1. Introduction

The existing two-step positioning method consists of two processes: parameter estimation and position calculation. The commonly used parameters mainly include time of arrival (TOA), time delay of arrival (TDOA), direction of arrival (DOA), and Doppler shift parameters [1–3]. The DOA parameter estimation is mainly based on array signal processing methods. The array signal processing methods have high accuracy, and they are simple to implement [4–7]. At present, the technology of DOA estimation based on array signal processing is relatively well developed, typical representatives are the MUSIC method, ESPRIT method, and root-MUSIC method [8–12].

Localization is another important step to obtain the position coordinates of the target sources. Localization is generally based on parameter estimation. It mainly uses the geometric relationship between the observation position coordinates and the positioning parameters to establish the cost function of the target sources, and the source position coordinates can be obtained by solving the cost function [13, 14]. At present, typical methods include the least square method, Chan method, and Taylor-series expansion method. The least square method is used to solve the position coordinates in [15], and the position coordinates obtained by this method are suboptimal, not global optimal. In order to improve the least square method, [16,17] give a double least square position method, namely, the Chan method, which can obtain noniterative closed-form solutions, but it is sensitive to parameter errors and has poor performance under a low signal-to-noise ratio (SNR). The Taylor-series expansion method uses the Taylor series criterion to convert the nonlinear cost function into the linear form, and the target source coordinates can be obtained through the iterative processes. The authors of [18–21] introduced the general steps of Taylor series expansion position methods and analyzed the performance. The Taylor series expansion method can obtain better positioning accuracy through approximation and iteration, but it is sensitive to the initial values. In order to improve the robustness of the Taylor
series expansion method, experts and scholars adopt different methods to optimize the initial values of Taylor series expansion. Li et al. [18] introduced the restriction conditions of the convergence region to ensure the convergence of iteration results. The authors of [19–21], respectively, adopted the steepest descent method and factor graph method to obtain a good initial value of iteration to improve the convergence probability of iteration process. Xiaowei et al. [22] combined the Chan method, Taylor-series expansion method, and Kalman filtering method to give a high-precision multipoint positioning method based on the TOAs. Location calculation is carried out by combining DOA parameters and TOA parameters or TDOA parameters in [23–27], which improve positioning accuracy. However, the observation stations are required not only to accurately obtain DOA parameters of target sources, but also to obtain TOA information.

The above-mentioned methods are mainly aimed at the single-source model, and the position precision is not high. In order to solve the DOA-based multisource location, this paper presents a multiscreening K-means clustering localization method. This method firstly establishes a cost function of the source position according to the coordinates of each observation position and DOA parameters and then solves the cost function to obtain a complete set of the source position coordinates. Because the obtained complete set contains “pseudotargets,” in order to eliminate the influence of the “pseudotargets,” the K-means clustering method is adopted to classify the real coordinates and the “pseudotargets.” To improve the positioning accuracy, screening samples are used as the input clustering data by the length of Euclidean distance, and the influence of fuzzy coordinates are gradually removed. Finally, the changeless clustering center can be obtained, which is the position coordinate of all target sources.

The rest of this paper is organized as follows: Section 2 introduces the localization calculation model based on DOA parameters; Section 3 presents the localization based on multiple screening K-mean clustering for multiple sources in detail and provides the step summary and the complexity analysis of this method; Section 4 presents simulation experiments; and Section 5 summarizes this work.

## 2. Localization Model

It is assumed that there exist \( D \) target sources transmitting plane waves to the measuring array. In order to calculate the position of the receiving signal source, the receiving station moves \( Q \) measuring positions. Each measuring position measures the DOAs of the target sources. The azimuth information of the target sources observed at the \( q \)th measuring position is \( \{ \theta_1^{(q)}, \theta_2^{(q)}, \ldots, \theta_D^{(q)} \} \). For the convenience of expression, a two-dimensional coordinate system is established, and all observation positions are in the first quadrant. To simplify the calculation, the line of the measuring array is always parallel to the \( x \) coordinate axis. In practice, the line of the array may be nonparallel to the \( x \) axis; generally, it can be converted into the corresponding angle between the incoming wave direction and the positive \( x \) axis through simple coordinate rotation. If there is no special explanation, the following azimuth angles all refer to the angles between the incoming wave direction and the positive \( x \) axis. According to the spatial geometric relationship, one target source can be located by any two observation positions, which is located at the upper side or the lower side of the line determined by two observation positions. The target source only has one real position, and the other one is the “pseudotarget.” The spatial geometric relationship between observation point coordinates and target source position coordinates is shown in Figure 1.

