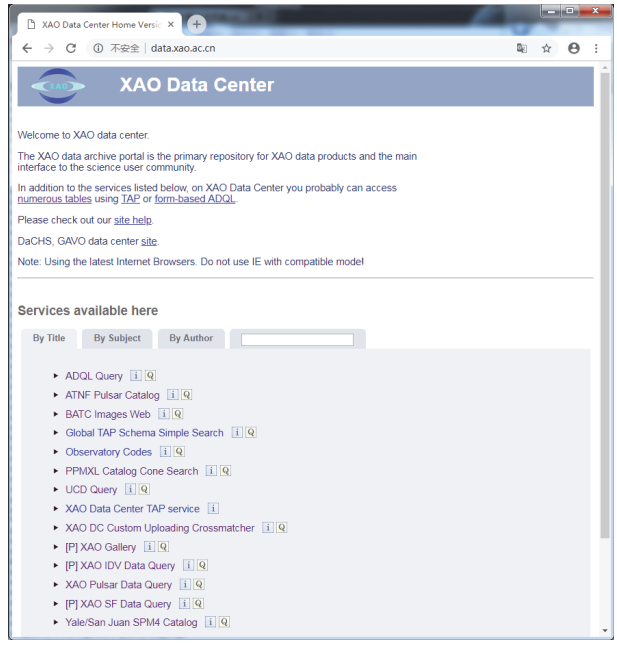


Figure 2: The main page of XAO data archive portal.

4. XAO Data Center and Pulsar Data Retrieval

The XAO data archive portal is the primary repository for XAO data products and the main interface to the science user community. In XAO Data Center (URL: http://data.xao.ac.cn), the interface is shown in Figure 2. Currently, several services such as ADQL (https://wiki.ivoa.net/internal/IVOA/ADQL/WD-ADQL-2.1-20171129.pdf) query, PPMXL [11] Catalog Cone Search, XAO DC Custom Uploading Crossmatcher, and XAO Pulsar Data Query, etc. have been released in the data center. The released data resources include Pulsar data, Active Galactic Nuclei data, and Molecular Line data.

4.1. Web Retrieval. The link of XAO pulsar data query page is http://data.xao.ac.cn/pul/pulsar/q/form. Users can also open the page through the "XAO Pulsar Data Query" link in the data center. A screenshot of the retrieval page is shown in Figure 3. At present, 84078 records of pulsar data have been archived with a maximum file size of 1GB. The data format is mainly PSRFITS, and subsequent data processing can be completed through PSRCHIVE.

The Query Interface:

(1) "Position/Name": "Position/Name": it is mainly used to input coordinates, the Right Ascension and Declination information of pulsars, or position information of SIMBAD-resolvable object, for example, RA ~ 278.4, Dec ~ -3.6, M31, etc.

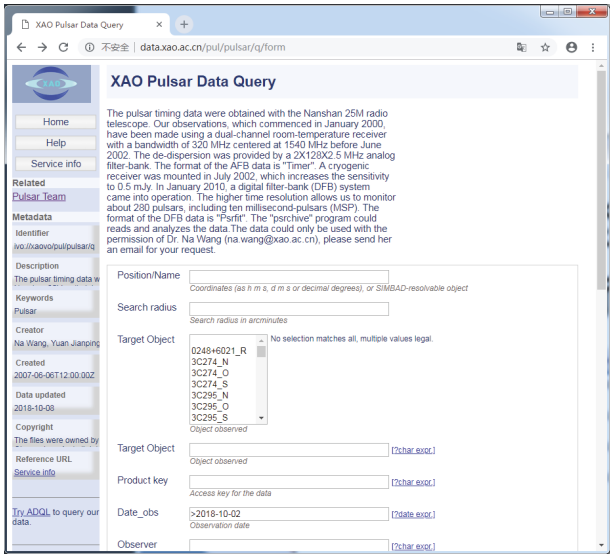


Figure 3: XAO pulsar data retrieval interface.

