Journal of Computers Vol. 35 No. 1, February 2024, pp. 1-16
doi: 10.53106/199115992024023501001

Wireless Sensor Networks Node Localization Algorithm Based on Range Optimization and Graph Optimization

Lei Wang, Ting-Ting Niu*, Wei-Hao Qiao, Song Cui

College of Computer Science and Technology, Henan Polytechnic University,
Jiaozuo, Henan Province, People's Republic of China
2769175036@qq.com, 2215251585@qq.com, hpucs2021@163.com, ntthpu@163.com

Received 3 April 2023; Revised 26 July 2023; Accepted 9 October 2023

Abstract. To address the problem of low localization accuracy in the node localization algorithms of wireless sensor networks (WSN) based on received signal strength indication (RSSI) ranging, a WSN node localization algorithm based on ranging optimization and graph optimization is proposed. In terms of RSSI ranging, the outliers are removed using the Grubbs method, and the data are processed using a moving average smoothing-Gaussian hybrid filter to establish a Bessel function ranging model to reduce the ranging error; in terms of node localization, the signal strength data are employed to construct a distance cost term, a cost function model is built based on these cost terms, and graph optimization is adopted to minimize this function. Then, the node position is estimated to minimize the overall observation error. Simulation results indicate that the proposed algorithm has higher ranging and localization accuracy than existing ranging and localization algorithms, and it can meet the requirements of node localization in large-scale WSN.

Keywords: wireless sensor networks, figure optimization, RSSI, filtering, grubbs method

1 Introduction

Over the past few years, Internet of Things (IoT) applications have exploded in various fields. Wireless Sensor Networks (WSN) are the most widely used end devices for sensing and collecting data in IoT. WSN consist of distributed, self-organized, energy-constrained wireless sensor nodes that automatically collect the data required for a specific application from a monitored area to improve performance and reduce the total cost of the monitoring system [1]. Because of the importance of WSN in various application areas such as industry, agriculture, military, environmental monitoring, healthcare, transport, and security, their research has received much attention. Localization is an important aspect of WSN application development and protocol design. In many applications, the data collected by sensors has low validity and value if the location information is not known, as a lack of location information may lead to misinterpretation of the data [2].

Satellite navigation systems such as the Global Positioning System (GPS) can solve the localization problem. However, for large-scale WSN, it is impractical and costly to equip each sensor node with a localization system module such as GPS. Meanwhile, satellite signals hardly work in indoor networks, and radio waves are strongly absorbed in underwater environments [3]. Therefore, localization techniques that can replace satellite navigation systems such as GPS in WSN have been a research hotspot. There are two main types of localization problems in WSN: one is the self-localization of sensor nodes, and the other is the use of WSN to obtain the location information of the target. The target can be a sensor node, a device, or an event [4]. Since localization system modules such as GPS are too expensive to be installed on every node in the network, only a small fraction of the nodes has a known location, and these nodes are called anchor nodes. Therefore, localization usually refers to estimating the position of the node to be located by interacting with some anchor nodes.

Localization techniques for WSN involve many aspects, including localization accuracy, power consumption, network availability and scalability, and latency. The advantage of low power consumption of sensor nodes limits the use of localization algorithms. Due to resource constraints such as limited energy and processing power of sensor nodes, it is critical to balance the localization accuracy and power consumption of WSN localization algorithms. Meanwhile, complex application scenarios will bring many thorny problems and increase the difficulty of localization, and solving these problems is the focus of WSN localization research. Therefore, developing fast, scalable, and robust localization algorithms with high localization accuracy is the focus and difficulty of WSN

* Corresponding Author

localization research.

To solve the above problems, this paper proposes a WSN node localization algorithm based on ranging optimization and graph optimization. This algorithm proposes a complete ranging scheme to improve ranging accuracy and improves localization accuracy through graph optimization. In the distance measurement scheme, the Grubbs method is used to remove outliers so that the RSSI data approximately conforms to the Gaussian distribution, and then the moving average smoothing-Gaussian mixed filter is employed to further process the data, and the Bessel function distance measurement model is constructed. In the node self-location method based on graph optimization, the distance cost items between nodes are established through signal strength data, and these cost items are utilized to construct the cost function model. Then, by minimizing this function through graph optimization, unknown nodes are obtained in the sense of least squares position estimate.

The rest of this paper is organized as follows: Section 2 presents the related work, Section 3 introduces the RSSI ranging scheme and its improvement method, Section 4 describes the proposed node localization algorithm based on graph optimization in detail, and the simulation results are provided in Section 5, and finally, Section 6 concludes this paper.

2 Related Works

At present, a variety of WSN localization algorithms have been proposed. According to whether to measure the distance between nodes, the localization algorithms are divided into ranging localization algorithms and non-ranging localization algorithms [5]. Typical ranging localization algorithms include those based on angle information [6], time difference of arrival [7], and received signal strength [8] (RSSI), where RSSI ranging does not require additional hardware devices and synchronization, so its cost and power consumption is lower. Non-ranging algorithms include the centroid algorithm, DV-Hop algorithm [9], etc. Compared with ranging localization algorithms, non-ranging localization algorithms have lower accuracy. Considering the difficulty of system deployment, localization accuracy, cost, etc., RSSI-based node localization technology is widely used in practical applications.

RSSI-based indoor localization technology is one of the classic methods in WSN node localization. This method is composed of an RSSI ranging model and a localization algorithm. Many scientific researchers have conducted in-depth studies on it. For the RSSI ranging scheme, Wang et al. [10] first used the Dixon test method to filter and process the data and then established the Shadowing model (DT-S), but the Dixon test method does not work when two suspicious values appear on the same side of the maximum (small) value, and the fitting effect of the shadowing model is poor. Liao et al. [11] used the particle filter to process data and build a shadowing model (PF-S), but the particle filter requires a large amount of sample data and the algorithm is relatively complex. Jiang et al. [12] used convex optimization to extract relationship trend items and established a Bessel ranging model (CO-B), but high-order calculations would cause problems such as high computational overhead and difficult solutions. Zhang et al. [13] employed Gaussian-median filtering to process data and established a shadowing model (GF-S), but the Gaussian filtering effect was poor due to the presence of outliers in the collected signal strength data.

