A Hierarchical Method for Locating the Interferometric Fringes of Celestial Sources in the Visibility Data (Postprint)

Authors: Rong Ma, Ruiqing Yan, Hanshuai Cui, Xiaochun Cheng, Jixia Li, Fengquan Wu, Zongyao Yin, Hao Wang, Wenyi Zeng and Xianchuan Yu

Date: 2024-03-29T00:00:00+00:00

Abstract

In source detection in the Tianlai project, locating the interferometric fringe in visibility data accurately will influence downstream tasks drastically, such as physical parameter estimation and weak source exploration. Considering that traditional locating methods are time-consuming and supervised methods require a great quantity of expensive labeled data, in this paper, we first investigate characteristics of interferometric fringes in the simulation and real scenario separately, and integrate an almost parameter-free unsupervised clustering method and seeding filling or eraser algorithm to propose a hierarchical plug and play method to improve location accuracy. Then, we apply our method to locate single and multiple sources' interferometric fringes in simulation data. Next, we apply our method to real data taken from the Tianlai radio telescope array. Finally, we compare with unsupervised methods that are state of the art. These results show that our method has robustness in different scenarios and can improve location measurement accuracy effectively.

Full Text

Preamble

A Hierarchical Method for Locating the Interferometric Fringes of Celestial Sources in the Visibility Data

Rong Ma[1], Ruiqing Yan[1], Hanshuai Cui[1], Xiaochun Cheng[2], Jixia Li[3], Fengquan Wu[3], Zongyao Yin[1], Hao Wang[1], Wenyi Zeng[1], and Xianchuan Yu[1]

[1] School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China; yuxianchuan@163.com

[2] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Received 2023 June 1; revised 2023 October 8; accepted 2023 October 30; published 2024 February 26

Abstract

In source detection for the Tianlai project, accurate localization of interferometric fringes in visibility data critically influences downstream tasks such as physical parameter estimation and weak source exploration. Conventional localization methods are time-consuming, while supervised approaches require large quantities of expensive labeled data. This paper first investigates the characteristics of interferometric fringes in both simulated and real scenarios, then integrates an almost parameter-free unsupervised clustering method with a seed filling algorithm to propose a hierarchical plug-and-play method that improves localization accuracy. We apply our method to locate single and multiple source interferometric fringes in simulation data, and subsequently to real data from the Tianlai radio telescope array. Comparisons with state-of-the-art unsupervised methods demonstrate that our approach exhibits robustness across different scenarios and effectively improves localization measurement accuracy.

Key words: methods: data analysis –techniques: image processing –techniques: interferometric

1. Introduction

Dark energy detection represents a crucial topic in cosmology (Korytov et al. 2019; Tanoglidis et al. 2021; Everett et al. 2022). As radio telescope technology achieves increasingly higher resolution and sensitivity, astronomers can observe the universe across broader frequency ranges. The Tianlai project (Chen 2011, 2012) aims to explore large-scale structure by measuring the redshifted 21 cm emission line of neutral hydrogen. Located at the Hongliuxia Observing Station in northeastern Xinjiang province, China, Tianlai records source signals as visibility data. Fringes can be obtained directly from raw visibility data by extracting either the amplitude or phase component of the complex values.

We treat the phase component of visibility data within specific time intervals and the 700–800 MHz frequency band as a single image because the phase is typically more sensitive than amplitude. Through calibration and data analysis, specific source parameters and sky maps can be recovered (Li et al. 2020; Zuo et al. 2021). A critical step in this pipeline is locating interferometric fringes of sources in raw visibility data. These images generally contain fringes at various low signal-to-noise ratios (S/N). High-accuracy fringe localization directly determines the precision of downstream physical parameter estimation, yet developing effective measurement methods for these weak signal processing tasks remains challenging.

Historically, interferometric fringe localization and physical parameter estimation relied on technician and researcher experience, consuming substantial labor while achieving low efficiency and accuracy. With artificial intelligence (AI) development, AI-based methods have demonstrated remarkable accuracy in weak signal processing, while high-performance computing has drastically reduced runtime. Consequently, we consider applying AI methods to fringe localization in raw visibility data.

