Long Cheng^{1*}, and Jiayin Guan¹

Department of Computer and Communication Engineering, Northeastern University, Qinhuangdao 066004, Hebei Province, China. [e-mail: chenglong@neuq.edu.cn, 2172159@stu.neu.edu.cn] *Corresponding author: Long Cheng

Received June 16, 2023; revised September 21, 2023; accepted October 26, 2023; published November 30, 2023

Abstract

As an important technology of the internetwork, wireless sensor network technique plays an important role in indoor localization. Non-line-of-sight (NLOS) problem has a large effect on indoor location accuracy. A location algorithm based on improved particle filter and directional probabilistic data association (IPF-DPDA) for WSN is proposed to solve NLOS issue in this paper. Firstly, the improved particle filter is proposed to reduce error of measuring distance. Then the hypothesis test is used to detect whether measurements are in LOS situations or NLOS situations for *N* different groups. When there are measurements in the validation gate, the corresponding association probabilities are applied to weight retained position estimate to gain final location estimation. We have improved the traditional data association and added directional information on the original basis. If the validation gate has no measured value, we make use of the Kalman prediction value to renew. Finally, simulation and experimental results show that compared with existing methods, the IPF-DPDA performance better.

Keywords: Wireless Sensor Network, Indoor Localization, NLOS, Improved Particle Filter, Directional Probabilistic Data Association.

1. Introduction

Wireless sensor network (WSN) is an important technology of the Internet, which is more and more widely used in daily life [1]. Indoor positioning is becoming more and more important in WSN. Indoor localization technology based on distance general uses time difference of arrival (TDOA), received signal strength (RSS), time of arrival (TOA), angle of arrival (AOA), [2] [3] [4], etc.

In this paper, the Euclidean distance measurements are obtained between mobile node (MN) and beacon nodes (BNs) through TOA method. In a NLOS case, by reason of the presence of barriers, the measured time will be longer than in LOS. This will lead to measured distance to be larger than Euclidean distance, which is called NLOS error [5]. In practice, all kinds of indoor objects can be regarded as obstacles. Therefore, the designed algorithm is necessary to reduce the NLOS errors.

There are some traditional mitigate NLOS errors algorithms, Particle Filter (PF), Average Filter (AF), Gaussian Filter (GF), Kalman Filter (KF), Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) so on. These algorithms can't achieve high positioning accuracy. PF [6] uses some random sampling points, namely particles, to describe the probability of target location. Then, through the weighted value of each particle, the estimated value of the MN is obtained.

In LOS case, EKF algorithm [7] has high positioning exactitude, but the positioning exactitude is dramatically reduced when there is NLOS interference. The interactive multimodel (IMM) algorithm has strong robustness, so many new algorithms are derived from the IMM algorithm to solve the problem of NLOS interference [8] [9] [10]. The disadvantage of IMM algorithm is that the statistical error needs to be known in advance.

REKF algorithm is a semi-parametric estimation algorithm based on TOA estimation [11], which has certain robustness. The REKF algorithm uses the Huber robust M estimator [12] to generate a linear regression model to process the estimated position, and to find the estimated value that meets the preset accuracy through multiple iterations, thereby mitigating the influence of NLOS. Replace NLOS filter in IMM algorithm with REKF, which is RIMM algorithm [13]. Although the statistical information of NLOS errors of RIMM does not need to be known in advance, c_1 and c_2 are set by ourselves. The MPDA [14] performs well in LOS. However, it is often easy to lose the target when the NLOS probability increases.

Due to the separate operation of each BN in the original PF, there is no connection between the various BNs. We hope to improve the existing PF to enable BNs to work together, thereby further reducing the impact of NLOS. Traditional PDA filters only use distance information when calculating correlation probabilities, and we found that direction is also an important factor. Therefore, we attempted to add direction information to the data association. In this paper, an algorithm IPF-DPDA is proposed for the above problems, this paper mainly makes the following contributions:

1) We have improved the traditional data association and added directional information on the original basis. The DPDA uses Mahalanobis distance and trajectory direction to compute the association probabilities to obtain more accurate location estimate.

2) Parallel with the previous particle filter BNs working alone, the improved particle filter BNs work together to alleviate negative impact of NLOS error. Improved particle filter significantly reduces the measurement distance error caused by NLOS.

3) The IPF-DPDA does not need to know the prior knowledge of NLOS errors in advance. Nevertheless, simulation and experimental results display that IPF-DPDA is better than those algorithms that need to know a priori information and performs better when the NLOS errors follow different distributions.