For the RSSI localization algorithm, various algorithms have been proposed. Zhang et al. [14] combined the advantages of multiple filtering algorithms based on the ZigBee technology, and they proposed a new hybrid filtering (HF) algorithm. Then, the path loss model was estimated with the least square method, and finally, the actual position was obtained by weighted centroid trilateration. Nie et al. [15] developed a weighted localization method based on RSSI probability distribution and Bayesian estimation (PDBE). This method introduces the prior RSSI probability distribution into the weight calculation through Bayesian estimation, gives outliers a lower weight, reduces the impact of environmental noise and external uncertain factors on the localization accuracy, and takes the position with the largest weight as the localization result. Based on the received signal strength indication, Zhang et al. [16] proposed a joint algorithm, which introduces a triangular localization auxiliary algorithm into the K-nearest neighbor position fingerprint matching (K-NPFM) algorithm. First, the triangular localization is performed to obtain the reference area, and then the position fingerprint matching is performed for precise localization, which effectively improves localization accuracy. Li et al. [17] presented a weighted centroid location (WCL) algorithm with power value correction. The algorithm uses information such as the distance between the anchor nodes and the unknown node so that the anchor node can adaptively obtain the optimal weight coefficient to improve localization accuracy. Hu et al. [18] utilized the BP neural network to fit the RSSI data received by the anchor node to determine the parameter value of the loss model. Meanwhile, they used the geometric relationship to transform between nodes and improved the three-point localization method to the six-point centroid localization algorithm. Long et al. [19] combined RSSI ranging to transform the localization problem into an optimization problem of maximum likelihood estimation (MLE) and form a trust subdomain problem, which was then solved by the binary search method. Wang et al. [20] used the randomness and ergodicity of chaotic search and the multi-population of the chicken swarm algorithm to optimize the solution process of the particle swarm optimization (PSO) algorithm and improve the traditional centroid localization algorithm, thereby achieving higher localization accuracy. Guo et al. [21] used the PSO fitting signal attenuation curve (PSO-FSAC) algorithm combined with the weighted K nearest neighbor algorithm to remove poor experimental data and used a parametric model to locate the user's location. Yin et al. [22] solved the problem that the traditional sparrow search algorithm (SSA) is easy to fall into the local optimum by introducing a good point set and a levy flight strategy, which helped to accelerate the convergence speed, correct the position deviation of the target node, and improve localization accuracy. Alfawaz et al. [23] used the improved Rat Swarm Optimization (RSO) algorithm instead of the least square method to estimate the position of the node, and the communication radius was doubled to reduce the distance estimation error and improve localization accuracy.

Table 1 summarizes the basic properties of different localization methods. It can be seen from the comparison that although the above solutions have improved localization accuracy, there are still some shortcomings. The main goal of this paper is to overcome the shortcomings of the above schemes, improve ranging accuracy and localization accuracy through ranging optimization and localization optimization, and reduce energy loss, thereby meeting the needs of node localization in large-scale WSN. The main contributions of this paper can be summarized as follows:

(1) A complete set of ranging schemes is proposed, which first use the Grubbs method to remove outliers. After the outliers are removed, the RSSI data approximately conforms to the Gaussian distribution, and then moving average smoothing-Gaussian mixed filtering is utilized to further process the data to obtain the most realistic distance data.

(2) A novel distance conversion function is proposed, which uses the least square method to identify the specific parameters of the ranging model and uses the convex optimization algorithm to make the relationship between the RSSI value and the corresponding attenuation of the distance clearer, thereby improving the ranging accuracy.

(3) An accurate self-location method for nodes is proposed. This method establishes distance cost items between nodes through signal strength data. These cost items are employed to construct a cost function model, and this function is minimized through graph optimization to obtain unknown nodes in the sense of least squares position estimation, thereby improving localization accuracy.

Table 1. Comparison of different localization methods

Algorithm	Unknown node location	Time synchronization	Ranging method	Advantage	Shortcoming
HF	Distributed	need	RSSI	Consideration of actual sound velocity propagation characteristics to minimize errors	Ignores background noise and target speed variations
PDBE	Distributed	need	RSSI	Estimates are easily obtained	Time synchronization required
K-NPFM	distributed	need	RSSI	High localization accuracy and low average localization error	Complexity of the calculation process
WCL	centralized	unnecessary	RSSI	No time synchronization required, high localization accuracy	Large errors at short distances and complex calculations
BP	centralized	need	RSSI	Greater coverage and low communications overhead	Application scenarios have limitations
MLE	centralized	unnecessary	RSSI	The case of n circles not intersecting with a point can be solved	Cannot be used when anchor nodes are co-located

PSO	distributed	unnecessary	RSSI	High localization accuracy and low cost	Failure to consider the real environment
SSA	distributed	unnecessary	TOA	High localization accuracy and fast convergence	Problems with empty solution sets
RSO	distributed	unnecessary	TOA	High localization coverage with low error	The iterative process is computationally intensive

3 RSSI Ranging Solutions

The RSSI is a commonly used distance estimation method in WSNs. RSSI measures the strength of the radio signal between the transmitter and receiver and uses it to estimate the distance between them. However, in practical signal propagation environments, the RSSI value obtained by the receiver is easily affected by various interfering factors, including non-line-of-sight, weather, and sensor node height. In such scenarios, simply averaging multiple RSSI measurements and converting them into distance will result in large ranging errors. To address these issues, this paper proposes several methods to improve the accuracy of RSSI ranging.

Firstly, there are outliers in the RSSI values from multiple measurements due to interference factors, and this paper uses the Grubbs method to remove the outliers. Secondly, the RSSI data are filtered using a moving average smoothing-Gaussian mixture filter, as the outliers are approximately normally distributed. Finally, after the RSSI values from multiple measurements are processed, this paper establishes a Bessel function ranging model based on the literature [12], and based on this more accurate ranging model, a more effective reflection of the RSSI value decay relationship is achieved, and the ranging accuracy is improved. The overall flow chart for ranging is shown in Fig. 1.