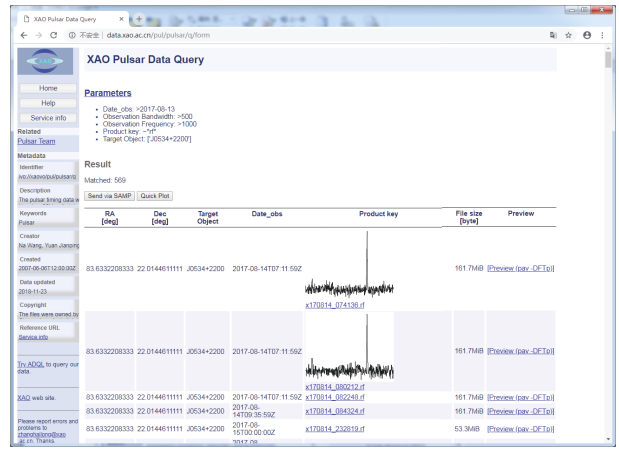


Figure 4: Observational data retrieval results of J0534+2200.

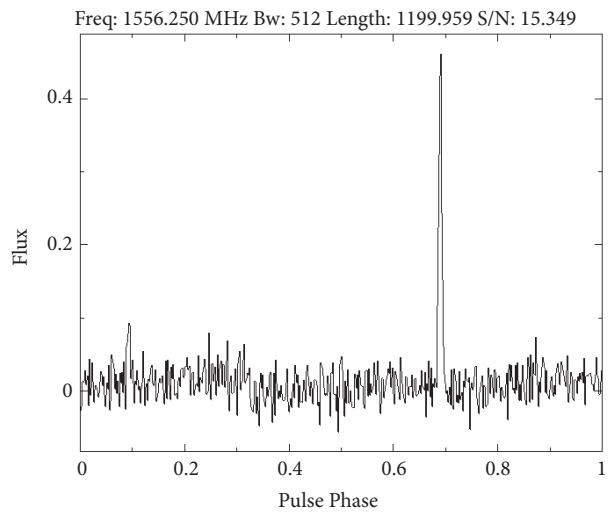


Figure 5: Pulse profile of PSR J0534+2200 (file name: x170814_074136.rf, the flux is not calibrated.)

(2) "Search radius": Search radius is according to the information in "Position/Name" and used to define the scope of data retrieval, for example, SR ~ 6; the unit of Search radius is arcminutes.

(3) "Target Object": "Target Object": it can let the user directly select or enter a pulsar name in the input form, for example, Target Object ~ J2157+4017, or ~ ∗j2157∗, and other related information; the ~ sign means contains and ∗ sign stands for all characters. The _R in target objects represents calibration file.

(4) "Date_obs": observation time. Enter "> year-month-day" which means query the data after a certain day of a certain year. For example, "<2018-09-20" means select values earlier than September 20th, 2018; "year-month-day +/- num" means that select values num days around year-month-day. For example, "2018-09-20 +/- 3" means select 6 days from September 17th, 2018, to September 23rd, 2018.

(5) "Observation Frequency": observation frequency is an important factor affecting observation of pulsars, which has certain influence on pulsar's dispersion effect, scattering effect and spectrum, and can be filled according to actual needs; the value type is numeric.

(6) "Output format": "Output format" includes HTML (default), Text (fixed columns), tar, Text (with Tabs), JSON, VOTable, FITS table, and CSV.

Taking J0534+2200 as an example, 569 items of relevant data were obtained with the conditions: observation frequency greater than 1000MHz, observation bandwidth greater than 500 MHz, and the Date-obs after August 13th, 2017; the file name includes "∗.rf". For example, the pulsar data is on August 14th, 2017, the recording time is 2017-08-140T07:11:59Z, and the data is named "x170814_074136.rf". The search results are shown in Figure 4. Click "[Preview (pav -DFTp)]" to view pulsar pulse profile (Figure 5).

4.2. VO Tools Retrieval. The VO Registry provides a mechanism through which VO applications can discover and select resources related to specific scientific issues. The XAO has completed the registration of pulsar data services.

(1) TOPCAT (http://www.star.bris.ac.uk/~mbt/topcat/). TOPCAT is an interactive graphic editor for tabular data, which is convenient for astronomers to analyze astronomical data. Thanks to the use of VO standards, it can work smoothly with other tools, service systems and data sets inside and outside VO. It has been widely used in Starlink, AstroGrid, VOTech, AIDA, GAVO, GENIUS, DPAC, and other projects.

