

Article

Underwater Source Counting with Local-Confidence-Level-Enhanced Density Clustering

Yang Chen , Yuanzhi Xue, Rui Wang * and Guangyuan Zhang

School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213159, China; chenyang.heu@gmail.com (Y.C.); 15895064102@163.com (Y.X.); 20080902029@smail.cczu.edu.cn (G.Z.)

* Correspondence: wangrui@cczu.edu.cn; Tel.: +86-18941196409

Abstract: Source counting is the key procedure of autonomous detection for underwater unmanned platforms. A source counting method with local-confidence-level-enhanced density clustering using a single acoustic vector sensor (AVS) is proposed in this paper. The short-time Fourier transforms (STFT) of the sound pressure and vibration velocity measured by the AVS are first calculated, and a data set is established with the direction of arrivals (DOAs) estimated from all of the time–frequency points. Then, the density clustering algorithm is used to classify the DOAs in the data set, with which the number of the clusters and the cluster centers are obtained as the source number and the DOA estimations, respectively. In particular, the local confidence level is adopted to weigh the density of each DOA data point to highlight samples with the dominant sources and downplay those without, so that the differences in densities for the cluster centers and sidelobes are increased. Therefore, the performance of the density clustering algorithm is improved, leading to an improved source counting accuracy. Experimental results reveal that the enhanced source counting method achieves a better source counting performance than that of basic density clustering.

Keywords: acoustic vector sensor; source counting; DOA estimation; density clustering; local confidence level



Citation: Chen, Y.; Xue, Y.; Wang, R.; Zhang, G. Underwater Source Counting with Local-Confidence-Level-Enhanced Density Clustering. *Sensors* **2023**, *23*, 8491. <https://doi.org/10.3390/s23208491>

Academic Editor: Iren E. Kuznetsova

Received: 13 August 2023

Revised: 4 October 2023

Accepted: 11 October 2023

Published: 16 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Acoustic vector sensors (AVSs) can co-pointedly and synchronously measure sound pressure and particle velocity vectors [1–4]. AVSs are widely used in underwater target detection, direction finding, tracking, and communication [5–7] due to their abilities of spatial filtering [8], DOA estimating [9], and resisting the interference of isotropic noise [10].

The target directing methods based on AVSs include the average sound intensity detector, sound intensity flow DOA histogram, cross-spectral DOA histogram, and so on [11,12]. AVSs can also be used in multi-channel arrays [13]. Each of these methods has its advantages and disadvantages. The histogram algorithm is widely used in engineering due to its better robustness compared with other algorithms. It can suppress narrowband and strong line spectrum interference and has a certain degree of multi-target resolution ability [14,15]. In [11], a DOA histogram based on sound intensity flow was used to estimate the DOA of targets, and the weighted DOA histogram of the line spectrum was used to realize the resolution and DOA estimation of multiple line spectrum targets. In [16], a DOA histogram was used to realize the resolution and DOA estimation of multiple wideband targets in the time–frequency domain and instantaneous frequency domain of the Huang transform. In an AVS sea trial experiment based on the Argo buoy platform in [17], it was also found that the DOA histogram could distinguish two wideband targets, the test ship and the engineering ship, whose adjacent interval was about 80 degrees. In reference [18,19], the windowed-disjoint orthogonality (WDO) of the target signal was further introduced to explain the mechanism of the wideband multi-target resolution ability of the DOA histogram.

The components of ship-radiated noise are very complex, including line spectrum, stationary ergodic random signals, and transient signals [20]. Therefore, it is considered to be broadband noise with energy distribution at any time and any frequency from the macro point of view. However, from the micro point of view, the energy intensity at different time–frequency points varies, which leads to a significant difference in the energy of different sources at some time–frequency points when multiple sources are synthesized at the sensor receiver, and a target playing a dominant role. This phenomenon is called the window-disjoint orthogonality (WDO) of signals and is widely used in blind signal separation for multi-source resolution and separation [21,22]. The higher the WDO characteristic of the signal in a certain time frequency point, the greater the energy of the dominant signal is compared to the sum of other signals, and the direction estimation result of the frequency point will be biased to the target direction of the dominant signal. If significant numbers of TF points possess the WDO property, in the DOA histogram the DOA estimates will cluster around the actual DOAs of the sources to achieve multi-target resolution of the histogram.

With multi-target resolution, it is possible to achieve source counting through multiple source detection. Traditional multiple source detection methods such as Minimum Description Length (MDL) [23], Akaike Information Criterion (AIC) [24], and Random Matrix Theory (RMT) [25] are overdetermined. They work well when the sensor channels outnumber the sources. However, for two-dimensional AVSs, these methods will fail when there are more than two sources, in which case the applicability of the AVS will be greatly limited.

As a result, an underdetermined method based on the density clustering algorithm was formed to solve the source counting problem for single AVSs in [18]. The density clustering algorithm was used to classify the DOAs of all the time–frequency points, where the cluster centers and the number of the clusters were obtained as the DOA estimations and target number, respectively. However, WDO is the inherent characteristic of the signal. With the increase in targets, the WDO decreases, which limits the multi-target resolution performance of the DOA histogram, and the performance of the density clustering algorithm deteriorated.

For this reason, a multi-source detection based on multimodal fusion was developed in [10]. The output of the AVS was decomposed into multiple modes by intrinsic time-scale decomposition (ITD), so that the source number in each mode decreased. For each mode, the WDO increased, leading to a better source counting accuracy. Therefore, the fused source counting performance was improved. However, the counting performance varied with the number of modes employed and it was difficult to determine the optimal number of employed modes.

In this paper, the local confidence level is adopted to weigh the density of each DOA data point to highlight the samples with the dominant source and downplay those without, so that the differences in densities for the cluster centers and sidelobes are increased. Therefore, the performance of the density clustering algorithm is enhanced. An analysis of lake trial data is conducted by comparing the proposed local-confidence-level-enhanced density clustering method with the methods in [10] and [18]. The results confirm the availability of the proposed local-confidence-level-enhanced density clustering method in improving source counting performance. The source counting accuracy of the proposed method is better than the methods it is compared with. The probability distribution of the source counting result for the local-confidence-level-enhanced method is more concentrated to the source number. As the SNR decreases, the proposed method undergoes slower degradation.

2. Model and Cross-Spectral DOA Histogram of Single-Vector Sensor

The output of an AVS contains information on both sound pressure and vibration velocity. The model of two-dimensional AVSs in the horizontal free sound field can be expressed as:

$$\mathbf{y}(t) = \begin{bmatrix} P(t) \\ V_x(t) \\ V_y(t) \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^N x_n(t) + n_p(t) \\ \sum_{n=1}^N x_n(t) \cos \alpha_n + n_x(t) \\ \sum_{n=1}^N x_n(t) \sin \alpha_n + n_y(t) \end{bmatrix} = \sum_{n=1}^N x_n(t) \begin{bmatrix} 1 \\ \cos \alpha_n \\ \sin \alpha_n \end{bmatrix} + \mathbf{e}(t), \quad (1)$$

where N is the number of sources, $P(t)$ is the sound pressure component of the AVS output, and $V_x(t)$ and $V_y(t)$ are the components of vibration velocity on two axes, respectively. $x_n(t)$ represents the signal radiated from the n th source to the AVS, $\alpha_n(t)$ indicates the angle of the n th sound source relative to the vector hydrophone (x axis is zero-degree orientation), $\mathbf{e}(t) = [n_p(t), n_x(t), n_y(t)]^T$ is the noise vector, and $n_p(t)$, $n_x(t)$, and $n_y(t)$ are noises of the sound pressure, x -axis, and y -axis vibration velocities, respectively.