The whole paper is arranged as follows. Section II describes the related work. Section III introduces proposed algorithm. Simulations and experiment are described in Section IV, and Section V gives the conclusion.

2. Related Works

Many scholars have done research in the field of indoor location for the sake of improving positioning accuracy. In [15], Zhou proposed a method of Asymptotic Relative Efficiency (ARE), and based on this idea, the new hybrid hypothesis test is designed, which is applied to raise the accuracy of Wi-Fi indoor location. Compared to traditional WIFI positioning, the author not only considers the diversity of signal distribution but also considers the error of RSS measurement values, thus improving the positioning accuracy. In [16], Generative countermeasure network is proposed for RSSI data enhancement. The network uses a small part of real labeled data to generate false RSSI data. This algorithm also improves the location accuracy of indoor positioning. Compared to traditional methods, this method reduces the cost of data collection and reduces training data. A new ultrasonic signal classification and suppression algorithm is proposed for NLOS [17]. The algorithm processes the amplitude and residual of the received signal, and uses these two characteristics to sort and suppress NLOS. This algorithm has great advantages compared to other algorithms in classifying narrowband signals containing severe multipath effects. In [18], Zhou studied the location error based on hybrid fingerprint. The final results show that using the relevance between different forms of fingerprints can significantly mitigate positioning errors. When the number of fingerprints decreases, the location errors will increase. This is the drawback of this algorithm.

Combinational localization algorithm is also a current research trend [19-23]. In [19], A Pedestrian Dead Reckoning (PDR) and TDOA idea are integrated for indoor position. Launch a wall mirror of a NLOS BSs to generate a virtual base station (VBS), which is considered LOS BSs. After completing this process, more LOS BSs will participate in positioning. Compared to traditional algorithms, this algorithm solves the problem of unknown regions being unable to locate. In [20], Chang combined RSS and AOA methods to solve the positioning of mobile node in 3D. Simulations demonstrate their algorithm can achieve great performance in different situations. The main advantage of this algorithm is that it can work effectively regardless of whether the target's transmission power is known or unknown, which reflects the universality of the algorithm. In [21], identifies and selects the LOS and NLOS measurements which have better channel condition and geometric dilution of precision (GDOP) that derived by the hybrid Round-Trip Time (RTT), TDOA and direction of arrival (DOA). More measurement information fusion can reduce positioning errors, which is also a current research trend. In [22], an indoor location method combining Wi-Fi RTT and PDR is proposed. For the sake of increasing positioning accuracy of Wi-Fi RTT, an adaptive filtering system is formed by combining multiple EKF with a new anomaly test means. Tomic transformed the location issue into a generalized trust region subproblem based on TOA and RSS [23], and then solves it by bisection program. Because the difficulty of the research content, converting to a generalized trust domain is only an approximate transformation, which may affect the accuracy of the algorithm.

3. Proposed Method

3.1 Signal Model

In the two-dimensional positioning area, it has M BNs around MN. There may or may not be obstacles between MN and BNs. The state vector of MN is $X(l) = [x(l) \ y(l) \ \dot{x}(l) \ \dot{y}(l)]^T$, (x(l), y(l)) denotes the position of the MN and $(\dot{x}(l), \dot{y}(l))$ denotes the velocity of MN. Then the state equation is modeled as:

$$X(l) = FX(l-1) + w \tag{1}$$

(1) where, $F = \begin{bmatrix} 1 & 0 & T_S & 0 \\ 0 & 1 & 0 & T_S \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ denotes the state transition matrix of MN, and w denotes

measurement noise, T_s is sampling interval. We suppose M BNs to detect the location of MN. We let $d_i(l) = [d_1(l), d_2(l), ..., d_M(l)]$ express that the measured distance is between M BNs and MN:

$$l_i(l) = h(X(l)) + n, l = 1, \dots, L; i = 1, \dots, M$$
(2)

where, $h(X(l)) = [h_1(X(l)), h_2(X(l)), ..., h_M(X(l))]^T$ indicate the Euclidean distance between MN and the m - th BN with the position $(x_{bn,m}, y_{bn,m})$ is:

$$h_m(X(l)) = \sqrt{(x(l) - x_{bn,m})^2 + (y(l) - y_{bn,m})^2}, m = 1, ..., M$$
(3)

The noise n is defined as:

$$n = \begin{cases} n_{LOS} & \text{LOS condition} \\ n_{LOS} + n_{NLOS} & \text{NLOS condition} \end{cases}$$
(4)

where, n_{LOS} follows the Gaussian distribution $n_{LOS} \sim N(0, 1^2)$ The n_{NLOS} represents the NLOS error, as the environment changes, it may follow different distributions.