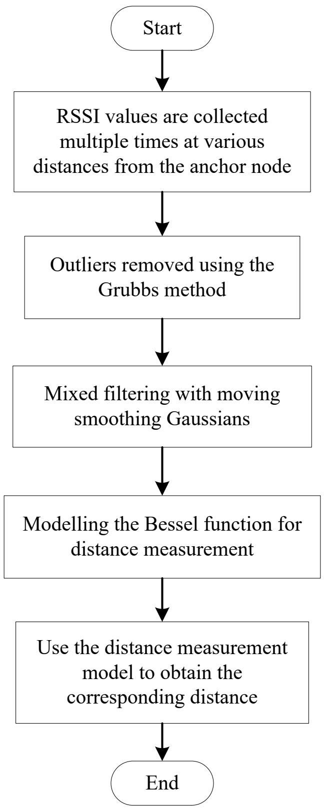

Fig. 1. The flow chart of distance measuring

3.1 Grubbs Method to Remove Outliers

At present, various statistical testing methods for judging outliers have been proposed, such as the Grubbs test method, Dixon method, skewness-kurtosis method, etc. However, all of these methods are susceptible to two major errors commonly in statistical hypothesis testing. Among these methods, the Grubbs test has the lowest probability of suffering from both types of errors. Therefore, the Grubbs test is preferred for removing outliers from data. The Grubbs test assumes that the data follows a normal distribution, and it works by comparing the distance between a suspect point and the mean of the sample to the standard deviation of the sample. If the distance exceeds a certain threshold, the suspect point is regarded as an outlier and is removed from the sample. The specific steps are as follows:

Step 1. Sorting the RSSI data collected multiple times in ascending order: $rssi_1 < rssi_2 \ldots < rssi_i < \ldots < rssi_n$.

Step 2. Calculate of the mean $\overline{rssi}$ and standard deviation S_{rssi}.

$$\overline{rssi} = \frac{\sum_{i=1}^{n} rssi_i}{n}, \tag{1}$$

$$S_{rssi} = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(rssi - \overline{rssi})^2}. \tag{2}$$

Step 3. Calculate the deviation: the difference between the mean and the minimum is $c_{\min} = \overline{rssi} - rssi_{\min}$ and the difference between the mean and the maximum is $c_{\max} = rssi_{\max} - \overline{rssi}$.

Step 4. Determine a suspicious value *sus*: compare $c_{\min}$ with $c_{\max}$, and if $c_{\min} > c_{\max}$, $rssi_{\min}$ is considered suspicious; otherwise, $rssi_{\max}$ is considered as suspicious.

Step 5. Calculate the G value: $G = (sus - \overline{rssi}) / Srssi$, where $sus - \overline{rssi}$ is the residual.

Step 6. Compared the value of G with the critical value $G(n)$ given in the Grubbs table, and if $G > G(n)$, the value is considered to be an outlier. The critical value $G(n)$ is related to the confidence probability P and the amount of RSSI data n, where P generally takes the value 0.95.

Step 7. Reject the RSSI outliers and repeat steps 1-7 until no outliers are detected.

3.2 Filter Processing

It was found that the random distribution of the RSSI data after removing the outliers approximately conforms to the Gaussian distribution, so using Gaussian filtering can effectively fit the distribution of RSSI values in real environments and thus eliminate the small probability of RSSI values. However, the use of Gaussian filtering only cannot eliminate the fluctuations in the data completely and process the data effectively. In contrast, the 9-point moving average smoothing filter can eliminate the effect of value fluctuations on the final measurement to a greater extent, but this method is less effective when the amount of RSSI data is small and fluctuating. Considering the complexity and feasibility of the ranging scheme and the complementary advantages of Gaussian filtering and moving average smoothing filtering, this paper adopts Gaussian filtering and moving average smoothing filtering to filter the RSSI data, and the filtering process is described below.

a) Build a Gaussian model as follows.

$$F(rssi) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{-\frac{(rssi-\mu)^2}{2\sigma^2}}, \tag{3}$$

$$\mu = \frac{1}{n_1}\sum_{i=1}^{n_1} rssi_i, \tag{4}$$

$$\sigma^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n}(rssi_i - \mu)^2. \tag{5}$$

where $rssi_i$ denotes the ith data in the RSSI data after removing the outliers; n_1 denotes the amount of RSSI data after removing outliers; μ is the mean value; σ is the standard deviation, which indicates the degree of data dispersion.

b) Excluding low probability values from the RSSI data. The interval of the normal distribution with a high probability of RSSI values is $(\mu - \sigma, \mu + \sigma)$, which has a probability of 68.26% to occur, so the RSSI values outside this interval are excluded from the RSSI data.

c) For the data processed by Gaussian filtering, a moving average smoothing filtering process is performed, the basic idea of which is that the RSSI value fluctuates up and down around a certain central value with small fluctuations. $rssiG_i$ is taken as the center, the mean value $rssi\ G_i'$ is found for k pieces of data before and after $rssiG_i$, and $rssi\ G_i'$ is substituted for $rssiG_i$, as shown in equation (6):

$$rssiG_i' = \frac{1}{2k}\sum\nolimits_{i=-k}^{k} rssiG_{i+1}, \tag{6}$$

where $rssiG_i$ denotes the ith RSSI value in the RSSI data after Gaussian filtering; the larger the value of k, the better the smoothing effect, and the greater the distortion. k is set to 9 in this study.

d) Finally, the output of the Gaussian-moving average smoothing filter $rssi_{filter}$ is obtained.

$$rssi_{filter} = \frac{1}{n_2}\sum\nolimits_{i=1}^{n_2} rssiGY_i, \tag{7}$$

where n_2 denotes the amount of filtered RSSI data, and $rssiGY_i$ denotes the ith RSSI value in the filtered RSSI data.

3.3 Building a Bessel Function Distance Measurement Model

The Bessel function, as a special function, has a wide range of applications in physics and is crucial for solving various potential field and electromagnetic wave propagation problems. Compared with the traditional empirically constructed models, the ranging model based on the Bessel function has practical physical significance. Accordingly, a ranging model based on the Bessel function was proposed and experimentally verified in the literature [12]. The results demonstrate that the Bessel function-based ranging model has a significantly better curve-fitting effect than the Shadowing model and the segmented ranging model, and it has a higher ranging accuracy. In addition, the model has less influence on the interference factors such as non-line-of-sight, weather, and sensor node height in the signal propagation environment, and it can still maintain high accuracy under complex propagation environments. In this regard, the Bessel function-based ranging model is used in this paper, and it is shown in equation (8).

$$RSSI(d) = k_1 J_0(k_2 d) + k_3 J_1(k_4 d), \tag{8}$$

where $RSSI(d)$ denotes the signal strength in dBm obtained by the receiver end at any distance d from the transmitter; $J_0(x)$ and $J_1(x)$ are the 0th and 1st-order functional expressions of the nth-order first-class Bessel function; k_1, k_2, k_3, and k_4 are parameters to be determined in the model.