By taking the short-time Fourier transform (STFT) of each channel of the mixture, the AVS model can be expressed in the time–frequency domain as:

$$\mathbf{Y}(\omega, m) = [P(\omega, m), V_x(\omega, m), V_y(\omega, m)]^T = \sum_{n=1}^N X_n(\omega, m) \begin{bmatrix} 1 \\ \cos \alpha_n \\ \sin \alpha_n \end{bmatrix} + \mathbf{E}(\omega, m), \quad (2)$$

where ω and m are the frequency bin and time frame indices, respectively. $P(\omega, m)$, $V_x(\omega, m)$, $V_y(\omega, m)$, $X_n(\omega, m)$, and $\mathbf{E}(\omega, m)$ are the STFTs of $P(t)$, $V_x(t)$, $V_y(t)$, and $\mathbf{e}(t)$, respectively. The DOA estimation at each time–frequency point can be obtained by:

$$\theta(\omega, m) = \angle \left[\text{Re}\{P(\omega, m)V_x^*(\omega, m)\} + j \cdot \text{Re}\{P(\omega, m)V_y^*(\omega, m)\} \right], \quad (3)$$

where \angle represents the phase angle, and $V_x^*(\omega, m)$ and $V_y^*(\omega, m)$ represent the complex conjugates of $V_x(\omega, m)$ and $V_y(\omega, m)$, respectively. The DOA histogram can be obtained by counting the estimation results of all time–frequency points. Supposing that the azimuth interval of histogram statistics is $\Delta\theta$, we count the number of DOA estimation results in the interval $[\theta - \Delta\theta/2, \theta + \Delta\theta/2]$: when $\theta - \Delta\theta/2 < \theta(\omega, m) \leq \theta + \Delta\theta/2$, the amplitude of the histogram at the corresponding position is increased by 1:

$$H_b(\theta) = \sum_{\theta - \Delta\theta/2 < \theta(\omega, m) \leq \theta + \Delta\theta/2} 1. \quad (4)$$

3. Local-Confidence-Level-Enhanced Density Clustering Source Counting

3.1. Density-Clustering-Based Source Counting

The density clustering algorithm depends on two important hypotheses: (1) The density for the center of a cluster is higher than that of the neighboring points; (2) The minimum distance between the center of a cluster and a higher-density point is relatively large [26,27]. The data set $\Theta = [\theta_1, \theta_2, \dots, \theta_K]$ is established with the DOA estimations obtained by Equation (3) in the time–frequency domain, where K is the number of the DOA estimations. Since the distance between two DOA estimations does not exceed 180° , the distance d_{kl} between each two samples in Θ is defined as:

$$d_{kl} = \begin{cases} |\theta_l - \theta_k|, & |\theta_l - \theta_k| < 180 \\ 360 - |\theta_l - \theta_k|, & |\theta_l - \theta_k| \geq 180 \end{cases}, \quad k, l \in \{1, \dots, K\} \quad (5)$$

where θ_k and θ_l are two different samples in the data set Θ . They are angles limited within the interval of $[-180^\circ, 180^\circ]$. The absolute difference between two angle values $|\theta_l - \theta_k|$ is

within the interval of $[0^\circ, 360^\circ]$. The distance between two angles is periodic with the period of 180° . So, when $|\theta_l - \theta_k| < 180$, the distance d_{kl} is $|\theta_l - \theta_k|$ itself. When $|\theta_l - \theta_k| \geq 180$, the distance d_{kl} is $360 - |\theta_l - \theta_k|$.

Then, the local density of each sample ρ_k is calculated as:

$$\rho_k = \sum_{l \in \{1, \dots, K\}} \chi(d_{kl} - d_c), \quad (6)$$

where $\chi(x) = 1$ if $x < 0$ and $\chi(x) = 0$ otherwise, and $d_c > 0$ represents the cut-off distance used to keep a region for each sample. ρ_k is the number of the samples within a limited distance of d_c from the sample θ_k , i.e., ρ_k is equal to the number of points that are closer than d_c to the sample θ_k . The more samples there are near the sample θ_k , the larger the local density of the sample θ_k is. d_c can be chosen so that the average number of neighbors is around 1 to 2% of the total number of points in the data set [28]. A sample is assigned to the same cluster of its nearest sample with a higher density only if their distance is smaller than d_c . The minimum distance δ_k between the sample θ_k and any other sample with a higher density is measured as:

$$\delta_k = \min_{l: \rho_l > \rho_k} (d_{kl}). \quad (7)$$

For the point with the highest density, we conventionally take δ_k as:

$$\delta_k = \max_{l \in \{1, \dots, K\}} (d_{kl}). \quad (8)$$

The product $\gamma_k = \rho_k \delta_k, k = 1, 2, \dots, K$ is used as the feature, and its value for a cluster center sample is obviously larger than that for the other sample. Thus, the differences among the ordered feature sequence are used to find the samples with significantly larger features, whose number corresponds to the number of clusters or targets [29]. In this procedure, the features $\gamma_k, k = 1, 2, \dots, K$ are sorted in descending order, i.e.,

$$\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_K. \quad (9)$$

Suppose that there are L targets in the detection range of the AVS. Since the features of the cluster center samples are much larger than the other samples, the first L features $\{\gamma_1, \gamma_2, \dots, \gamma_L\}$ are significantly larger than the other $K - L$ features $\{\gamma_{L+1}, \gamma_{L+2}, \dots, \gamma_K\}$. Consequently, the difference between γ_L and γ_{L+1} is relatively large. To obtain L , the differences in the ordered features are computed as:

$$\Delta\gamma_i = \gamma_i - \gamma_{i+1}, \quad i = 1, 2, \dots, K-1, \quad (10)$$

And the variance of the sequence $\{\Delta\gamma_i\}_{i=n}^{K-1}$ is calculated as:

$$\sigma_n^2 = \frac{1}{K-n} \sum_{i=n}^{K-1} \left(\Delta\gamma_i - \frac{1}{K-n} \sum_{i=n}^{K-1} \Delta\gamma_i \right)^2, \quad (11)$$

where $n = 1, 2, \dots, K-2$. Further, the second-order statistic of the features is defined as:

$$S_n = \begin{cases} \frac{\sigma_{n+1}^2}{\sigma_n^2}, & \sigma_n^2 > 0 \\ +\infty, & \sigma_n^2 = 0 \end{cases} \quad (12)$$

Finally, the target (or cluster) number is estimated as:

$$L = \underset{n=1, \dots, K-3}{\operatorname{argmin}} S_n, \quad (13)$$

And the L DOA samples with the L largest features are obtained as the final multi-target DOA estimation results.