Recent years have seen AI methods provide effective tools for astronomical problems. Traditional machine learning and deep learning methods address both small- and large-scale scenarios. Cavaglia et al. (2018) proposed a random forest and genetic programming method for gravitational wave detection, while Gheller et al. (2018) utilized convolutional neural networks to detect extragalactic sources. Deep learning has achieved significant progress in image classification (He et al. 2016), segmentation (Shelhamer et al. 2017), object detection (Redmon et al. 2016), and signal detection. Awni et al. (2019) introduced deep convolutional neural networks for large-scale stellar spectra classification and supernova remnant candidate detection. Yan et al. (2022) applied channel attention shrinkage networks to weak source fringe detection. Wang et al. (2019) adopted ResNet for pulsar selection, and researchers have introduced transfer learning across various fields (Xu et al. 2015; He et al. 2022; Kuang et al. 2022). In practical applications, novel deep neural network architectures have been developed for multiple scenarios (Zeng et al. 2021; Fu et al. 2022; Lin et al. 2022; Peng et al. 2022).

However, supervised learning typically requires large quantities of labeled training data and computation. While high-performance computing has mitigated computational constraints, obtaining labeled data remains problematic, particularly for unsupervised and class-imbalanced scenarios. Novel solutions have been proposed for unsupervised industrial measurement scenarios (Cao et al. 2022; Fong & Narasimhan 2022; Zhu et al. 2022). Given the expense of labeled data, unsupervised methods are preferable for interferometric fringe localization.

Fuzzy clustering serves as a powerful tool in practical applications. Since Zadeh (1965) proposed fuzzy sets, numerous unsupervised methods have been developed based on this theory. Dunn (1973) introduced the fuzzy C-means (FCM) algorithm to replace hard clustering and improve accuracy. Krinidis & Chatzis (2010) constructed an objective function with a local fuzzy factor G_{ki} to enhance accuracy. Zeng et al. (2020) introduced hesitant fuzzy theory with the hesitant fuzzy C-means (HFCM) algorithm. Gong et al. (2013) merged kernel methods with local information through weights to propose the kernel metric weighted fuzzy C-means algorithm with local information (KWFLICM). Generalizing KWFLICM, Memon & Lee (2018) proposed neighbor searching methods for high-dimensional data. These effective fuzzy clustering variants have been successfully applied to image segmentation tasks.

While acquiring visibility data in simulation scenarios is straightforward through programming, real-scenario labeled data are expensive due to reliance on expert experience. To overcome this limitation, we employ unsupervised methods. Given the substantial uncertainty and noise in both simulated and real visibility data, fuzzy clustering is particularly suitable. Among variants, the KWFLICM algorithm demonstrates representative robustness and satisfactory accuracy in segmentation tasks.

Despite KWFLICM providing good segmentation results, other factors and noise still affect downstream physical parameter estimation. To mitigate these adverse influences, we further select representative regions as signals. This paper investigates image characteristics in simulated and real scenarios and proposes a hierarchical method for interferometric fringe localization. To distinguish fringe signals from background, we utilize KWFLICM for image segmentation, then apply a seed filling algorithm to remove noise influence while retaining most fringe features for downstream processing. We regard the width of the maximum connected region as the fringe width, thereby providing a complete plug-and-play solution with high accuracy for locating interferometric fringes in raw visibility data.

This paper is organized as follows: Section 2 reviews classical unsupervised clustering methods and advanced variants. Section 3 proposes our novel hierarchical fringe localization method. Section 4 presents experiments in simulated and real scenarios to validate effectiveness, including comparisons with state-of-the-art unsupervised methods. Section 5 concludes the paper.