Based on the literature [12], the Bessel function distance measurement model is established in the following process:

a) The collected series of RSSI values and the corresponding distance data are pre-processed by the Grubbs method and the filtering method. In this way, the quality and reliability of the data are improved, and some unavoidable errors and disturbances are eliminated.

b) According to the processed data set, more RSSI values can be obtained with the corresponding change in the distance decay relationship. As the test distance becomes larger, the RSSI value fluctuation magnitude becomes larger, and there will be abnormal RSSI values, making it difficult to clarify the RSSI value and the distance correspondence decay relationship. Therefore, the convex optimization algorithm is employed to make the RSSI value and distance correspondence decay relationship clearer.

c) The least squares algorithm is used to identify the parameters k_1, k_2, k_3, and k_4, in the ranging model to determine the unknown parameters and obtain the specific ranging model expressions.

4 Graph Optimization Based Node Localization Algorithm

4.1 Foundations of Graph Optimization

Graph optimization is essentially a non-linear least squares optimization process, which has been proven to have better accuracy than Kalman and particle filters as an alternative to filtering methods, and it is now widely used in simultaneous localization and map building (SLAM) [24]. Graph optimization $G = \{V, E\}$ mainly consists of nodes $V$and edges E. The quantities to be estimated are taken as vertices of the graph, and the inter-observations to be estimated are taken as edges connecting the vertices. The estimation is a process of finding the best vertices in the state space that are in best agreement with the overall observations of the system.

The main process of chart optimization is as follows.

(1) Establish a cost function based on the observation model, which is generally in the form of a sum of squares.

(2) Find the minimum value of this cost function by using non-linear least squares optimization methods such as the second-order gradient method, Gauss-Newton method, Levenberg-Marquardt method, etc., which in turn find the quantity to be estimated. The general form of the cost function is as follows.

$$E = \sum_i e_i^T \Omega_i e_i, \tag{9}$$

where e_i denotes the error between the quantity to be estimated and the observation; Ω_i denotes the weighting factor on this error, the larger the weighting factor, the less noisy the observation corresponding to the error term e_i ; the subscript $i$denotes the traversal of all observations. Generally, the error term e_i is a function of the quantity to be estimated s_i and its corresponding observation m_i .

$$e_i = g(s_i, m_i). \tag{10}$$

Many mature runtime libraries are available to implement graph optimization, such as G2O and other runtime libraries. As graph optimization is well established, this paper describes the cost function establishment and the localization method in the context of WSN node localization.

4.2 Cost Function Establishment

The RSSI data collected by node several times is processed according to Sections 3.1 and 3.2 to obtain the received signal strength of this node, and then the distance $d_{i,j}$between it and the transmitting signal node j is obtained by equation (8). Based on this distance, the error function can be obtained as follows:

$$e_{i,j} = d_{i,j} - \sqrt{(x_i - x_j)^2 + (y_i + y_j)^2}, \tag{11}$$

where (x_i, y_i) is the position of the ith unknown node, and (x_j, y_j) is the position estimate of the jth unknown or anchor node. At this point, the corresponding cost function can be expressed as follows.

$$C = \sum_{i,j} e_{i,j}^2, \Omega_{i,j}, \tag{12}$$

where $\Omega_{i,j}$ is a weighting factor for the error $e_{i,j}$ difference; $\Omega_{i,j}$ is a normalized weight taken according to the inverse of the corresponding distances of all edges in the graph optimization, as the measured distance error increases with the distance between sensor nodes. The larger the distance between sensor nodes, the lower the confidence level, and the weight should be as smaller; conversely, the smaller the distance between sensor nodes, the higher the confidence level, and the corresponding weight should be larger.

By minimizing equation (12), the distance observation information between nodes can be fused, and the position estimation of each unknown node with the smallest overall observation error can be obtained.

4.3 Node Self-Localization Methods

In the context of large-scale WSN, the nodes are generally categorized as anchor nodes and unknown nodes. Typically, anchor nodes obtain their precise positions through methods such as high-precision GPS or manual measurements and serve as reference points with fixed positions in graph optimization techniques [25]. When global graph optimization is performed to determine the locations of all unknown nodes simultaneously, this incurs a significant computational burden due to the huge number of operations involved and the low performance of node operations, so it is impractical to use this method in WSN that experience frequent node movements. In practice, anchor nodes are typically deployed first, followed by the continuous deployment of unknown nodes. To ensure that each unknown node obtains its position immediately after deployment, this study proposes a local graph optimization strategy. Specifically, when locating an unknown node m in the WSN, all nodes with known positions within its communication radius r, including both anchor nodes and unknown nodes with position estimates, are identified. This enables only the nodes within the communication range of node m to participate in the graph optimization process, thereby avoiding the generation of large graphs that require extensive computational resources. Consequently, there are two scenarios in the overall WSN node localization process.

a) At the early stage of node localization, nodes with known positions within the communication radius r of node m are all anchor nodes. In this regard, node m is first taken as the vertex to be estimated, and then the corresponding error function between node m and each anchor node is established based on equation (11); finally, the cost function is constructed based on equation (12), and graph optimization is performed to obtain the position estimate of node m. At this point, using graph optimization to solve for the position of node m is similar to using the Levenberg-Marquardt method for weighted least squares problems, as the gradient descent method set up in this paper for graph optimization is the Levenberg-Marquardt method. The literature [26] has shown that the localization error is minimized when the number of anchor nodes involved in the least squares method is 4. Therefore, the $N = 4$ anchor nodes closest to node m are selected to participate in the graph optimization process, as shown in Fig. 2.

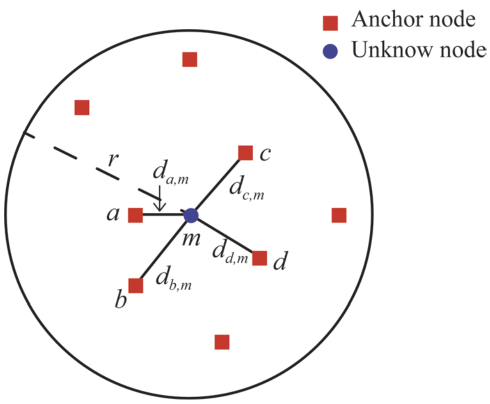

Fig. 2. The schematic diagram of network node localization situation 1

b) In the middle and late stages of localization, the nodes with known locations within the communication range of node m generally include anchor nodes and unknown nodes, for which location estimates have been obtained. The node localization process at this point consists of the following steps.

Step 1: Assume that there are r unknown nodes P_i ($i = 1, 2, \ldots, r$) within the communication range of node m for which position estimates have been obtained, and construct a data table in node m with node P_i and its corresponding number of connected edges n_i .