3.2 General Concept

The flow-process diagram of IPF-DPDA is showed in Fig. 1. We take for that incipient value and covariance of state estimation are computed ahead of time in this essay. We make use of M BNs and P particles to locate MN. At time l, the measured distance between BN and MN is recorded as d_i^l . The distance between the p-th particle and the i-th BN is recorded as $d_{i,p}^l$. Then calculate the residuals of d_i^l and $d_{i,p}^l$. Select P particles according to the obtained residuals. The first choice is to calculate the sum of residuals from each particle to each BN, and then normalize it. Particles with normalized weight greater than the average weight are retained. The second choice is to calculate the residual of the remaining particles again. If it is not less than the threshold (LOS) of the calculated, the particles will be retained, otherwise they will be discarded. Distance selection: carry out residual processing between the estimated distance and the measured distance. If the residual outweigh the average distance residual (regarded as NLOS), measured distance is updated with the estimated distance obtained from remaining particles. On the contrary, the distance is still measured.

Then, using the grouping idea and the least-squares method, C_M^3 measurements are obtained, $C_M^3 = N$. These N different position estimates are detected through the validation gate, and what inside validation gate is correct location estimated. Furthermore, those left outside the validation gate are discarded. If there are measured values inside validation gate, the corresponding association probabilities are applied to weight retained position estimate to

obtain final position estimation. If the number of inside validation gate is 0, update with predicted state estimate.

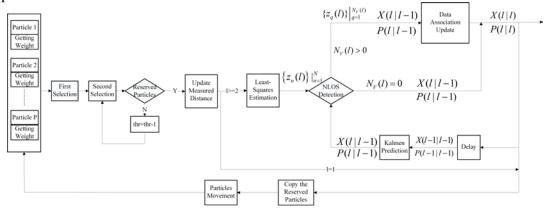


Fig. 1. The flowchart of IPF-DPDA

3.3 Improved Particle Filter

1) Particles Weighting: The weight of particles is related to the final distance calculation and update. The Weighting Particles has three steps:

Firstly, at time l, the distance difference between the p - th particle and the MN to the i - th BN is calculated. The distance difference can be represented as $res_{i,p}^{l}$. The specific expression is as follows:

$$res_{i,p}^{l} = |d_{i,p}^{l} - d_{i}^{l}|, i = 1, ..., M, p = 1, ..., P$$
(5)

Secondly, we compute the total residual from the p - th particle to each BN, which can be represented as res_p^l . The specific expression is as follows:

$$res_{p}^{l} = \sum_{i=1}^{M} |d_{i,p}^{l} - d_{i}^{l}|, p = 1, ..., P$$
(6)

Finally, the reciprocal of the total of the residuals of the p - th particle is obtained. The specific expression is as follows:

$$W_p^l = 1 / res_p^l \tag{7}$$

2) First Selection: We can know that when the particle is closer to MN, res_p^l is smaller and W_p^l is larger. When making the first selection, the particles with greater weight are retained.

$$H_{0}: \mathbf{W}_{j}^{l} \geq \sum_{i=1}^{P} W_{i}^{l} / P, j = 1, ..., P$$

$$H_{1}: \mathbf{W}_{j}^{l} < \sum_{i=1}^{P} W_{i}^{l} / P, j = 1, ..., P$$
(8)

If H_1 is false, the j - th particle is retained. Otherwise, the j - th particle is discarded. We assume that after the first selection, S particles are retained. However, if multiple BNs misjudge the particles, the particles can also be selected for the first time. Therefore we make a second selection.

3) Second Selection: We can use the residuals of the first step to approximate the probability of NLOS error.

$$H_{0}: res_{i,j}^{l} \geq \sum_{p=1}^{p} res_{p}^{l} / M * P, i = 1, ..., M; j = 1, ..., P$$

$$H_{1}: res_{i,j}^{l} < \sum_{p=1}^{p} res_{p}^{l} / M * P, i = 1, ..., M; j = 1, ..., P$$
(9)

We define the starting value of m as 0. If H_0 holds true, then m plus one, the cumulative value of m is obtained at last. Finally, the percentage of the value of m in the total is the probability of NLOS, which is recorded as λ .

$$\lambda = m / P * M \tag{10}$$

So, there are approximately $M * (1 - \lambda)$ BNs are in LOS. $thr = M * (1 - \lambda)$ (11)