2. Preliminaries

Since Dunn (1973) proposed FCM and Hathaway & Bezdek (2000) provided extensions, FCM has attracted considerable research interest. Fuzzy clustering offers effective image segmentation, and compared with deep learning methods, FCM and its variants can produce segmentation results iteratively, making them more efficient for scenarios with limited samples. In image segmentation, images are divided into regions based on features such as gray level, local information, and texture.

The classical FCM algorithm constructs objective function J_m through the sum of squared error functions:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ji}^{m} d_{ji}^{2}$$

where $X = \{x_1, x_2, \dots, x_N\}$ represents the dataset of N samples. In image segmentation, each sample is one-dimensional, though other scenarios may involve higher dimensions. The parameter c ($c \in [1, N]$) represents the number of clusters (different regions in segmentation). u_{ji} signifies the membership degree describing the degree to which sample x_i belongs to cluster v_j. m is an exponent factor ensuring algorithm convergence to an optimal value. d_{ji} describes the distance between pixel x_i and cluster center v_j. Algorithm termination is controlled by threshold ϵ when $\epsilon > \sum |V^{(b+1)} - V^{(b)}|$.

Objective function J_m can be solved via Lagrange multipliers, yielding segmentation results through the update equations. The membership degree from the b-th iteration is denoted as $u_{ji}^{(b+1)}$.

Gong et al. (2013) merged kernel methods with local information carried by neighboring pixels, enhancing outlier handling and improving FLICM robustness. This algorithm constructs a novel objective function with local similarity factor G'_{ki}:

$$G'_{ki} = \sum_{j \in N_i} \sum_{l=1}^{c} w_{ij}(1 - K(x_j, v_k))$$

where w_{ij} is the fuzzy factor describing the weight between center pixel i and neighbor pixel j. $K(x_j, v_k)$ represents distance based on a kernel function. This serves as a penalty term accelerating algorithm convergence.

3. Methodology

This section investigates fringe image characteristics and introduces the KWFLICM algorithm for processing. We treat this task as image segmentation in the first stage. After obtaining segmentation results, we further process them with a seed filling algorithm to locate interferometric fringes in the second stage. For scenarios with multiple fringes, we propose an eraser algorithm to iteratively find all fringe locations. Our hierarchical method consists of these two stages, detailed below.

3.1. Characteristics of Interferometric Fringe in Simulated and Real Scenarios

Images containing interferometric fringes differ between simulated and real scenarios. In simulation, fringes are obvious with minimal noise, making localization easier compared to real scenarios at the same S/N. Intuitively, fringe localization becomes more difficult at lower S/N, while higher S/N yields more explicit fringe shapes. In real scenarios, fringe signals and noise are typically interlaced, and in special cases, low-S/N fringes may disappear into the noise.

3.2. Kernel Metric Weighted Fuzzy C-means Algorithm with Local Information

For fringes with varying S/N, we adopt an unsupervised fuzzy clustering method that provides segmentation results with low computational cost. FCM is a popular clustering algorithm, and since its proposal by Dunn (1973), many variants have been developed. KWFLICM is representative, fusing kernel methods with local information through weights to achieve effective clustering.

Data can be difficult to handle in low dimensions, and projecting into higher dimensions is computationally expensive. The kernel method addresses this by reducing computation. Traditional fuzzy clustering algorithms applied to image segmentation often neglect local information, yielding suboptimal results. Pixel relationships critically influence final segmentation outcomes, as neighboring pixels typically carry similar information to central pixels. KWFLICM incorporates both local information and kernel methods, providing higher accuracy.

From the objective function, we compute partial derivatives with respect to membership degree u_{ki} and cluster centers v_k separately. Update equations for clusters and membership degrees are obtained by setting these derivatives to zero:

$$v_k^{(b+1)} = \frac{\sum_{i=1}^{N} u_{ki}^m K(x_i, v_k) x_i}{\sum_{i=1}^{N} u_{ki}^m K(x_i, v_k)}$$

$$u_{ki}^{(b+1)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{1-K(x_i,v_k)}{1-K(x_i,v_j)} \right)^{\frac{1}{m-1}}}$$

We process fringe images by first transforming color images to grayscale to reduce runtime. After initializing important parameters, the algorithm produces membership and center matrices according to image pixels, iterating until convergence using the update equations. Upon termination, the resulting image contains only two regions representing signals and noise (see proof in Gong et al. 2013). Algorithm 1 summarizes KWFLICM.