3.2. Local-Confidence-Level-Enhanced Density Clustering Algorithm

With the increase in sources, the WDO decreases, which limits the multi-target resolution performance of the DOA histogram. DOA estimations of different time–frequency points have different contributions to density clustering. For a certain time–frequency point, the larger the proportion of the dominant signal energy is, the closer the DOA estimation is to the truth value, and the greater the contribution this sample provides to the clustering. A local-confidence-level-based density enhancement algorithm is proposed in which the contribution of the samples with a high proportion of the dominant signal energy are exaggerated, so that the accuracy of the clustering and the precision of the target number estimation are improved.

For a certain time–frequency point (ω, m) (red circle) in the STFT domain, define the rectangular area $\Omega_{\omega, m}$ (Dotted Box) around it, as illustrated in Figure 1. The width and length of this area are l_m time points and l_ω frequency points, respectively. Therefore, there are $l_\omega l_m$ time–frequency points in the rectangular area $\Omega_{\omega, m}$. The local confidence $\Gamma(\omega, m)$ of this time–frequency point (ω, m) is estimated by performing a principal component analysis (PCA) [30] on the snapshot vector $Y(\Omega_{\omega, m})$ in the region $\Omega_{\omega, m}$, which reflects the strength of the dominant signal at the time–frequency point (ω, m) .

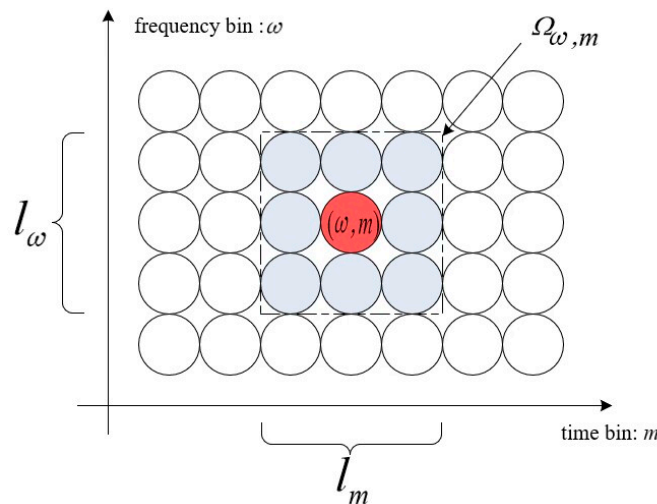


Figure 1. Time–frequency point (ω, m) and the surrounding rectangle area $\Omega_{\omega, m}$.

For each time–frequency point, a snapshot $Y(\Omega_{\omega, m})$ is formed with all snapshots $Y(\omega, m)$ in the region $\Omega_{\omega, m}$. It is a $3 \times l_\omega l_m$ matrix with columns $Y(\omega, m)$, $(\omega, m) \in \Omega_{\omega, m}$. A positive semidefinite complex Hermitian matrix is then constructed with $Y(\Omega_{\omega, m})$ for the region $\Omega_{\omega, m}$ as:

$$R(\omega, m) = Y(\Omega_{\omega, m})Y^H(\Omega_{\omega, m}). \quad (14)$$

For a two-dimensional AVS, the rank of the matrix $R(\omega, m)$ is 3. Performing an eigenvalue decomposition on $R(\omega, m)$, we obtain three real-valued positive eigenvalues in decreasing order $\lambda_1(\omega, m) \geq \lambda_2(\omega, m) \geq \lambda_3(\omega, m)$ of the matrix $R(\omega, m)$. The local confidence level of the time–frequency point (ω, m) is expressed as:

$$\Gamma(\omega, m) = \frac{2\lambda_1(\omega, m)}{\lambda_2(\omega, m) + \lambda_3(\omega, m)}. \quad (15)$$

If there is only one dominant signal whose power is much larger than the noise in the time–frequency point (ω, m) , λ_1 is the eigenvalue of the dominant signal which is much larger than λ_2 and λ_3 , which are the eigenvalues of the noise. Otherwise, if there are two similar signals whose power is much larger than the noise, λ_2 will be much larger than λ_3 , and the local confidence will become much smaller than the one dominant signal case. Therefore, it can be used to represent the WDO.

It is obvious that the local confidence level is proportional to the ratio of the dominant signal.

Then, the local confidence is used to enhance the cross-spectral DOA histogram. In the original orientation histogram statistics, the density of each sample is 1. The local-confidence-weighted DOA histogram with a sample density of $\Gamma(\omega, m)$ is:

$$H_e(\theta) = \sum_{\theta - \Delta\theta/2 < \theta(\omega, m) \leq \theta + \Delta\theta/2} \Gamma(\Omega_{\omega, m}). \quad (16)$$

Substituting Equation (4) into Equation (16) yields:

$$H_e(\theta) = H_b(\theta) \cdot \left[\sum_{\theta - \Delta\theta/2 < \theta(\omega, m) \leq \theta + \Delta\theta/2} \Gamma(\Omega_{\omega, m}) \right] = H_b(\theta) \bar{\Gamma}(\theta), \quad (17)$$

where $\bar{\Gamma}(\theta)$ is the average local confidence level.

The enhanced local density ρ'_k of the k th sample is weighted with the corresponding local confidence level $\Gamma(k)$ as:

$$\rho'_k = \Gamma(k) \rho_k. \quad (18)$$

The minimum distance is calculated according to the weighted local density ρ'_k , and the target number is estimated by the ordered feature sequence afterward. The remaining steps are the same as in Section 3.1.

With the enhanced local density, the density of the sample points that are closer to the cluster center will increase while the density of the sample points far away from the cluster center will decrease. Consequently, the performance of the density-clustering-based multi-target detection is improved. To summarize, the proposed source counting method based on the local confidence-enhanced density clustering is depicted in Algorithm 1.

Algorithm 1: Local-confidence-level-enhanced density clustering source counting

Input: Output $\mathbf{y}(t)$ of AVS.

Output: Target number L and the DOAs $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_L)$ of the targets.

1 Compute the STFT $Y(\omega, m)$ of $\mathbf{y}(t)$ with Equation (2);

2 Define the rectangle area $\Omega_{\omega, m}$ for each (ω, m) point;

3 Compute positive semidefinite complex Hermitian matrix $\mathbf{R}(\Omega_{\omega, m})$ with Equation (14);

4 Perform eigenvalue decomposition on $\mathbf{R}(\Omega_{\omega, m})$ to obtain the eigenvalues $\lambda_1(\omega, m)$, $\lambda_2(\omega, m)$, and $\lambda_3(\omega, m)$;