Step 2: Obtain the nearest N nodes to node m (which may include anchor nodes and unknown nodes with position estimates), such as the four nodes connected by black lines in Fig. 3. After the nodes are obtained, if they contain unknown nodes with position estimates, the number of edges connected to the corresponding nodes in the table is added by 1. The node m and the unknown nodes with position estimates among the obtained nodes are taken as the vertices to be optimized in the graph, the anchor node is taken as a fixed node, and the distance observations between node m and the N nodes are taken as edges.

Step 3: If there is an unknown node with a position estimate among the obtained N nodes, such as P_1, P_2, and P_3 in Fig. 3, cycle through the $N - n_i$ nodes closest to each node, and if there is an unknown node or node m with N connected edges among the first $N - n_i$ nodes, skip this node and find the next closest node, such as the node connected by the grey line in Fig. 3. After the node is obtained, if it contains an unknown node with position estimate, add the number of connected edges of the corresponding node in the table by 1, and update the number of connected edges of nodes P_1, P_2, and P_3 in the table to N. Then, the unknown node with position estimate among the acquired nodes is taken as the vertex to be optimized in the graph, the anchor node is taken as a fixed node, and nodes P_1, P_2, and P_3 are taken as edges with their corresponding $N - n_i$ inter-node distance observations taken as edges.

Step 4: If there are still unknown nodes in the nodes obtained in the third step, process them in the same way as Step 3 until there are no unknown nodes in the obtained nodes.

Step 5: Finally, graph optimization is performed on the constructed undirected graph to obtain the latest position estimates of node m and each unknown node in the graph for which a position estimate has been obtained.

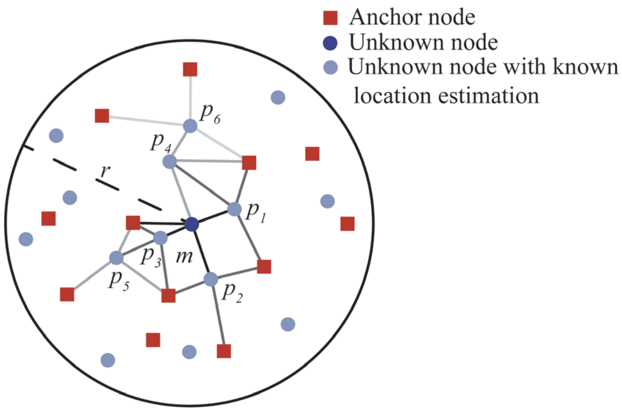

Fig. 3. The schematic diagram of network node localization situation 2

5 Verification of Simulation Experiments

The experiment consists of two main parts: feasibility and ranging performance analysis of the ranging scheme proposed in this paper, and localization performance analysis of the graph optimization-based node self-localization method. Both parts use MATLAB 2019a software to construct the simulation platform and compare the proposed algorithm with similar algorithms.

5.1 Feasibility and Performance Analysis of Ranging Solutions

Data acquisition was conducted in a 33 m corridor indoors. The specific acquisition plan was as follows: the total test distance was 30 m, and the wireless signal strength data was collected every 0.5 m using a Zigbee module model E18-MSI-IPX. There were 60 sampling points in total, and the wireless signal strength values were collected 300 times at each sampling point. The parameter setting of the Zigbee module is shown in Table 2.

Table 2. The parameter setting of the Zigbee module

Parameter name	Parameter values
Transmitting power	-20dBm
Launch interval	1ms
Number of RSSI acquisitions	300
Range of distance variation	0~30m

Then, the data collected from various sampling points are subjected to the removal of outliers and filtering. To verify the feasibility of these methods, the data collected from a sampling point 4 m away from the transmitting node is processed, and the results are presented in Fig. 4. The figure reveals that the use of the Grubbs method effectively removed the outliers in the original data. Meanwhile, using a moving average-smoothed Gaussian mixture filter to process the RSSI data demonstrated better results than using a moving average-smoothed filter or Gaussian filter alone.

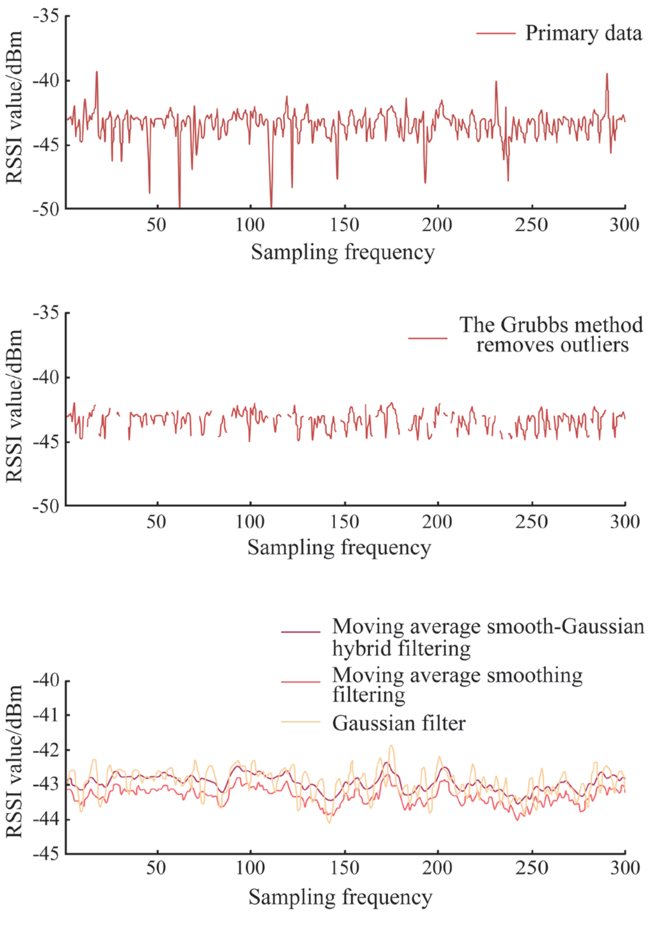

Fig. 4. Outlier removal and filtering

After processing the data collected at each location, based on the literature [12], the relational trend terms were first extracted using the convex optimization algorithm, and then the unknown parameters in equation (6) were identified using the least squares algorithm based on the convex optimization processed data, and the identification results were k_1 = -32.3, k_2 = 0.095, k_3 = 150.6, and k_4 = 0.042. As shown in Fig. 5, the convex optimization made the overall variation exhibit an obvious decay relationship and reduced the influence of fluctuating RSSI values when determining the unknown parameters of the Bessel ranging model, and the Bessel ranging model obtained a better fitting effect.