(2) TOPCAT Retrieval. Start TOPCAT and searching for the keyword "xaovo" through Cone Search in VO menu can find 4 XAO registered services including XAO Pulsar Data Query, PPMXL Catalog Cone Search, etc.

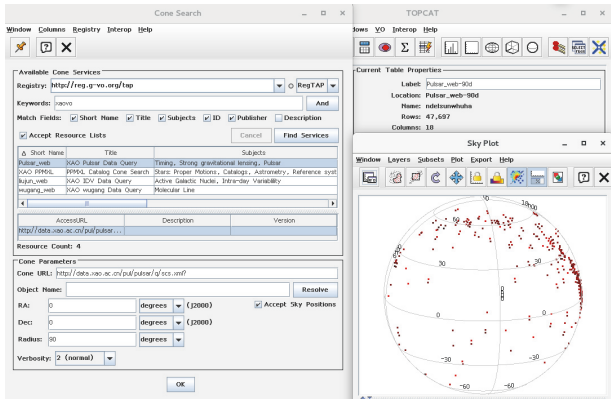


Figure 6: VO tool retrieval results.

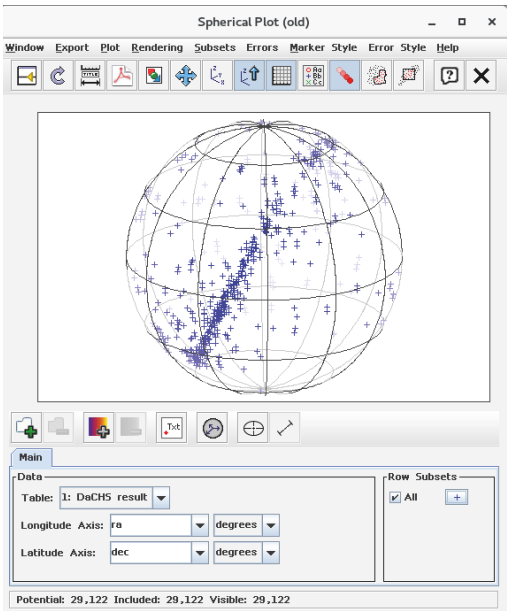


Figure 7: TOPCAT visualization results.

As shown in Figure 6, this example selects XAO Pulsar Data Query service to automatically generate Cone URL "http://data.xao.ac.cn/pul/pulsar/q/scs.xml?" to retrieve RA~0, Dec~0, Radius~90, Verbosity~3(maximum). 47697 items of relevant data can be retrieved (the left panel on Figure 6). Among them, "J2000" with the year 2000 epochs (named after the letter "B" with year 1950 epochs).

4.3. Pulsar Data Visualization. Through TOPCAT, the retrieved pulsar data can be visualized: turn on TOPCAT and select "Send via SAMP" in pulsar retrieval result interface to transfer the data to TOPCAT for visualization. In this example, TOPCAT version 4.6-1 was used to visualize 29122 results with Date-obs: < 2014-02-01 and Product key: "~ ∗.rf" (PULSAR data format generated by DFB3 system after January 2010); as shown in Figure 7, the data is marked in 3D galactic coordinates by Spherical Plot.

4.4. Pulsar Data Access. Since the launch of the pulsar data release server in 2016, the data access statistical information is shown in Figure 8. The left part of Figure 8 shows the total number of accesses with an average of about 10,000 accesses per month, and the right part of Figure 8 shows the number of hosts accessed with an average of nearly 500 hosts accessing the data server per month.

5. Conclusion

The XAO data archiving and release system is available to worldwide researchers. Data release is based on VO protocols. Users can retrieve and access data through browsers, various scripts, and standard VO tools. The pulsar data service has been registered to the IVOA.