We use the particles left by the first selection to make the second selection. The second selection avoids possible errors in the first selection. The specific explanation is as follows:

$$H_{0}: res_{i,j}^{l} \leq \sum_{p=1}^{p} res_{p}^{l} / P * M, i = 1, ..., M; j = 1, ..., S$$

$$H_{1}: res_{i,j}^{l} > \sum_{p=1}^{p} res_{p}^{l} / P * M, i = 1, ..., M; j = 1, ..., S$$
(12)

We define the starting value of n as 0. Calculate whether the residual from the j - th particle to M BNs meets H_0 . If H_0 holds true, then n plus one, the cumulative value of n is obtained at last. If $n \ge thr$, particle reserved, otherwise, particle discard. After calculating the j - th particle, redefine n as 0 and continue to calculate the next particle. Repeat this process until all saved particles are calculated for the first time.

When the two choices are completed, the algorithm uses the retained particles to calculate the distance to each BN. If the second selection does not leave any particles. Then, we select particles again by reducing the threshold by 1. Suppose U particles are retained after two selections.

4) Distance Update: We normalize weight of the remaining U particles, after normalization, weight of the p - th particle is MW_p^k at time l. The estimated distance dis_i^l between MN and i - th BN is fitted by weight of the reserved particle and the distance to the i - th BN.

$$dis_{i}^{l} = \sum_{p=1}^{U} MW_{p}^{l} * d_{i,p}^{l}, i = 1, ..., M$$
(13)

After obtaining the estimated distance, we use the actual measured distance and estimated distance to calculate the residual and average residual as follows:

$$res1_{i}^{l} = |dis_{i}^{l} - d_{i}^{l}|, i = 1, ..., M$$
(14)

We apply the acquired distance residuals to judge whether BN is in LOS relative to MN. It should be emphasized here that the i - th residual at this time is greater than the average residuals this time. If this condition is met, $d_i^l = dis_i^l$. That is the estimated distance obtained by the improved particle filter is be applied to update measured distance. Otherwise, the measured distance remains unchanged. The reason for this is to avoid the measurement distance being incorrectly updated.

5) Particle Copy and Movement: After the distance update is complete, the particles retained after two selections will be used in the next distance update. We constructed a replication layer for this purpose. The length of each layer is weight value of the particles, and total length is 1. We set layer L as follows:

$$L = (0, MW_1^l, MW_1^l + MW_2^l, \dots, \sum_{i=1}^{L-1} MW_i^l, 1)$$
(15)

Then generate a random number between 0 and 1, we set the generated random number as r. On which layer is the random number located, the corresponding particle is copied once, P times in total. This method ensures that the larger weight, the greater probability of particles will be copied.

The resulting new particles move in a small range, preventing these particles from concentrating in one area. The moved particles are used for the next distance update. The following equation represents the movement of particles:

$$(X_p^{l+1}, Y_p^{l+1}) = (X_p^l, Y_p^l) + N(0, \sigma_p^2)$$
(16)

3.4 Grouping and Positioning

We use LS method to estimate the position of MN, and estimate one position coordinate every three distance measurements. We assume the position of MN is(x(l), y(l)), There is the following relation between BNs and MN:

$$\begin{cases} (x_1 - x(l))^2 + (y_1 - y(l))^2 = (d_1(l))^2 \\ \vdots \\ (x_M - x(l))^2 + (y_M - y(l))^2 = (d_M(l))^2 \end{cases}$$
(17)

Equation (17) can be represented by matrix AP = B.

$$A = 2 \begin{bmatrix} (x_{1} - x_{2}) &, (y_{1} - y_{2}) \\ (x_{1} - x_{3}) &, (y_{1} - y_{3}) \\ \vdots \\ (x_{1} - x_{M}) &, (y_{1} - y_{M}) \end{bmatrix} P = \begin{bmatrix} x(l) \\ y(l) \end{bmatrix}$$

$$B = \begin{bmatrix} d_{2}(l)^{2} - d_{1}(l)^{2} - (x_{2}^{2} + y_{2}^{2}) + (x_{1}^{2} + y_{1}^{2}) \\ d_{3}(l)^{2} - d_{1}(l)^{2} - (x_{3}^{2} + y_{3}^{2}) + (x_{1}^{2} + y_{1}^{2}) \\ \vdots \\ d_{M}(l)^{2} - d_{1}(l)^{2} - (x_{M}^{2} + y_{M}^{2}) + (x_{1}^{2} + y_{1}^{2}) \end{bmatrix}$$
(18)