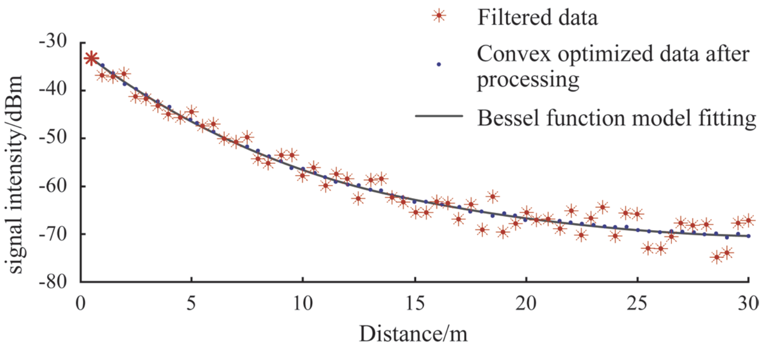

Fig. 5. The convex optimization of the processed data and the Bessel ranging model

To verify the performance of the distance measurement scheme proposed in this paper, data was collected at distances of 1-10m from the signal transmitter node in increments of 1m. This was repeated 100 times, and the average value was taken as the final experimental result. In order to evaluate the performance of our proposed distance estimation scheme, a comparison was made with several existing RSSI-based To verify the performance of the distance measurement scheme proposed in this paper, data was collected at distances of 1 ~ 10 m from the signal transmitter node in an increment of 1m. This data collection process was repeated 100 times, and the average value was taken as the final experimental result. To evaluate the performance of our proposed distance estimation scheme, a comparison was made with several existing RSSI-based distance estimation schemes including RSSI-DT-S proposed in the literature [10], RSSI-PF-S proposed in the literature [11], RSSI-CO-B proposed in the literature [12], and RSSI-GF-S proposed in the literature [13]. The comparison results are illustrated in Fig. 6.

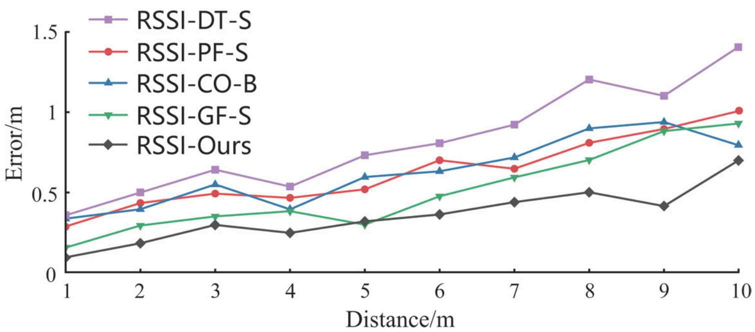

Fig. 6. The performance comparison of five ranging solutions

According to the result in Fig. 6, the measurement errors of all distance measurement schemes increase with the distance from the signal transmission node. This is because the RSSI values gradually decay due to environmental factors as the distance from the signal transmission node increases, leading to a reduction in the number of effective RSSI signals received by the localization node. Therefore, the errors of all five compared RSSI distance measurement algorithms increase. By contrast, the proposed distance measurement scheme in this study exhibits the smallest absolute difference between the measured distance and the actual distance among the five distance measurement algorithms, and the average error within 10 m is about 0.359 m, demonstrating superior ranging performance.

5.2 Localization Performance Analysis

The proposed localization algorithm is verified by MATLAB 2019a simulation software. To obtain good experimental results, each experiment was independently run 100 times in the same environment, and the average value is taken as the final result of the simulation experiment. The specific experimental parameters are shown in Table 3. In the experiment, the signal propagation attenuation model adopts the Bessel function model. The simulation area is a cube area of 100 m × 100 m × 100 m. The specific deployment strategy is: first distribute the anchor nodes evenly in the experimental area, and then continuously deploy unknown nodes. During the placement of unknown nodes, each time a node is placed, its position is obtained immediately.

Table 3. Simulation experiment parameter setting

Parameter name	Value
Simulation area	100m*100m*100m
Number of anchor nodes	10-35
Total number of nodes	50-100
Communication radius	30m-50m
Reference distance d_0	1m
Reference range signal stre $P(d_0)$	-15dbm
Number of experiments	100

To verify the performance of the proposed algorithm in terms of localization, the proposed algorithm is compared with the traditional RSSI localization algorithm in the literature [8], the RSSI-HF localization algorithm in the literature [14], and the RSSI-PSO localization algorithm in the literature [20]. The localization accuracy of the four algorithms is analyzed by comparing the effects of the number of anchor nodes and communication radius on the average localization error. The distribution of WSN in the simulation experiment is shown in Fig. 7, and the effect graph of the obtained error is shown in Fig. 8.

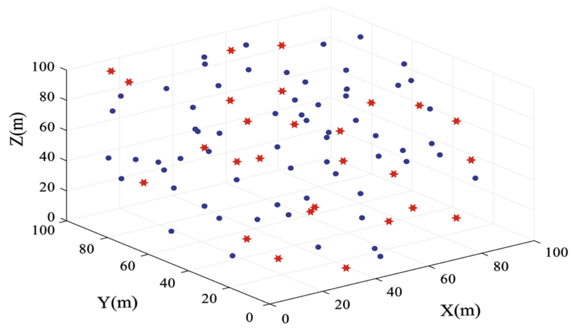

Fig. 7. The 3D distribution of WSN nodes

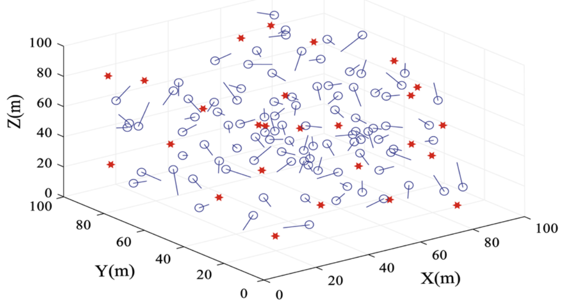

Fig. 8. The 3D node localization error effect

Additionally, the average localization error $e$and the normalized average localization error $\overline{e}$ are used as localization performance indicators, which are defined in equation (13).

$$\begin{cases} e = \dfrac{1}{KN}\sum_{j=1}^{N}\sum_{i=1}^{K}\left|p-\hat{p}\right| \\ \overline{e} = \dfrac{e}{r} \end{cases}, \tag{13}$$

where p and $\hat{p}$ denote the real and estimated positions of the unknown nodes respectively; K is the number of unknown nodes; N is the number of simulations, and $N = 100$; r is the node communication radius.