According to the geometric relationship of two-dimensional space,

\[
\begin{align*}
(x^{(p,q)} - x_p) \tan \theta_1^{(p,q)} &= y^{(p,q)} - y_p, \\
(x^{(p,q)} - x_p) \tan \theta_1^{(q)} &= y^{(p,q)} - y_q,
\end{align*}
\]

(1)

and

\[
\begin{align*}
(x^{(p,q)} - x_p) \tan \theta_1^{(p)} &= y_p - y^{(p,q)}, \\
(x^{(p,q)} - x_q) \tan \theta_1^{(q)} &= y_q - y^{(p,q)},
\end{align*}
\]

(2)

where \((x_p, y_p)\) and \((x_q, y_q)\) are the coordinates of the \( p \)th and \( q \)th observation positions; \( \theta_1^{(p)} \) and \( \theta_1^{(q)} \) are the \( p \)th and \( q \)th observation azimuths, respectively; \( \theta_1^{(p,q)} \) is the \( \theta_1 \) index and \( \theta_1^{(q)} \) represents the \( \theta_1 \) subscript.

The target position coordinates appear in pairs; that is, \( i \in \{1, 2\} \), \( j \in \{1, 2, \ldots, D\} \), and \( p, q \in \{1, 2, \ldots, Q\} \). According to the formulas (1) and (2), the position coordinates of the \( i \)th target source can be obtained as

\[
\begin{align*}
x_{(i,1)}^{(p,q)} &= \frac{y_p - y_q - x_p \tan \theta_1^{(p)} + x_q \tan \theta_1^{(q)}}{-\tan \theta_1^{(p)} + \tan \theta_1^{(q)}}, \\
y_{(i,1)}^{(p,q)} &= \frac{-y_p \tan \theta_1^{(q)} + y_q \tan \theta_1^{(p)} + (x_p - x_q) \tan \theta_1^{(p)} \tan \theta_1^{(q)}}{\tan \theta_1^{(p)} - \tan \theta_1^{(q)}},
\end{align*}
\]

(3)

and

\[
\begin{align*}
x_{(i,2)}^{(p,q)} &= \frac{y_p - y_q + x_p \tan \theta_1^{(p)} - x_q \tan \theta_1^{(q)}}{\tan \theta_1^{(p)} - \tan \theta_1^{(q)}}, \\
y_{(i,2)}^{(p,q)} &= \frac{y_p \tan \theta_1^{(q)} - y_q \tan \theta_1^{(p)} + (x_q - x_p) \tan \theta_1^{(p)} \tan \theta_1^{(q)}}{-\tan \theta_1^{(p)} + \tan \theta_1^{(q)}},
\end{align*}
\]

(4)

Through the above analysis, it can be seen that as for the symmetry of DOAs, there is position ambiguity when two station positions are used for location calculation; that is, any two observation positions can obtain a group of source coordinates according to the DOAs, which include a true
source location and a fuzzy source location. The fuzzy one is the “pseudotarget.”

3. Localization Based on Multiple Screening

K-Mean Clustering for Multiple Sources

3.1. Method Introduction. According to the above analysis, the position coordinate of a target source can be determined by using two observation stations with a fuzzy location. Therefore, three observation positions can be used to uniquely determine a source coordinate and achieve the purpose of eliminating the false targets and preserving the true targets. The positioning schematic with three observations is shown in Figure 2.

As shown in Figure 2, three stations can determine six positions, and three of them have overlapped with each other: \((x_{i,1}^{(p,q)}, y_{i,1}^{(p,q)})\), \((x_{i,2}^{(p,q)}, y_{i,2}^{(p,q)})\), and \((x_{i,1}^{(q,g)}, y_{i,1}^{(q,g)})\) are all real target locations in Figure 2. Red positions represent the “pseudotargets”; that is, \((x_{i,2}^{(p,q)}, y_{i,2}^{(p,q)})\), \((x_{i,1}^{(p,q)}, y_{i,1}^{(p,q)})\), and \((x_{i,2}^{(q,g)}, y_{i,2}^{(q,g)})\) are fuzzy positions. Since there are scenarios with multiple sources in the real environment, it is impossible to effectively match azimuth parameters by using the traditional three-station positioning method. In order to solve the problem of nonfuzzy solution of multitarget source position coordinates, the K-means clustering algorithm is introduced in this section. According to the information theory, the more useful the information obtained, the more accurate the source position coordinates estimated. Similarly, any two observation stations in a space can determine one group of source position coordinates, if all observation position coordinates are placed in the same two-dimensional space, the positions of true targets would be more and more concentrated, and the positions of pseudotargets would be more and more divergent. This is also the basis of the K-means clustering algorithm used in this paper, which can realize the multitarget position calculation without angle matching.