Algorithm 1. KWFLICM Algorithm—First Stage of Our Method

Input: Gray-level image with interferometric fringe.
Output: Image segmented into two regions.

Steps: 1. Set parameters: number of clusters c, fuzzy exponent factor m, window size N, and stopping threshold ϵ. 2. Randomly initialize c clusters and membership degree matrix U. 3. Set counter $b = 0$. 4. Compute fuzzy weight factor w_{ij} and update distance $V^{(b+1)}$ using the cluster update equation. 5. Update center matrix using Equation (6). 6. Update membership degree matrix $U^{(b+1)}$ using Equation (7). 7. If $\epsilon > \sum |V^{(b+1)} - V^{(b)}|$, terminate; otherwise set $b = b + 1$ and return to Step 4.

3.3. Maximum Connection Region Finding Algorithm

Segmentation results retain fringe features. To extract the most important feature, we adopt the seed filling algorithm to obtain the maximum connected region. Although signals and noise are separated, noise regions still influence results. We compute the area of each enclosed region Reg_s $(s = 1, 2, \ldots, m)$ by counting pixels in the segmentation result, where m represents the number of connected regions. Large-area regions typically contain signals with important features, while small-area regions are usually noise. We sort these regions by area in descending order and select the maximum connected region as the algorithm output. Algorithm 2 details the seed filling process.

Algorithm 2. Seed Filling Algorithm—Second Stage of Our Method

Input: Output image from Algorithm 1.
Output: Maximum connected region of segmentation result.

Steps: 1. Input image segmented by KWFLICM algorithm. 2. Compute areas of different connected regions Reg_s. 3. Sort all connected regions by area and select the largest region $MAX(Reg_s)$. 4. Output image containing only the maximum connected region.

3.4. Eraser Algorithm

Special cases may involve multiple interferometric fringes in one image. To address this, we propose the eraser algorithm (Algorithm 3), which operates iteratively. In each iteration, the algorithm finds and retains a maximum connected region, then erases all related regions within the width range of that maximum region until obtaining the specified number of maximum connected regions.

The parameter *num* controls algorithm termination. To determine this parameter, we count signal pixels by column and generate a statistical curve. After smoothing and peak detection, we treat the number of peaks as the *num* value.

Algorithm 3. Eraser Algorithm—For Multiple Interferometric Fringe Scenario

Input: Output image from Algorithm 1, number of interferometric fringes num, iteration counter $counter = 0$.
Output: Locations of num maximum connected regions.

Steps: 1. Input segmentation result from KWFLICM algorithm. 2. Find and retain maximum connected region using Algorithm 2. 3. Increment counter. 4. If $counter < num$, erase connected regions within the width range and return to Step 2; otherwise proceed to Step 5. 5. Terminate algorithm and compute locations of different fringes.

Example: Figure 1 [Figure 1: see original paper] illustrates this process. The images show the phase component of raw visibility, with horizontal axis representing time and vertical axis representing frequency (origin at bottom left). For simplicity, coordinates are omitted and image size is set to 224×224. Assuming three fringes exist, the algorithm completes three iterations. The first maximum connected region is found and retained, its width $((166, 0), (190, 223))$ is obtained, and all regions in this width range are erased (red boxes). The process repeats to find subsequent regions $((23,0),(62,223))$ and $((99,0),(127,223))$. After three iterations, three maximum connected regions are obtained and the algorithm terminates. This generalizes our method' s applicability. For single-fringe images, the eraser algorithm degenerates to the classical seed filling algorithm. Figure 2 [Figure 2: see original paper] shows the hierarchical method flowchart.

4. Experiments

This section describes simulation data production, validates our method across scenarios, and compares it with state-of-the-art unsupervised methods.