The influence of the number of anchor nodes on the average localization error. Firstly, the normalized average localization error changes with the proportion of anchor nodes in sensor nodes δ. Setting r = 35m, the range of the anchor node ratio is 0.1-0.5, and the results are shown in Fig. 9.

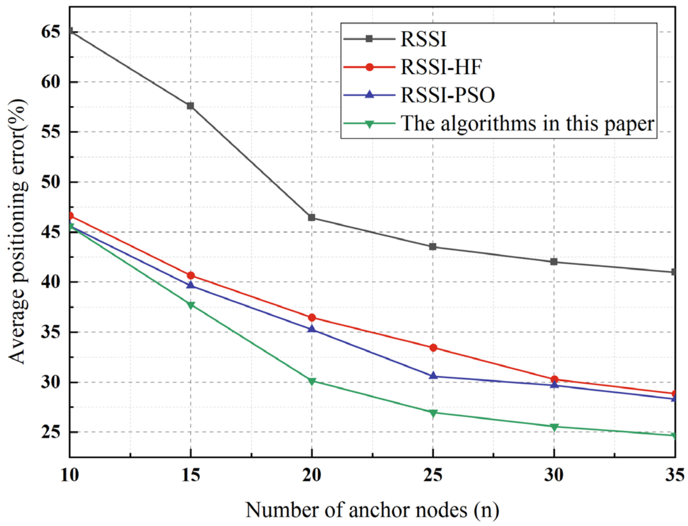

Fig. 9. The curve of normalized average localization error with the proportion of anchor nodes

As shown in Fig. 9, the normalized average localization errors of all four localization algorithms show a decreasing trend with the increase in the proportion of anchor nodes. This is because the more anchor nodes, the more distance information is provided within the communication radius, and the higher the probability of obtaining ranging data with small errors, and thus the higher the localization accuracy. Meanwhile, when the proportion of anchor nodes increases to a certain level, the error decline tends to slow down. Compared with the traditional RSSI localization algorithm, RSSI-HF localization algorithm, and RSSI-PSO localization algorithm, the average localization accuracy of the localization algorithm proposed in this paper is improved by 20.81%, 6.38%, and 5.51%, respectively. These results indicate that the proposed algorithm can use fewer anchor nodes to achieve a better localization effect, thereby reducing the deployment cost.

Influence of communication radius on average localization error. Next, the average localization error is analyzed as a function of the communication radius. Setting $\delta = 0.2$, and the results are shown in Fig. 10.

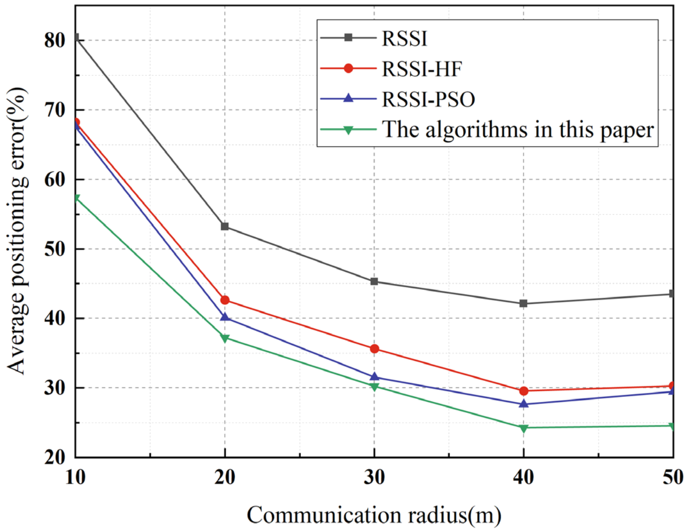

Fig. 10. The average localization error curve with communication radius

As shown in Fig. 10, the algorithm proposed in this paper performs significantly better than the other three algorithms at different communication radii. Compared with the original RSSI algorithm, RSSI-PSO algorithm, and RSSI-SSA algorithm, the average localization errors of the proposed algorithm are reduced by 22.62%, 7.21%, and 6.81%, respectively. When r < 40m, the average localization error of all four algorithms shows a decreasing trend; when r > 40m, the average localization error of the traditional RSSI localization algorithm, RSSI-HF localization algorithm and RSSI-PSO localization algorithm shows an increasing trend. This is due to the increase in the communication radius and because the anchor nodes that are farther away are involved in the localization. The further the distance between the nodes, the larger the ranging error, which in turn leads to a larger average localization error. By contrast, the algorithm proposed in this paper gradually reduces the average localization error as the communication distance increases. This is because the larger the communication radius, the greater the number of nodes involved in graph optimization, and the larger the graph size, the higher the stability of the graph and the smaller the overall error. However, with the increase in the graph size, the computing power rises, and the comprehensive computing power and localization accuracy also increase, so the communication radius is generally set to 40 m.

Comparison of the calculation time of localization algorithms. Finally, the calculation time of several localization algorithms is compared. When the node communication radius is set to 40 m, the proportion of beacon nodes is 30%, and the total number of nodes is 60, 70, 80, and 90 respectively, the calculation time of the algorithm is the average value of running 100 times. The simulation time of the four algorithms is shown in Table 4. As the number of nodes increases, the calculation time of the algorithm increases. Meanwhile, the traditional RSSI algorithm takes the shortest time among the four algorithms. While improving the precision of the optimization algorithm, the complexity and running time of the algorithm are increased. The RSSI-HF localization algorithm only uses several filtering algorithms to optimize the ranging stage, so the calculation time of the algorithm increases less than that of the traditional RSSI localization algorithm. The RSSI-PSO localization algorithm and the localization algorithm proposed in this paper optimize not only the ranging phase but also the localization phase, and the calculation time of the algorithm increases more than the traditional RSSI localization algorithm. Although the calculation time of the proposed algorithm has increased, it is shorter than that of RSSI-PSO, and its accuracy is the highest among the four algorithms.