5 Compute the local confidence level $\Gamma(\omega, m)$ with Equation (15);

6 Compute the DOA estimations $\theta(\omega, m)$ at all the time–frequency points with Equation (3);

7 Consist data set Θ with the DOA estimations;

8 **for** $k = 1$ **to** K

9 Compute the enhanced local density ρ'_k of each sample with Equation (18);

10 **for** $l = 1$ **to** K

11 Compute the distance d_{kl} between each two samples in Θ with Equation (5);

12 **end for**

13 **end for**

14 **for** $k = 1$ **to** K

15 Compute the minimum distance δ_k with Equations (7) and (8);

16 Compute features as $\gamma_k = \rho'_k \delta_k$;

17 **end for**

18 Sort the features γ_k ($k = 1, 2, \dots, K$) in descending order;

19 Compute $\Delta\gamma_i$ of the ordered features with Equation (10);

20 Compute the variance of $\{\Delta\gamma_i\}_{i=n}^{K-1}$ with Equation (11);

21 Compute the second-order statistic S_n of the ordered features with Equation (12);

22 Estimate the target number L with Equation (13);

23 Search DOAs $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_L)$ with the L largest features.

4. Lake Trial and Analysis

4.1. Experimental Settings

The experimental data were collected in FuXian Lake. The experimental scenario is depicted in Figure 2, in which the location of a two-dimensional co-vibration AVS is in the middle of the lake. The sensitivity of the velocity channel decreased with increasing frequency (as the slope of -6 dB/octave), and there was a 90-degree phase difference between the vibration velocity channel and the sound pressure channel. The AVS was rigidly fixed about 4 m underwater on the side of the survey ship. The output signals of the AVS were collected and stored by a multi-channel synchronous data acquisition system. In the experiment, four ships (yachts) acted as moving targets located 1–2 km around the AVS. The ships were distributed in four directions of about 40° , 140° , 210° , and 330° at the beginning. The sampling rate was 48 kHz, the length of the STFT window was 8192 points, the frequency band was 0.5–8 kHz, the integral length of the azimuth histogram was 1 s (six windows), the sliding step was 8192 points (one window), and the rectangular area was $l_m = l_w = 3$. The signals of two vibration velocity channels were compensated in the frequency domain according to the slope and phase of sensitivity after STFT.

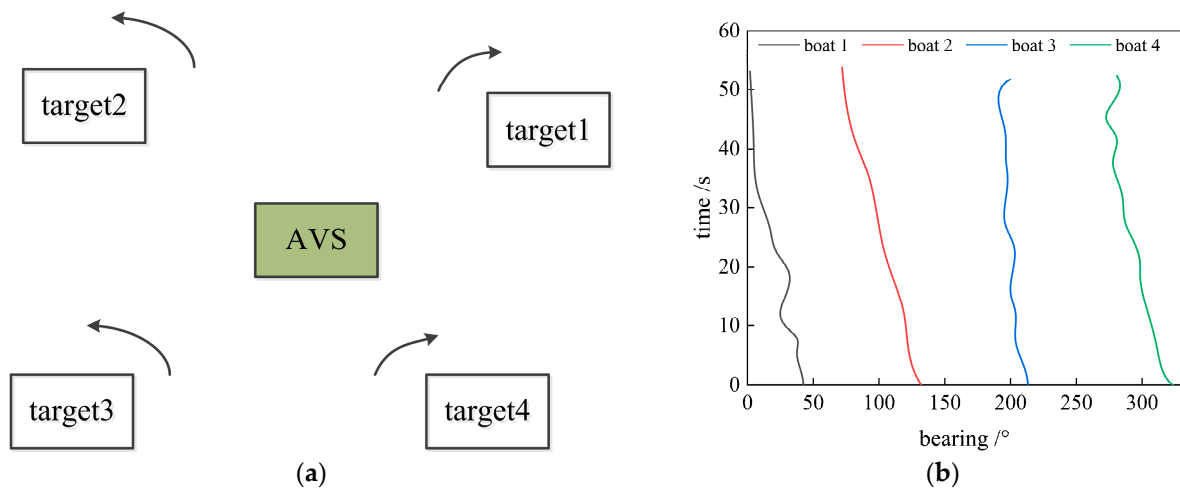


Figure 2. Experimental setup and traces of 4 boats: (a) Experimental setup; (b) Experiment azimuth waterfall map.

4.2. Resolution Performance

Figure 3 illustrates the temporal course of the cross-spectral DOA histogram computed with Equation (4) at each time bin. Four targets marked as <1>, <2>, <3>, and <4> are presented in the histogram. The red color stands for the peaks of the histogram, which indicate the locations of the targets. Continuous peaks in time compose the time-bearing course of a target, which is illustrated roughly with straight black, red, green, and blue dash-dot lines, respectively, for each target. As shown in Figure 3, the azimuths of targets 3 and 4 gradually become closer over time, so the spectral peaks of these two targets become hard to distinguish (red color connects in one piece around time (d)). Since the signal from the target ship is weak when the ship decelerates or stops, the spectral peaks of targets 1 and 2 are unapparent or even submerged by the background during 38–40 s in the DOA histogram, as with the time points 44–50 s of target 3 and 50–53 s of target 4 (red color absent). Thus, the target courses have obvious discontinuity.

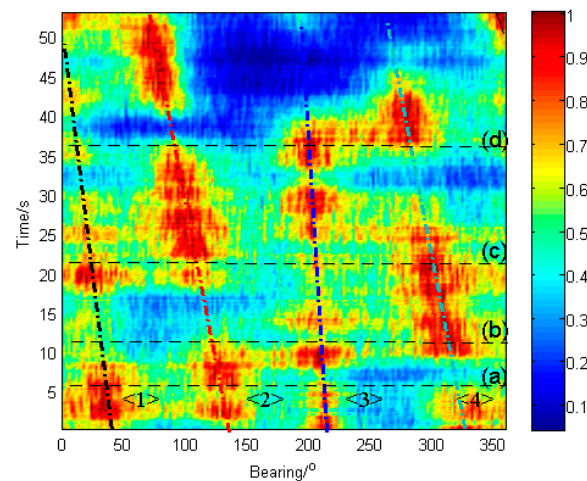


Figure 3. The temporal course of the cross-spectral DOA histogram of vector hydrophone from lake trial. ((a), (b), (c) and (d) denote the time of 6.7 s, 11.6 s, 21.3 s and 36.0 s).

Figure 4 depicts the temporal course of the local-confidence-level-enhanced cross-spectral DOA histogram computed with Equation (16) at each time bin. It can be seen that the course of target 3 is significantly enhanced (red color is more continuous in time and peaks are higher than those in Figure 3), and the course of other targets is also enhanced relative to the original ones in Figure 3. It indicates that by using the local confidence level enhancement, the DOA courses of these targets become apparent, and both the resolution of targets and the continuity of the courses are improved (red color areas are thinner in bearing axis direction and more clearly separated compared to Figure 3).

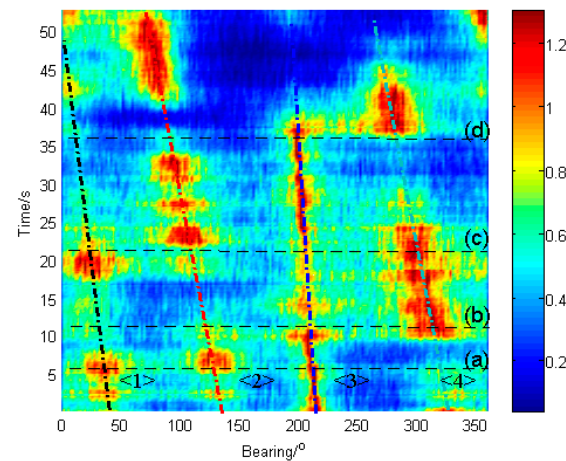


Figure 4. The temporal course of the local-confidence-level-enhanced cross-spectral DOA histogram. ((a), (b), (c) and (d) denote the time of 6.7 s, 11.6 s, 21.3 s and 36.0 s).