4.1. Simulation Data

Visibility for an antenna pair is expressed as (Thompson et al. 1991):

$$V = Ae^{j2\pi f \mathbf{B}\cdot\mathbf{n}_s} + N$$

where j is the imaginary unit, f is frequency, $\mathbf{B}$ is the baseline vector, and $\mathbf{n}_s$ is the direction vector from antenna to radio source, reflecting the beam pattern. The visibility amplitude A represents signal strength when a point source transits the antenna beam. In simulation, amplitude is typically modeled as a Gaussian function whose peak value is proportional to source brightness. Beyond pure source visibility, interferometer systems inevitably receive instrumental noise. In our simulation, we model noise with a normal distribution added to visibility as N in Equation (8). Real instrumental noise is more complex, including cross-couplings between adjacent feeds, system gain variations from temperature fluctuations, and transient radio interference from human activity. S/N is determined by the ratio of source peak value to instrumental noise standard deviation.

We implement computer programming based on Equation (8) to produce simulation data and label pairs. Our method' s results are compared with labels to compute accuracy. For real scenarios, we use Tianlai telescope array visibility data, employing experts for manual labeling due to the lack of ground truth.

Figure 3 [Figure 3: see original paper] shows randomly generated fringe images varying in S/N, baseline length, and source location. Some images have explicit fringes (g, e), while others have dim fringes (c, d). Table 1 provides detailed parameters.

Table 1 Baseline, Source Location, Fringe Length and S/N of Simulation Images

Simulation Image	Baseline Length (m)	Source Location (%)	Fringe Length (%)	S/N

4.2. Simulation Scenario

4.2.1. Single Interferometric Fringe We produce visibility images detailed in Table 1. Baseline Length represents antenna separation (10–50 m range). Source Location indicates position percentage relative to 224 time points. Fringe Length is the ratio of fringe width to 224 time points.

Original simulated images appear in Figure 3 [Figure 3: see original paper]. Low-S/N images (\$ 0.7)$includeFigures3(c), (d), and (f), whilehigh - S/Nimages($1.1) include Figures 3(e), (g), and (i). KWFLICM parameters are set as: membership exponent $m = 2$ (providing good results in most variants), threshold $\epsilon = 0.00001$ (commonly used to balance runtime and accuracy), window size $winSize = 3$ (3\$×\$3 window captures local information with acceptable runtime), and clusters $c = 2$ (representing signal and background) for both simulated and real scenarios. These nearly fixed parameters make our method convenient and effectively parameter-free.

Remark 1: After first-stage processing, segmentation results appear in Figure 4 [Figure 4: see original paper], dividing images into two regions. Most images show clear fringe shapes, though low-S/N cases yield weaker results. High S/N produces more explicit shapes and better segmentation (Figures 4(e, g)), while lower S/N still yields usable results (Figures 4(h, i)). Notably, Figures 4(a, b) show half-shaped signals at image borders, yet the algorithm still provides excellent segmentation. Even for the lowest-S/N image (Figure 4(c)), distinct results are obtained. For S/N < 0.5, fringes become unrecognizable, establishing 0.5 as the limiting threshold. For S/N values between 1.5 and 10, the algorithm demonstrates very high effectiveness. We conclude the first-stage algorithm is effective for all S/N visibility data containing interferometric fringes.

After obtaining segmentation results, we apply the seed filling algorithm to find maximum connected regions, shown in Figure 5 [Figure 5: see original paper].

Remark 2: Finding the maximum connected region aims to retain maximal fringe features. This region represents the most salient fringe characteristic, minimizing influence from other factors and noise. After identification, we use the maximum region' s width to represent the interferometric fringe width, thereby obtaining final fringe location.

Remark 3: Table 2 presents numerical experimental results. Accuracy is computed through predicted and real location widths:

$$\text{Accuracy}(I) = \frac{S(\text{PreLoc} \cap \text{RealLoc})}{S(\text{PreLoc} \cup \text{RealLoc})}$$

where $S(\cdot)$ denotes area from image pixels, $\text{PreLoc} \cap \text{RealLoc}$ and $\text{PreLoc} \cup \text{RealLoc}$ represent intersection and union of predicted and real location regions, respectively.