Table 4. Comparison of average computing time of localization

Algorithm	Average computing time/s			
	The number of nodes is 60	The number of nodes is 70	The number of nodes is 80	The number of nodes is 90
RSSI	0.19	0.28	0.36	0.45
RSSI-HF	0.26	0.37	0.42	0.49
RSSI-PSO	0.49	0.56	0.64	0.68
The algorithms in this paper	0.31	0.37	0.48	0.54

6 Conclusion

In this paper, for the problem that the RSSI ranging-based node localization algorithm for WSN has low localization accuracy and cannot meet the demand of real applications, a WSN node localization algorithm based on ranging optimization and graph optimization is proposed. Firstly, the data are processed using the Grubbs method and moving average smoothing-Gaussian hybrid filtering, then a Bessel function-based ranging model is established, and finally, the graph optimization-based node localization algorithm is used to obtain an estimate of the unknown node position. The results of simulation experiments indicate that outliers can be effectively eliminated using the Grubbs method. The moving average smoothing-Gaussian hybrid filtering can effectively remove large jump values and obtain a set of stable and smooth data. Meanwhile, the Bessel function ranging model is constructed, which can fit the RSSI-distance curve better and obtain higher ranging accuracy. Besides, by using graph optimization to make full use of the inter-node distance observation data, an unknown node position estimation that minimizes the overall observation error is obtained. Compared with existing ranging schemes and localization algorithms, the algorithm proposed in this paper has significant improvement in both ranging and localization accuracy. However, this paper only considers the node localization of static targets in 3D space, and the application of the localization algorithm based on graph optimization to dynamic 3D node localization needs to be studied in future work.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 52174109), the Science and Technology Project of Henan Province (No. 212102210092), and the Key Research Foundation of Colleges and Universities in Henan Province (No. 20A520015).

References

[1] S. Chakraborty, N.K. Goyal, S. Soh, On area coverage reliability of mobile wireless sensor networks with multistate nodes, IEEE Sensors Journal 20(9)(2020) 4992-5003.
[2] J. Luo, Y. Yang, Z. Wang, Y. Chen, Localization algorithm for underwater sensor network: A review, IEEE Internet of Things Journal 8(17)(2021) 13126-13144.
[3] M. Li, F. Jiang, C. Pei, Review on positioning technology of wireless sensor networks, Wireless Personal Communications 115(1)(2020) 2023-2046.
[4] C.-Y. Wei, H.-Y. Pan, Adaptive zone–assisted iterative localization in energy-efficient wireless sensor networks, IEEE Sensors Journal 21(23)(2021) 27186-27194.
[5] H.-M. Shan, C.-X. Chen, RSSI gaussian filter-based artificial bee colony localization algorithm, Chinese Journal of Sensors and Actuators 34(7)(2021) 979-983.
[6] P. Sinha, I. Guvenc, Impact of antenna pattern on TOA based 3D UAV localization using a terrestrial sensor network, IEEE Transactions on Vehicular Technology 71(7)(2022) 7703-7718.
[7] A. Diewald, B. Nuss, M. Pauli, T. Zwick, Arbitrary angle of arrival in radar target simulation, IEEE Transactions on Microwave Theory and Techniques 70(1)(2022) 513-520.
[8] L.-F. Duan, S. Qin, Q. Wan, A Hybrid Localization Algorithm Based on RSSI Assisted Precise Distance Measurement, Journal of University of Electronic Science and Technology of China 48(3)(2019) 331-335.

[9] J. Cheng, Y.-L. Dong, J.-H. Chen, Z.-H. Liu, An improved three-dimensional DV-Hop algorithm with continuous hop value, Acta Electronica Sinica 48(11)(2020) 2122-2130.
[10] J.-Q. Wang, Y. Dai, Q.-F. Lei, Indoor positioning algorithm based on Dixon test filtering RSSI, Chinese Journal of Sensors and Actuators 34(1)(2021) 118-123.
[11] Y.-Q.-Z. Luo, W. Liao, RSSI range optimization newton localization algorithm based on particle filter, Instrument Technique and Sensor 5(6)(2017) 116-119+124.
[12] Y.-Y. Jiang, Y. Yu, M.-L. Shi, Y.-B. Liu, RSSI ranging method of Bessel function ranging model, Chinese Journal of Sensors and Actuators 33(2)(2020) 279-285.
[13] Y.-L. Zhang, J. Wang, H.-Z. Han, D. Yang, Research and precision analysis of indoor ranging model based on Bluetooth, Science of Surveying and Mapping 46(3)(2021) 1-7+66.
[14] Z.-C. Zhang, X.-E. Li, G.P. Gu, Y. Zhang, Indoor positioning algorithm based on ZigBee hybrid filtering RSSI, Transducer and Microsystem Technologies 42(5)(2023) 126-129+134.
[15] D.-W. Nie, H. Zhu, F. Wu, X.-F. Han, Weighted positioning method based on RSSI probability distribution and Bayesian estimation, GNSS World of China 47(2)(2022) 52-59.
[16] L.-C. Zhang, L.-W. Gu, X.-M. Li, X.-M. Mu, T. Yang, J. Tang, Design and implementation of indoor 2D positioning system based on RSSI joint algorithm, GNSS World of China 46(6)(2021) 78-83.
[17] C.-S. Li, Q.-L. Kong, T.-F. Wang, W.-H. Fang, L.-G Zhang, Simulation and accuracy analysis of the improved RSSI centroid location algorithm, Journal of Navigation and Positioning 10(1)(2022) 48-52.
[18] B.-B. Hu, X.-J. Ni, An Improved Indoor Localization Algorithm of Centroid with RSSI, Computer Technology and Development 27(9)(2017) 133-136.
[19] P. Long, J. He, RSSI ranging-based approximate maximum-likelihood estimator node localization algorithm, Journal of Navigation and Positioning 10(4)(2022) 187-191.
[20] G.-Y. Wang, L.-Y. Wang, H.-X. Lu, RSSI-based Centroid localization algorithm optimized by hybrid swarm intelligence algorithm, Computer Science 46(9)(2019) 125-129.
[21] H. Guo, H. Li, J. Xiong, M. Yu, Indoor positioning system based on particle swarm optimization algorithm, Measurement 134(2019) 908-913.
[22] L. Yin, D. Gu, F. Liu, Improved sparrow search algorithm based DV-Hop localization in WSN, Chinese Journal of Sensors and Actuators 17(5)(2021) 670-675.
[23] O. Alfawaz, W. Osamy, M. Saad, A.M. Khedr, Modified rat swarm optimization based localization algorithm for wireless sensor networks, Wireless Personal Communications 130(3)(2023) 1617-1637.
[24] G. Grisetti, R. Kümmerle, C. Stachniss, W. Burgard, A tutorial on graph-based SLAM, IEEE Intelligent Transportation Systems Magazine 2(4)(2010) 31-43.
[25] Y. Zhang, L. Leng, Graph optimization based indoor positioning method using bluetooth low energy beacons, Journal of Electronic Measurement and Instrumentation 33(6)(2019) 45-50.
[26] Y. Liu, B.-X. Xiao, B.-Q Yin, Quadrilateral weighted centroid localization algorithm based on RSSI of correction, Journal of Electronic Measurement and Instrumentation 34(10)(2020) 107-114.