Figure 5 shows the comparative results between the cross-spectral DOA histogram and the enhanced cross-spectral DOA histogram at different time points corresponding to (a), (b), (c), and (d) in Figures 3 and 4, respectively. As shown in Figure 5, both the original and the enhanced histograms present spectral peaks at the directions of the targets, while the enhanced one possesses glaringly obvious peaks. This reveals that the enhanced histogram is able to estimate the azimuths of targets at a higher resolution. In particular, when the signal from the target ship is weak, such as target 3 in Figure 5c, the enhanced cross-spectrum can significantly upgrade its peak, which makes it possible to detect weak targets. Additionally, as depicted in Figure 5d, the distinguishability of adjacent targets (targets 3 and 4) is also promoted by the enhanced cross-spectrum.

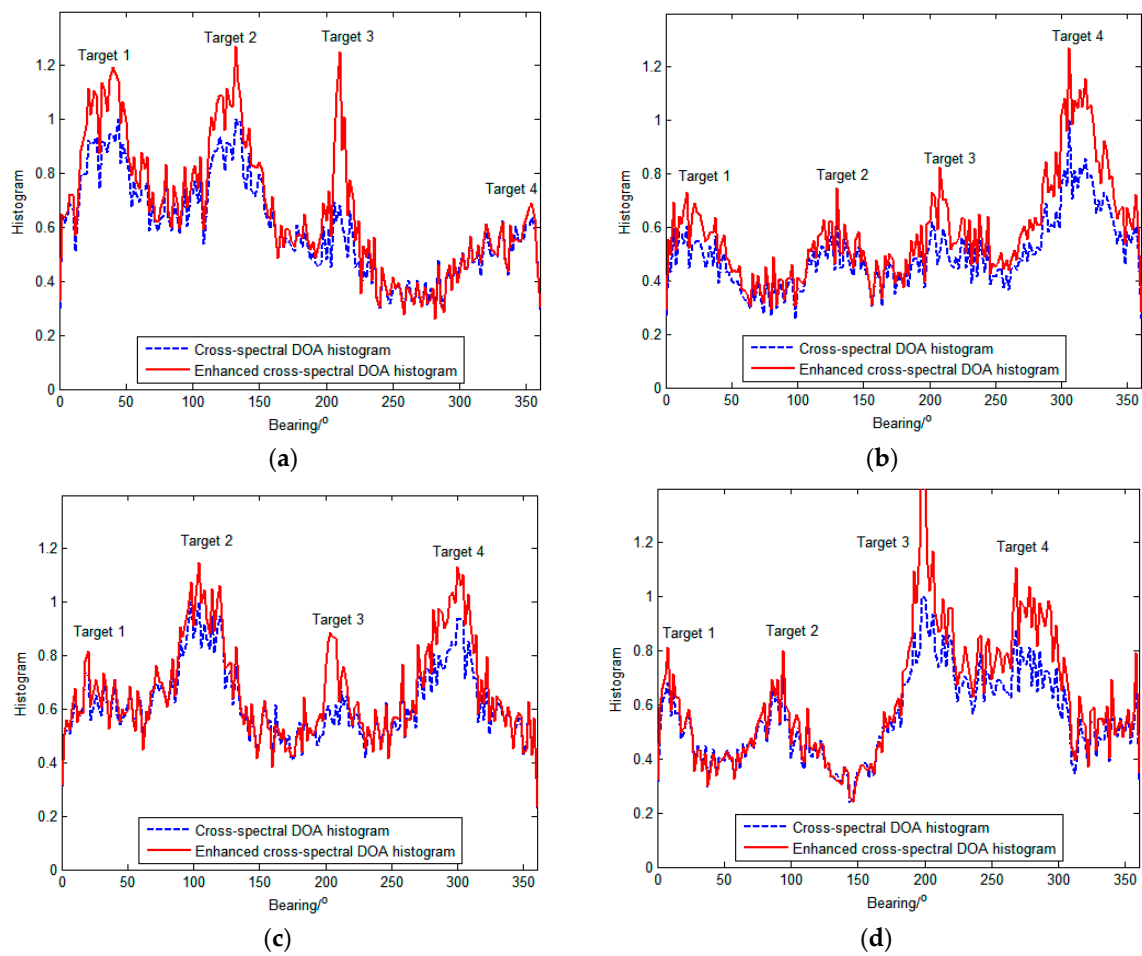


Figure 5. Comparison of cross-spectral DOA histograms and local-confidence-level-enhanced cross-spectral DOA histograms at different times. (a) DOA histograms at time (a); (b) DOA histograms at time (b); (c) DOA histograms at time (c); (d) DOA histograms at time (d).

Figure 6 explains the enhancement performance of the local confidence level by the DOA histogram at time (c). The average local confidences of the directions of the targets are larger than the others. As a result, the peaks of the targets become higher and sharper.

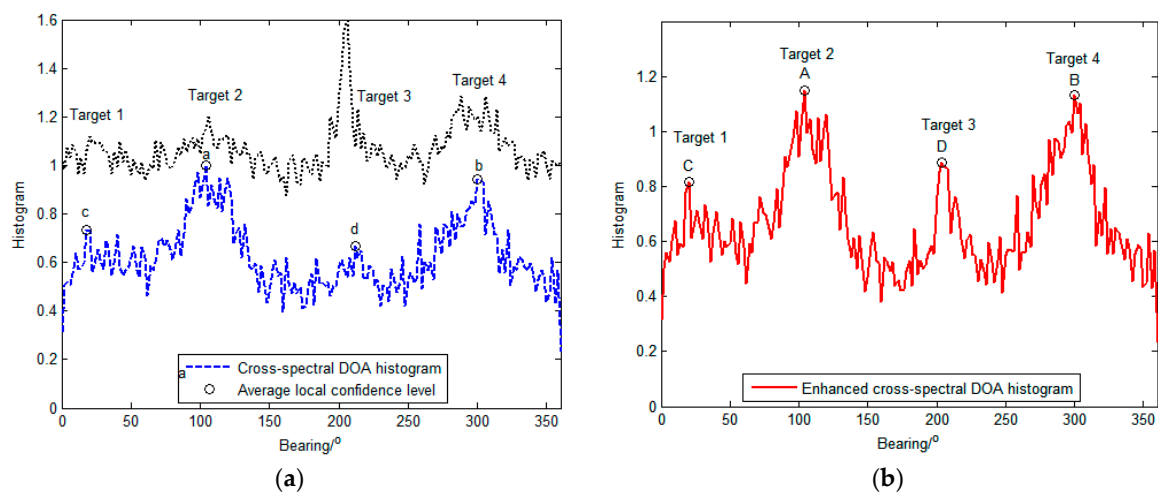


Figure 6. DOA histogram: (a) The cross-spectral DOA histogram and average local confidence level; (b) The enhanced cross-spectral DOA histogram. (The circles a–d and A–D denote the data samples considered as the clusters centers).