Table 2 Validation of Our Method with Simulation Data

Simulation Image	Prediction Location	Labeled Location	Accuracy (%)	Running Time (s)	Iterations

For low-S/N images (Figures 5(c, d, f)), effects are weaker, while high-S/N images (Figures 5(b, g)) yield better results. These numerical experiments align with intuition. Since all simulation images share the same dimensions, iteration runtime is comparable, with iteration count related to fringe characteristics and noise distribution. Some accuracies appear relatively low numerically (e.g., images (c, f) with lowest S/N). Though our method provides limited accuracy in these cases, it still yields effective fringe locations. The accuracy criterion—intersection-over-union ratio—magnifies small mismatches, as even slight time-axis misalignment causes significant area ratio drops. Thus, 47% accuracy for very weak S/N remains acceptable.

4.2.2. Multiple Interferometric Fringes In drift-scan observations, multiple nearby sources in right ascension can produce multiple fringes within a time interval. We explore our method' s generalization capability for such cases. Figure 6 [Figure 6: see original paper] shows segmentation and eraser algorithm results for multiple fringes, demonstrating effectiveness. When multiple fringes have similar S/N, our method yields explicit results. However, large S/N differences can cause high-S/N fringes to suppress low-S/N ones, potentially discarding weak fringes as noise.

The maximum number of effectively recognized fringes depends on image complexity. Non-overlapping fringes with similar S/N can all be recognized, with maximum number depending on source count. For overlapping fringes, recognition is restricted by overlap degree—some fringes combine into one, reducing recognized fringe count. Figure 7 [Figure 7: see original paper] illustrates this: image (a) with three low-S/N (0.5) fringes yields three recognized fringes, while image (c) with overlapping fringes recognizes them as a single fringe (maximum recognized fringes = 2). Though direct width measurement is impossible in overlap cases, our method provides a basis for further processing, such as developing unmixing methods for special downstream parameter estimation tasks. For non-overlapped fringes, our method still provides explicit results.

4.3. Real Scenario

We validate our method using real interferometric fringe images from the Tianlai Cylinder telescope array. Real data contain more noise sources: radio frequency interference (RFI) from the environment and cross-coupling effects in dense radio arrays. These pollutants range from slight to serious, impacting fringe recognition differently. We discuss both cases separately.

4.3.1. Slightly Polluted Images Figure 8 [Figure 8: see original paper] shows real images with varying S/N. Some have obvious fringes (a, c, e), others have weak fringes (b, d). Compared to simulations, noise is more pronounced, affecting localization effectiveness. However, these images are only slightly polluted by RFI and cross-coupling, allowing direct application of our method. First-stage segmentation results appear in Figure 9 [Figure 9: see original paper], and second-stage localization results in Figure 10 [Figure 10: see original paper].

Remark 4: Segmentation results differ from Figure 4 [Figure 4: see original paper]. While fringes are clearly distinguished in simulation results (Figure 4), real-scenario results (Figure 9) show less clear fringe distinction due to additional noise factors. However, after maximum connection region processing, fringe locations are successfully detected (Figure 10). Thus, our method demonstrates good effectiveness and robustness for slightly polluted images in both simulated and real scenarios without requiring pre-processing (numerical results in Table 3). Real scenarios consume more runtime and iterations, reflecting greater image complexity.

Table 3 Validation of Our Method with Real Data

Real Image	Labeled Location	Detected Location	Accuracy (%)	Running Time (s)	Iterations

Remark 5: Both scenarios allow intuitive assessment of localization effectiveness. Simulation scenarios provide defined accuracy metrics showing numeric differences across images. Our method achieves good localization results ($ 70 $80%).