Firstly, the possible position coordinates of all targets are calculated according to formulas (1) and (2), and there are \(2D^2C_{Q}^2\) groups of position coordinates, where \(C_{Q}^2\) represents the number of selected stations from \(Q\) observation positions. Therefore, the initial input sample number of the K-means clustering method is \(D^2Q(Q–1)\); that is, the size of the complete coordinate set is \(D^2Q(Q–1)\). The K-means clustering method firstly takes any \(D\) position coordinates in the complete set as an initial clustering center, and one initial clustering center represents a class. Then, the rest coordinate positions are classified according to the Euclidean distance. The Euclidean distance is defined as follows:

\[
d((x_{i,1}^{(p,q)}, y_{i,1}^{(p,q)}), (x_{mid}, y_{mid})) = \sqrt{(x_{i,1}^{(p,q)} - x_{mid})^2 + (y_{i,1}^{(p,q)} - y_{mid})^2},
\]

where \((x_{mid}, y_{mid})\) is a clustering center, and \((x_{i,1}^{(p,q)}, y_{i,1}^{(p,q)})\) is the element in a complete coordinate set. After classification of all the elements of the complete coordinate set, the statistical average of the coordinates in the class is taken as the new clustering center, which is the updated clustering center. The general K-means clustering method is to repeat formula (5) for reclassification until the clustering center no longer changes. Actually, it is impossible to keep the clustering center completely unchanged. A smaller value should be selected as the threshold, and if the change of clustering center is smaller than this threshold, it means that the clustering method has converged.

According to the introduction of the localization model, it can be seen that the true coordinate positions in the complete coordinate set are overlapped or distributed more intensively, while the “pseudotarget” positions are dispersed. Therefore, if the dispersed coordinates are forcibly classified in the clustering process, the final position accuracy will be decreased. In order to reduce the influence of “pseudo-targets” on the positioning results, before repeating the reclassification, the complete set should be screened. In the screening process, the positions that are closer to the clustering center will be retained, and the farther ones will be discarded. As each of the two observation positions...
determines a true source position; that is, $C_Q^2$ common positions in the complete set are the true locations of the target sources, so each cluster center can only retain $C_Q^2$ closest coordinate positions and discard the rest. In this way, the final results of the output clustering centers are the $D$ target source coordinates. The flow chart of the method is shown in Figure 3.

3.2. Summary of Method Steps. This paper gives a DOA-based localization method with multiple screening K-Means clustering for multiple sources, which can realize the position solution of DOA parameters and is the core content of the two-step positioning technology. This method can not only realize the position solution of DOA parameters, but also can be used to realize the target source position solution of TOAs or DOAs, and TOAs. This method gives a location screening method when the number of sources or the number of observations are large, which effectively improves the position estimation accuracy. Based on the above analysis, the steps of this method are summarized as follows:

**Step 1.** According to the observation positions and DOA information at each observation position, the cost function of the source coordinates is established as shown in formulas (3) and (4), and the complete set of position coordinates of all the sources is obtained.

**Step 2.** Get the initial clustering center based on the complete coordinate set and input data into the K-means clustering method. Calculate the Euclidean distances between the elements of the coordinate set and the clustering center according to formula (5), and then, recluster according to the Euclidean distances.

**Step 3.** Update the clustering center, then recalculate the Euclidean distances between each sample coordinate in the complete set and the new clustering centers, reserving the coordinate positions of the smallest $C_Q^2$ Euclidean distances as the new input data and re-clustering with K-means method.

**Step 4.** Update the cluster centers, and determine whether the updated cluster centers change with respect to the cluster centers in Step 3. Repeat Step 3 if the cluster centers change. If the cluster centers are stable, output the cluster centers as the location coordinates of the target sources.

3.3. Complexity Analysis. The complexity of this algorithm mainly comes from two parts: the solution of complete coordinate set and the application of multiple screening K-means clustering method. The computation of complete coordinate set mainly includes formulas (3) and (4). The computational complexity is $O(16C_Q^2)$; that is, $O(8Q^2 - 8Q)$. The K-means clustering method is an iterative method. The number of iterations depends on the selection of initial values of the input samples. For the sake of measuring the computational complexity conveniently, the number of iterations is set to $Y$. The complexity of the iteration process is mainly concentrated on the calculation of Euclidean distances of formula (5). The complexity of calculating Euclidean distances in an iteration operation is $O(3D^2Q(Q - 1))$, and the computational complexity of $Y$ iterations is $O(3YD^2Q(Q - 1))$. In summary, the computational complexity of this method is $O((3YD^3 + 8Q)(Q - 1))$.