4.3. Source Counting Performance

To explain the advantage of the local-confidence-level-enhanced density clustering algorithm, Figure 7 illustrates the decision graphs of the basic and enhanced density clustering algorithms with the data in Figure 6. Due to the inconspicuousness of the peaks of targets 1 and 3 (i.e., points c and d in Figure 6a), the cluster centers (i.e., decision points c and d in Figure 7a) of the basic density clustering are close to the cluster halos, which are recognized as the background. Nevertheless, in the enhanced density clustering algorithm, the local density is enhanced with the corresponding local confidence level which is represented by the size of the points in Figure 7. The local confidence level is higher around the cluster center than that in the cluster halo. Therefore, the background becomes more clustered and the cluster centers are far away from the background, especially for targets 1 and 3 (i.e., points C and D in Figure 7b). Therefore, the enhanced local density leads to a clearer distinction between the cluster centers and the background in the decision graph, which makes it easier to achieve correct counting.

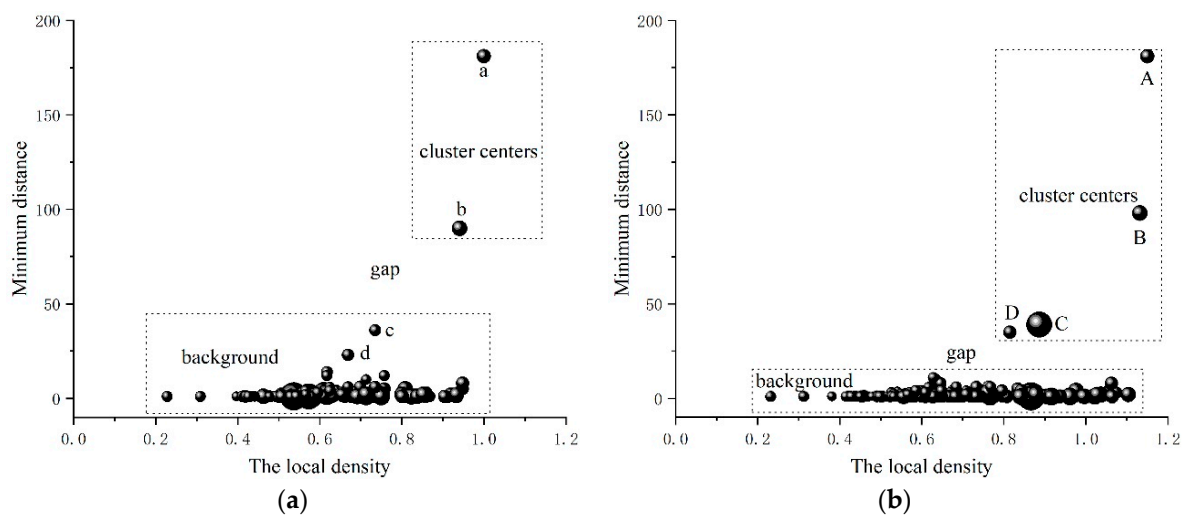


Figure 7. The decision graphs of the basic and enhanced density clustering algorithms for the data in Figure 6: (a) Basic density clustering; (b) Enhanced density clustering. (The points a–d and A–D denote the data samples considered as the clusters centers).

Figure 8 demonstrates the second-order statistic S_n of the basic and enhanced density clustering algorithms with the data in Figure 6. Obviously, the basic density clustering obtains the wrong counting result while the enhanced density clustering obtains the correct one. These results are consistent with the analysis of Figure 7.

Figure 9 illustrates the courses of the multi-target DOA estimations obtained by the basic density clustering and the local-confidence-level-enhanced density clustering. Compared with the courses estimated by the basic density clustering, the outliers in the courses estimated by the enhanced density clustering significantly reduce by weighting the density with the local confidence level, and the continuity of the courses of the multi-target detection is remarkably improved, especially in Regions A and B. Moreover, the distribution of the DOA estimations of targets is more compact for enhanced density clustering.

Figure 10 depicts the results of the source number estimations obtained by the basic density clustering and the local-confidence-level-enhanced density clustering. Obviously, in Regions A, B, and C, the enhanced density clustering achieves the correct estimation results while the basic density clustering does not. In Region D, although both methods obtain incorrect results, the outcomes of the enhanced density clustering are closer to the true value. Targets 3 and 4 stop for a while during the periods of 44–50 s and 50–53 s, respectively. During these periods, there are only three targets in the field. As a result, the source number estimation result in the period before 44 s was used to calculate the source counting accuracy.

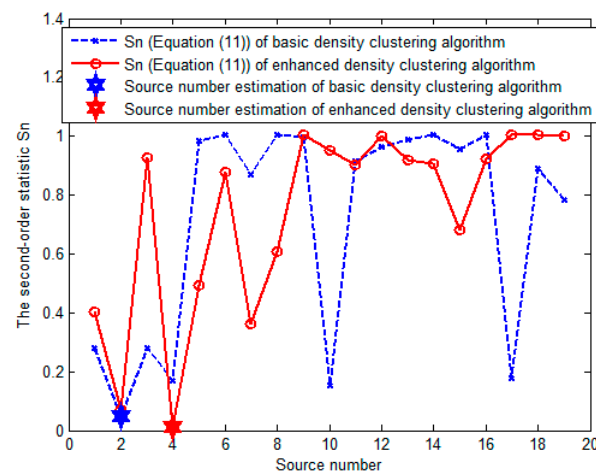


Figure 8. The second-order statistic S_n of the basic and enhanced density clustering algorithms for the data in Figure 6.

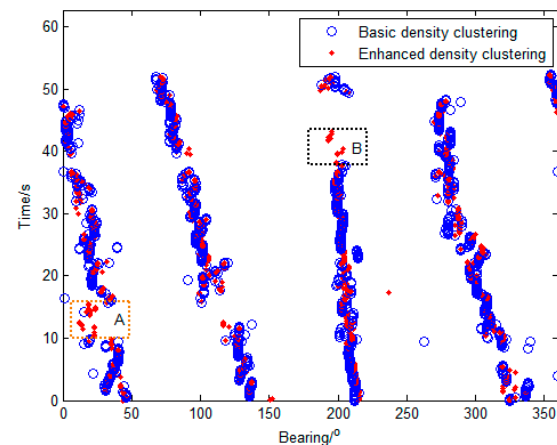


Figure 9. Courses of the multi-target DOA estimations with the basic density clustering, and the local-confidence-level-enhanced density clustering. (Comparing to the basic density clustering, the continuity of the courses of the enhanced density clustering is remarkably improved in Regions A and B).

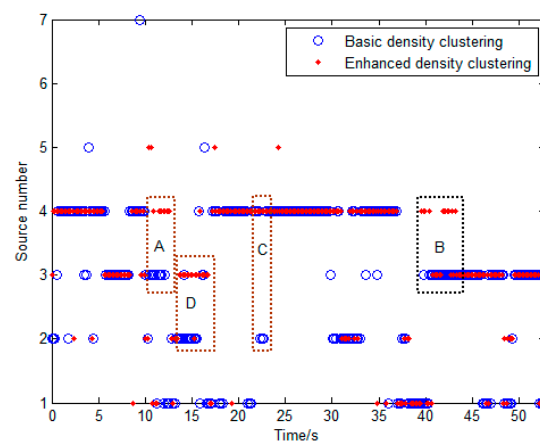


Figure 10. Courses of the source number estimations obtained by the basic density clustering and the local-confidence-level-enhanced density clustering. (The enhanced density clustering achieves the correct estimation results while the basic density clustering does not in Regions A, B, and C. The outcomes of the enhanced density clustering are closer to the true value in Regions D).