4.3.2. Seriously Polluted Images Actual data processing sometimes encounters very strong RFI, generating fake fringes that affect weak signal detection. Cross-coupling effects also behave stronger for short baselines. We conduct experiments on data polluted by strong RFI and cross-coupling, with results in Figure 11 [Figure 11: see original paper].

Figures (a1) and (b1) show raw phase images with strong cross-coupling effects (stable horizontal phases throughout time intervals) that bury interference fringes. However, cross-couplings are temporally stable, allowing removal via time-averaged phase subtraction (Figures a2, b2). Additionally, serious narrow-band RFI intermittently appears in the 760–788 MHz frequency range (top portions of a1, b1), breaking source fringe continuity in frequency direction, especially for weak sources. This polluted frequency range must be removed before processing (Figures a3, b3). Finally, our method completes fringe recognition (Figures a4, b4).

These results demonstrate that pre-processing—narrow-band RFI and cross-coupling removal—helps reduce noise influence for seriously polluted images. After pre-processing, our method handles fringes easily. Thus, for seriously polluted images, pre-processing is required before fringe detection, after which the problem reduces to the slightly polluted case. Since most real images suffer various interference types, pre-processing is generally necessary before applying our method.

4.4. Comparison with State-of-the-Art Methods

We compare our method with state-of-the-art unsupervised methods to illustrate adaptability and validation. Deep fuzzy k-means (DFKM) (Zhang et al. 2020) and unsupervised image segmentation by backpropagation (UnProp) (Kanezaki 2018) have shown good performance in other image processing fields. Figure 12 [Figure 12: see original paper] compares these methods on simulated image Figure 3(e). Resulting widths are: our method $[(19, 0), (56, 223)]$, DFKM $[(28, 0), (55, 223)]$, and UnProp $[(30, 0), (54, 223)]$. With labels $[(28, 0), (62, 223)]$, accuracies are 79.4%, 70.6%, and 65.1%, respectively, showing our method' s numeric advantage.

Comparison with Classical Image Segmentation Methods

Image segmentation is critical for interferometric fringe detection. Figure 13 [Figure 13: see original paper] compares our first-stage method with classical approaches on Figure 3(f): (a) Otsu method (threshold-based), (b) Canny method (edge-based), and (c) Watershed method (region-based). Our method (d) retains more fringe shape features. Otsu and Canny provide limited results where fringe shapes are hardly recognizable. Watershed yields rectangular regions lacking shape features, causing confusion in subsequent processing. Thus, compared to classical methods, our method provides a superior basis for further image processing.

5. Conclusion

For radio telescope arrays like Tianlai, effectively locating interferometric fringes in massive visibility data with high efficiency remains challenging. This paper investigates fringe characteristics across scenarios and proposes a hierarchical localization method. The first stage treats the task as image segmentation using unsupervised clustering; the second stage employs seed filling to find maximum connected regions, whose location represents the source' s interferometric fringe. We validate our method on real scenarios with slight and serious RFI/cross-coupling pollution and demonstrate superiority over state-of-the-art methods.

Future work will focus on distinguishing fringe signals from noise, extracting more features to enhance accuracy, and developing more effective localization methods, particularly automatic RFI and cross-coupling removal. We will also explore unsupervised learning for physical parameter estimation and implement GPU acceleration for real-time processing, hoping to enrich the community with novel weak source detection methods.

Acknowledgments

The authors thank Professor Xuelei Chen (Chief of Cosmology Dark Matter and Dark Energy Group, National Astronomical Observatories, Chinese Academy of Sciences), Professor Wenwu Tian (Chief of Comprehensive Astrophysics Research Group, National Astronomical Observatories, Chinese Academy of Sciences), Dr. Dan Wu and their teams for outstanding research contributions. We thank anonymous reviewers and Editor-in-Chief Professor Liang Gao for excellent comments and valuable suggestions that improved this paper. We also thank Professor Dan Hu at the University of North Carolina at Chapel Hill. This study is supported by the National Natural Science Foundation of China (NSFC, grant Nos. 42172323 and 12371454). The authors declare no conflict of interest.