Data Availability

All the pulsar data generated by XAO 25m radio telescope have been released in XAO data center, and the way to query the data was described in the article. The address of our data center is http://data.xao.ac.cn. Astronomers can repeat our research or query easily.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the support of the National Natural Science Fund of China (11873082, U1531125, 11803080, and 11503075), National Key Basic Research Program of China, 973 Program 2015CB857100, National Key Basic Research and Development Program 2018YFA0404704, and Youth Innovation Promotion Association CAS. The research is partly supported by the Operation, Maintenance and Upgrading Fund for Astronomical Telescopes and Facility Instruments, budgeted from the Ministry of Finance of China (MOF) and administrated by the Chinese Academy of Sciences (CAS). The algorithm and debugging work has

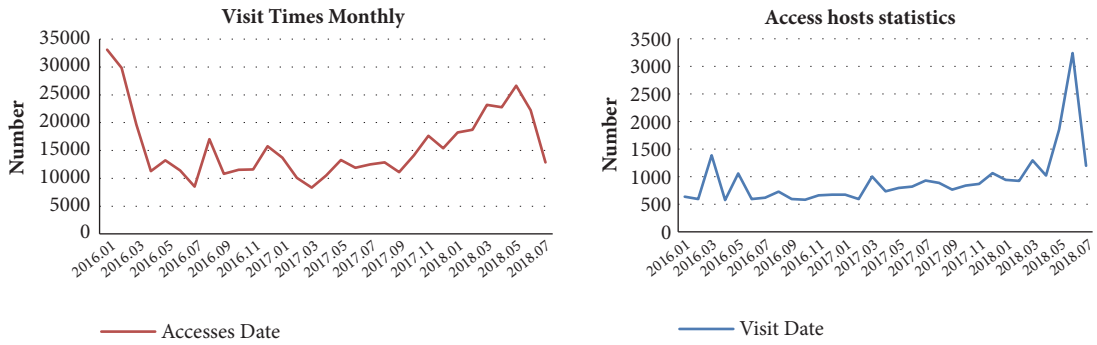


Figure 8: NSRT pulsar data service accessing statistics.

applied Taurus High Performance Computing Cluster of Xinjiang Astronomical Observatory, CAS.

References

[1] R. D. Lorimer and M. Kramer, *Handbook of Pulsar Astronomy*, Cambridge university press, 2005.

[2] Z. Weizhen, *Observation and Study of Pulsar Timing and Glitch*, Graduate University of the Chinese Academy of Sciences, 2009.

[3] Z. Hailong, Z. Yan, N. Jun et al., "Xinjiang astronomical observatory NSRT data storage system," *Astronomical Research & Technology*, vol. 15, no. 2, pp. 181–187, 2018.

[4] G. Hobbs, D. Miller, R. N. Manchester et al., "The parkes observatory pulsar data archive," *Publications of the Astronomical Society of Australia*, vol. 28, no. 3, pp. 202–214, 2011.

[5] N. Wang, R. N. Manchester, J. Zhang et al., "Pulsar timing at urumqi astronomical observatory: observing system and results," *Monthly Notices of the Royal Astronomical Society*, vol. 328, no. 3, pp. 855–866, 2001.

[6] J. Zhang and N. Wang, "Single dish astrophysics observation at urumqi observatory," *Progress of Astronnomy*, vol. 18, no. 4, pp. 271–282, 2000.

[7] N. Wang, *Observation and Study of Pulsar Timing and Interstellar Scintillation*, Peking University, 2001.

[8] G. Hampson and B. Andrew, *A 1GHz Pulsar Digital Filter Bank and RFI Mitigation System*, 2008, http://www.jb.man.ac.uk/pulsar/observing/DFB.pdf.

[9] A. W. Hotan, W. Van Straten, and R. N. Manchester, "PSRCHIVE and PSRFITS: An open approach to radio pulsar data storage and analysis," *PASA - Publications of the Astronomical Society of Australia*, vol. 21, no. 3, pp. 302–309, 2004.

[10] M. Demleitner, M. C. Neves, F. Rothmaier, and J. Wambsganss, "Virtual observatory publishing with DaCHS," *Astronomy and Computing*, vol. 7-8, pp. 27–36, 2014.

[11] S. Roeser, M. Demleitner, and E. Schilbach, "The ppmxl catalog of positions and proper motions on the ICRS. Combining USNO-B1.0 and the two micron all sky survey (2MASS)," *The Astronomical Journal*, vol. 139, no. 6, pp. 2440–2447, 2010.