A comparison was conducted between the proposed local-confidence-level-enhanced density clustering method, the multimodal-fusion-based method in [10], and the basic density-clustering-based method in [18]. As shown in Figure 11, the basic density-clustering-based method obtained 48.82% accuracy of target number estimation, the multimodal-fusion-based method obtained an accuracy of 51.15% with four modes employed, while the local-confidence-level-enhanced method achieved 63.39% accuracy. The probability distribution of the source counting result for the local-confidence-level-enhanced method is more concentrated to the source number.

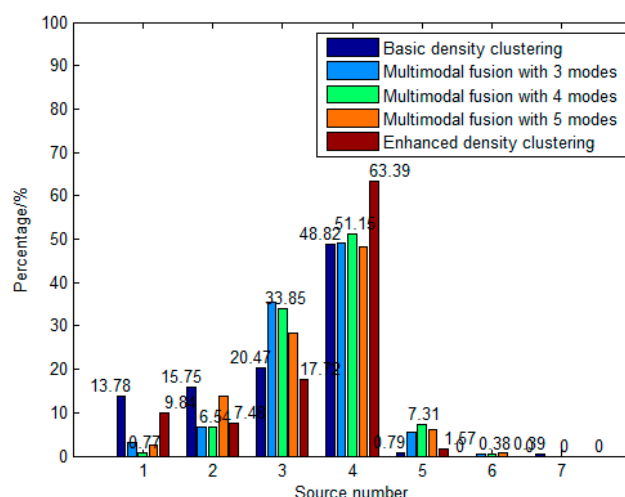


Figure 11. The estimated source number histogram of the basic density-clustering-based method, the multimodal-fusion-based method, and the enhanced density-clustering-based method for the experiment.

The multimodal-fusion-based method significantly reduces the probability of underestimating the errors of one and two sources and worsens the probability of overestimating errors of five sources. The WDO increases and sources are more detectable in each mode, so the probability of missed detections is reduced. However, a source signal may be divided into parts and be distributed in more than one mode. These parts may not be fully recognized as the same source during the fusion step, leading to an overestimation. As a result, compared to the basic density-clustering-based method, despite an advantage in the estimation accuracy, the probability distribution of estimated results is much better for the multimodal-fusion-based method. In addition, the estimation accuracy varies with the number of modes employed. Therefore, it can be seen that the presetting of the number of employed modes is quite important.

In contrast, the local-confidence-level-enhanced density clustering dramatically improves the accuracy of the source counting. In the enhanced method, the local density of the samples around the cluster centers with high local confidence is increased. Therefore, the clusters become more compact, which improves the source counting accuracy.

In order to evaluate the effect of the noise on the proposed method, an experiment with different SNRs was performed. The real SNR of the lake trial is unknown. It was assumed, empirically, to be 24 dB. The SNR was adjusted by adding simulated Gaussian white noise to the lake trial signal. The performance of the basic density-clustering-based method and the enhanced density-clustering-based method for 8, 12, 16, 20, and 24 dB (without simulated noise) is displayed in Figure 12 comparatively. As the SNR decreases, the performance of both methods degenerates gradually. However, the enhanced density-clustering-based method performs better than the basic density-clustering-based method and undergoes slower degradation.

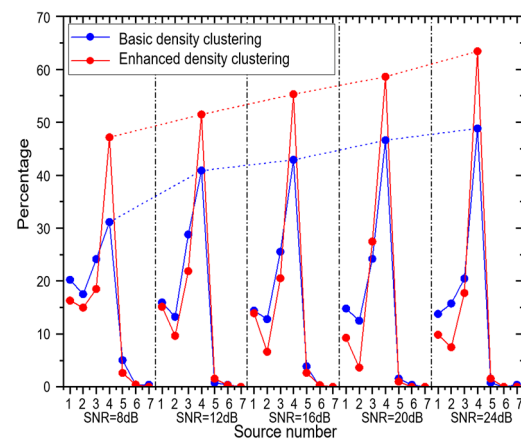


Figure 12. Source counting performance vs. SNR.

To investigate the frequency bands of the targets in the experiment, each TF point was assigned to a target according to its dominant signal. If the DOA estimation of a certain TF point was 2° away from the nearest cluster center, it was assigned to this target. Otherwise, it is considered that there is no dominant signal in this time–frequency point. Figure 13 shows the assignment of the first 1.4 s. The frequencies of all four targets spread over the whole broadband of 500–8000 Hz randomly, and clearly cannot be divided into four different separated sub-bands. This result indicates that the WDO assumption holds in this experiment.

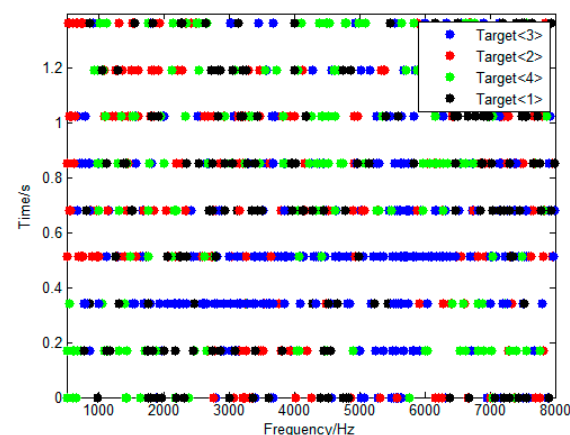


Figure 13. Frequency distribution of targets in time–frequency domain.

5. Conclusions

An underwater source counting method with local-confidence-level-enhanced density clustering is proposed. In this method, the STFT of the AVS output signal is calculated, and the DOAs of all time–frequency points consist of a data set. Then, the DOAs in the data set are classified with density clustering, where the cluster centers and the number of clusters are obtained as the DOA estimations and the target number, respectively. Finally, the local confidence level is applied to further enhance the performance of the density clustering algorithm. A lake trial is conducted to verify the local-confidence-level-enhanced source counting method which achieves a better source counting performance than the multimodal-fusion-based method and the enhanced density-clustering-based method.