4. Simulation Experiments

In this section, the simulation experiments of the multiple screening K-means clustering localization method are given. In order to measure the performance of the proposed method, the definition of the root mean square error (RMSE) is given, which is the deviation between the estimated source position coordinates (the final output cluster centers) and the real source position coordinates. The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{TD} \sum_{j=1}^{T} \sum_{i=1}^{D} \left[ (x_{i,j} - \bar{x}_{i,j})^2 + (y_{i,j} - \bar{y}_{i,j})^2 \right]},$$  

where $(x_{i,j}, y_{i,j})$ is the real location coordinate of the $i$th source, $(\bar{x}_{i,j}, \bar{y}_{i,j})$ is the estimated location coordinate of the $i$th source by the $j$th Monte Carlo experiment, and $T$ is the Monte Carlo number. In order to verify the effectiveness of the proposed method, the position distribution scatter plots of the screening coordinates, the RMSE with angle deviation, and the RMSE performance with different number of observation points are given. The simulation conditions are set as shown in Table 1.
Simulation 1. Location distribution of input samples with each screening.

In this Monte Carlo experiment, location distribution of the initial complete coordinate set and coordinates distribution with first, second, and third screening are given. The angle deviation is 0.5 degree, and the simulation results are shown in Figure 4; it can be seen from the scatter distribution figures that after screening of initial values, the interference of pseudo targets is effectively eliminated and the degree of aggregation of input samples is improved.

<table>
<thead>
<tr>
<th>Simulation parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source coordinates</td>
<td>(8000, 4000), (5000, −3000)</td>
</tr>
<tr>
<td>Observation locations</td>
<td>(0, 1400), (6000, 0), (10000, 50), (16000, 0), (21000, 10000)</td>
</tr>
<tr>
<td>Monte Carlo number</td>
<td>T = 200</td>
</tr>
<tr>
<td>DOA deviation</td>
<td>0–0.5°</td>
</tr>
</tbody>
</table>

Figure 4: Scatter plot of position coordinates of the input samples. (a) Initial sample, (b) with first screening, (c) with second screening, (d) with third screening.
Simulation 2. Performance simulation of RMSE with different angle deviations.

Localization methods based on DOA parameters are sensitive to angle deviation because even if the angle deviation is small, it will cause large position deviation in a long distance, and it changes approximately linearly with the increase of distance. The RMSE performance of the position accuracy with DOAs deviation of $0^\circ - 0.5^\circ$ is given in this Monte Carlo experiment, as shown in Figure 5. Simulation results show that when the angle deviation is small, the position estimation accuracy is high. With the increase of the angle deviation, the RMSE shows an upward trend. Whether small angle deviation or large angle deviation, this method can effectively calculate the position coordinates.

Simulation 3. RMSE performance simulation with different numbers of observation locations.

The more the number of observation locations, the more the sample coordinates obtained and the more valuable the information used. According to the knowledge of information theory, the positioning accuracy will be improved with adding useful information. This Monte Carlo simulation experiments of the RMSE performance when the number of observation positions is varied from 3 to 10. The simulation results are shown in Figure 6. The simulation results indicate that when the number of observation location increases, the RMSE decreases. Moreover, the coordinates of observation locations have a great influence on the RMSE performance.

5. Conclusion

In order to achieve high-precision location of multiple sources without ambiguity based on DOA parameters, this paper introduces the $K$-means clustering algorithm into location measurement and presents a multisource localization method based on multiscreening $K$-means clustering. Firstly, the cost function of the target source position coordinates is established according to the observation position coordinates and the DOAs parameters, and the complete coordinate set of source positions is obtained by solving the cost function. In order to remove the “pseudotargets” in the complete set, the $K$-means clustering method is adaptively improved in this paper. The aggregation degree of the input sample positions is increased through multiple screening and gradually narrows the distance between the clustering center and the real target coordinates. Finally, the location measurement without ambiguity and high precision of multisources can be
realized. Complexity analysis and simulation experiments are given to verify the effectiveness of the proposed method. This method gives a location screening method when the number of sources or the number of observations are large, which can not only realize the position solution of DOA parameters, but also can be used to realize the target source position solution of TOAs or DOAs, and TOAs.

Data Availability

The data, which are produced by simulations, 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

This work was supported by the National Natural Science Foundation of China (grant no. 61401513).

References

Submit your manuscripts at
www.hindawi.com