References

Awni, Y., Pranav, R., Masoumeh, H., Geoffrey, H., & Codie, B. 2019, Nature Medicine, 25, 65
Canny, J. 1986, ITPAM, 8, 679
Cao, X., Chen, B., & He, W. 2022, ITIM, 71, 1
Cavaglia, M., Staats, K., & Gill, T. 2018, CCoPh, 25, 1
Chen, X. 2011, SSPMA, 5304, 43
Chen, X. 2012, Int. J. Modern Phys.: Conf. Ser., 12, 256
Dunn, J. C. 1973, Journal of Cybernetics, 3, 32
Everett, S., Yanny, B., Kuropatkin, N., et al. 2022, ApJS, 258, 15
Fong, S., & Narasimhan, S. 2022, ITIM, 71, 1
Fu, Y., Fan, J., Xing, S., et al. 2022, ITIM, 71, 1
Gheller, C., Vazza, F., & Bonafede, A. 2018, MNRAS, 480, 3749
Gong, M., Liang, Y., Shi, J., Ma, W., & Ma, J. 2013, ITIP, 22, 573
Hathaway, R. J., & Bezdek, J. C. 2000, IEEE Transactions on Fuzzy Systems, 8, 576
He, J., Ouyang, M., Chen, Z., Chen, D., & Liu, S. 2022, ITIM, 71, 1
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (Las Vegas, NV: IEEE), 770
Kanezaki, A. 2018, in Icassp IEEE Int. Conf. on Acoustics (Calgary, AB: IEEE)
Korytov, D., Hearin, A., Kovacs, E., et al. 2019, ApJS, 245, 26
Krinidis, S., & Chatzis, V. 2010, ITIP, 19, 1328
Kuang, J., Xu, G., Tao, T., & Wu, Q. 2022, ITIM, 71, 1
Li, J., Zuo, S., Wu, F., et al. 2020, SCPMA, 63, 129862
Lin, A., Chen, B., Xu, J., et al. 2022, ITIM, 71, 1
Liu, W., Zhu, M., Dai, C., He, D., & Yu, X. 2019a, MNRAS, 483, 4774
Liu, W., Zhu, M., Dai, C., et al. 2019b, RAA, 19, 1
Memon, K., & Lee, D. 2018, Fuzzy Sets and Systems, 340, 91
Otsu, N. 1979, ITSMC, 9, 62
Peng, C., Liu, Y., Gui, W., Tang, Z., & Chen, Q. 2022, ITIM, 71, 1
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016, in IEEE Conf. on

Computer Vision and Pattern Recognition (Las Vegas, NV: IEEE), 779
Shelhamer, E., Long, J., & Darrell, T. 2017, ITPAM, 39, 640
Tanoglidis, D., Drlica-Wagner, A., Wei, K., et al. 2021, ApJS, 252, 18
Thompson, A., Moran, J., Swenson, G., & Aurass, H. 1991, Interferometry and Synthesis in Radio Astronomy (New York: Wiley)
Vincent, L., & Soille, P. 1991, ITPAM, 13, 583
Wang, H., Zhu, W., Guo, P., & Li, D. 2019, SCPMA, 62, 61
Xu, Z., Yu, F., Chang, S., & Wang, S. 2015, arXiv:1503.00591
Yan, R., Ma, R., Liu, W., et al. 2022, DSP, 129, 103663
Zadeh, L. A. 1965, InfCo, 8, 338
Zeng, W., Ma, R., Yin, Q., Zheng, X., & Xu, Z. 2020, JIFS, 39, 1
Zeng, X., Wei, S., Shi, J., & Zhang, X. 2021, ITIM, 70, 1
Zhang, R., Li, X., Zhang, H., & Nie, F. 2020, IEEE Transactions on Fuzzy Systems, 28, 2814
Zhu, Y., Liang, X., Wang, T., Xie, J., & Yang, J. 2022, ITIM, 71, 1
Zuo, S., Li, J., Li, Y., et al. 2021, A&C, 34, 100439

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