Author Contributions: Conceptualization, Y.C. and R.W.; methodology, Y.C.; software, Y.X. and G.Z.; validation, Y.C. and Y.X.; formal analysis, R.W.; investigation, Y.C. and R.W.; resources, Y.X.; data curation, Y.X.; writing—original draft preparation, R.W.; writing—review and editing, Y.C. and R.W.; visualization, Y.X.; supervision, Y.C. and R.W.; project administration, R.W.; funding acquisition, R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Office of Science and Technology of Changzhou, Grant Number CJ20220100; Postdoctoral Foundation of Jiangsu Province, Grant Number 2021K187B; National Postdoctoral General Fund, Grant Number 2021M701042; Foundation of State Key Laboratory of Automotive Simulation and Control, Grant Number 20210241, the National Natural Science Foundation of China under Project 61501061.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to express their gratitude to every reviewer of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, D.; Zhu, Z.; Tian, Y. Theoretical bases and application development trend of vector sonar technology. *J. Unmanned Undersea Syst.* **2018**, *26*, 185–192. (In Chinese)
2. Sun, Q.; Zhang, X.; Han, M.; Wang, W. Combined pressure-resistant co-vibration vector hydrophone for underwater glider platform. *Acta Armamentar* **2020**, *41*, 2063–2070. (In Chinese)
3. Zou, Y.; Liu, Z.; Ritz, C.H. Enhancing Target Speech Based on Nonlinear Soft Masking Using a Single Acoustic Vector Sensor. *Appl. Sci.* **2018**, *8*, 1436. [[CrossRef](#)]
4. Roh, T.; Yeo, H.G.; Joh, C.; Roh, Y.; Kim, K.; Seo, H.-S.; Choi, H. Fabrication and Underwater Testing of a Vector Hydrophone Comprising a Triaxial Piezoelectric Accelerometer and Spherical Hydrophone. *Sensors* **2022**, *22*, 9796. [[CrossRef](#)]
5. Felisberto, P.; Santos, P.; Jesus, S.M. Acoustic pressure and particle velocity for spatial filtering of bottom arrivals. *IEEE J. Ocean. Eng.* **2019**, *44*, 179–192. [[CrossRef](#)]
6. Kim, J. TDOA-Based Target Tracking Filter While Reducing NLOS Errors in Cluttered Environments. *Sensors* **2023**, *23*, 4566. [[CrossRef](#)] [[PubMed](#)]
7. Yin, J.; Zhu, G.; Han, X.; Ge, W.; Li, L.; Tian, Y. Iterative channel estimation-based soft successive interference cancellation for multiuser underwater acoustic communications. *Acoust. Soc. Am.* **2021**, *150*, 133–144. [[CrossRef](#)]
8. Mantouka, A.; Felisberto, P.; Santos, P. Development and Testing of a Dual Accelerometer Vector Sensor for AUV Acoustic Surveys. *Sensors* **2017**, *17*, 1328. [[CrossRef](#)]
9. Stinco, P.; Guerrini, P.; Tesei, A.; Ferri, G.; Biagini, S.; Micheli, M.; Garau, B.; Lepage, K.D.; Troiano, L.; Grati, A. Passive Acoustic Signal Processing at Low Frequency With a 3-D Acoustic Vector Sensor Hosted on a Buoyancy Glider. *IEEE J. Ocean. Eng.* **2021**, *46*, 283–293. [[CrossRef](#)]
10. Chen, Y.; Zhang, G.Y.; Wang, R.; Rong, H.; Yang, B. Acoustic Vector Sensor Multi-Source Detection Based on Multimodal Fusion. *Sensors* **2023**, *23*, 1301. [[CrossRef](#)]
11. Yao, Z.; Hui, J.; Yin, W.; Yang, J. Four approaches to DOA estimation based on a single vector hydrophone. *Ocean Eng.* **2006**, *24*, 122–127+131. (In Chinese)
12. Yan, H.; Chen, T.; Wang, P.; Zhang, L.; Cheng, R.; Bai, Y. A Direction-of-Arrival Estimation Algorithm Based on Compressed Sensing and Density-Based Spatial Clustering and Its Application in Signal Processing of MEMS Vector Hydrophone. *Sensors* **2021**, *21*, 2191. [[CrossRef](#)] [[PubMed](#)]
13. Agarwal, A.; Agrawal, M.; Kumar, A. Higher-order-statistics-based direction-of-arrival estimation of multiple wideband sources with single acoustic vector sensor. *IEEE J. Ocean. Eng.* **2020**, *45*, 1439–1449. [[CrossRef](#)]
14. Wang, W.; Wang, C.; Han, M.; Sun, Q.; Zhang, X. Research on the application of vector hydrophone onboard an underwater glider. *Acta Armamentar* **2019**, *40*, 2580–2586. (In Chinese)
15. Wang, C.; Sun, Q.; Zhang, L.; Zhang, W.; Tian, D. Analysis of target detection capability of “G-Argo” acoustic buoy in South China Sea. *Appl. Acoust.* **2019**, *38*, 1025–1032. (In Chinese)
16. Wang, Y.; Cai, P.; Xu, D. Separation of multiple targets using a single vector sensor based on Hilbert-Huang transform. *Tech. Acoust.* **2006**, *25*, 376–380. (In Chinese)
17. Zhang, Q.; Wang, C.; Zhang, X.; Wang, W. Two-dimensional vector hydrophone and its application in Argo buoy platform. *Acta Armamentar* **2020**, *41*, 1566–1572. (In Chinese)
18. Chen, Y.; Wang, J.; Yu, Y.; Zhang, X.; Yang, B.; Zhu, Y. The W-disjoint orthogonality of underwater acoustic signals and underdetermined source counting for acoustic vector sensor. In Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–4.

19. Chen, Y.; Wang, W.; Wang, Z.; Xia, B. A Source Counting Method Using Acoustic Vector Sensor Based on Sparse Modeling of DOA Histogram. *IEEE Signal Process. Lett.* **2019**, *26*, 69–73. [\[CrossRef\]](#)
20. Han, X.; Piao, S.; Fu, J. Time-varying loudness feature extraction of the audition rhythm of ship radiation noise. *J. Harbin Eng. Univ.* **2020**, *41*, 487–492.
21. Taha, L.Y.; Abdel-Raheem, E. Blind Source Separation: A Performance Review Approach. In Proceedings of the 2022 5th International Conference on Signal Processing and Information Security (ICSPIS), Dubai, United Arab Emirates, 7–8 December 2022; pp. 7–8.
22. Sgouros, T.; Mitianoudis, N. A novel directional framework for source counting and source separation in instantaneous underdetermined audio mixtures. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 2025–2035. [\[CrossRef\]](#)
23. Wax, M.; Adler, A. Detection of the Number of Signals in Uniform Arrays by Invariant-Signal-Subspace Matching. *IEEE Trans. Signal Process.* **2022**, *70*, 1270–1281. [\[CrossRef\]](#)
24. Ke, X.; Zhao, Y.; Huang, L. On Accurate Source Enumeration: A New Bayesian Information Criterion. *IEEE Trans. Signal Process.* **2021**, *69*, 1012–1027. [\[CrossRef\]](#)
25. Zhang, Z.; Tian, Y.; Liu, W.; Chen, H. Enumeration for a Large Number of Sources Based on a Two-Step Difference Operation of Linear Shrinkage Coefficients. *IEEE Trans. Signal Process.* **2023**, *71*, 2283–2295. [\[CrossRef\]](#)
26. He, Z.; Cichocki, A.; Xie, S.; Choi, K. Detecting the number of clusters in n-way probabilistic clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2006–2021.
27. Arberet, S.; Gribonval, R.; Bimbo, F. A robust method to count and locate audio sources in a multichannel underdetermined mixture. *IEEE Trans. Signal Process.* **2010**, *58*, 121–133. [\[CrossRef\]](#)
28. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Yang, B.; Zhan, W.; Wang, P.; Chan, C.; Cai, Y.; Wang, N. Crossing or not? context-based recognition of pedestrian crossing intention in the urban environment. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 5338–5349. [\[CrossRef\]](#)
30. Yang, B.; Yang, J.; Ni, R.; Yang, C.; Liu, X. Multigranularity scenarios understanding network for trajectory prediction. *Complex Intell. Syst.* **2023**, *9*, 851–864. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.