At time l, the estimated position coordinates can be derived from the above equation as follows:

$$z(l) = [x(l), y(l)]^{T} = (A^{T}A)^{-1}A^{T}B$$
(19)

So total estimated position coordinates are $z_n(l)$, n = 1, ..., N. Then we use the Gauss Newton algorithm [24] to iteratively discover minimum value of quadratic cost function, in which the nonlinear $h(\cdot)$ is approximated by the first-order Taylor series.

$$z_{n,s+1} = z_{n,s} + (H_n(z_{n,s})^T H_n(z_{n,s}))^{-1} * H_n(z_{n,s})(d - h_n(z_{n,s})),$$

$$n = 1, 2, \dots, N$$
(20)

where s is the iteration index, s plus one for each iteration. $h_n(z_{n,s})$ represents in subgroup n the Euclidean distances between the three corresponding BSs and $z_{n,s}$. The Jacobian matrix $H_n(z_n)$ is defined as:

$$H_n(\mathbf{z}_n) = \frac{\partial h_n([x, y]^T)}{\partial [x, y]^T} \Big|_{x=x, y=y}$$
(21)

The iterations are repeated until between $z_{n,s}$ and $z_{n,s+1}$ is smaller than a certain threshold and when iteration stops the value of *s* is less than 7. Finally, let $z_{n,s+1}$ be the final position estimate.

3.5 Algorithm

1) Kalman Filter Prediction: we assume that initial value $\hat{X}(0|0)$ and covariance of state estimation $\hat{P}(0|0)$ are computed ahead of time. The prediction state and covariance matrix estimation are obtained by KF:

$$\hat{X}(l \mid l-1) = F\hat{X}(l-1 \mid l-1)$$
(22)

$$P(l | l-1) = FP(l-1 | l-1)F^{T} + CQC^{T}$$
(23)

$$\hat{z}(l \mid l-1) = B \hat{X}(l \mid l-1)$$
(24)
$$E^{2} \hat{z}(l \mid l-1) = B \hat{X}(l \mid l-1)$$

where $B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, $C = \begin{bmatrix} T_S^2/2 & 0 \\ 0 & T_S^2/2 \\ T_S & 0 \\ 0 & T_S \end{bmatrix}$, $Q = I_2$, and $\hat{z}(l|l-1)$ represents the

predicted coordinates of MN.

2) NLOS Detection: First, we get the difference between N measurements and predicted values.

 \wedge

$$v_n(l) = z_n(l) - z(l | l - 1), n = 1, ..., N$$
 (25)

Then, we judge whether the N position estimation is in LOS or NLOS. If there are no NLOS errors, then there are:

$$v_n(l) \sim N(0, S_n(l)), n = 1, ..., N$$
 (26)

where, $S_n(l)$ is innovation covariance matrix of the n - th subgroup, calculated by the following equation:

$$S_n(l) = BP \left(l \mid l-1 \right) B^T + \sigma_L^{2} \left(H_n^T(z_n) H_n(z_n) \right)^{-1}$$
(27)

We can calculate the area of the validation gate through the innovation covariance:

$$V_n(l) = \gamma \pi |S_n(l)|^{0.5}$$
(28)

To test (26), we define N hypotheses and another N alternative hypotheses:

$$\zeta_{0,n}: v_n(l) \sim N(0, S_n(l), \ n = 1, ..., N$$
(29)

$$\zeta_{1,n}$$
: not $\zeta_{0,n}$, $n = 1, ..., N$

If $\zeta_{0,n}$ is true, BNs of the n - th subgroups are in LOS. Otherwise, the BNs are in NLOS. To detect $z_n(l)$ whether is in the LOS condition, that is, whether it's inside validation gate. We can compare test statistic $T_n(l)$ with threshold of the validation gate γ to judge the position estimation $z_n(l)$ Whether is in LOS or NLOS condition. $T_n(l)$ is calculated as follows:

$$T_n(l) = v_n^{T}(l)S_n^{-1}(l)v_n(l)$$
(30)

If $T_n(l)$ is smaller than γ , $z_n(l)$ is in LOS, $z_n(l)$ is inside validation gate, $\zeta_{0,n}$ is true. Otherwise, $\zeta_{1,n}$ holds true. γ is associated with the threshold probability P_G , P_G means the probability of a measurement from LOS BNs falls in the validation gate. The relationship between the two is as follows:

$$\int_{0}^{\gamma} f_{x^{2}(2)}(x) dx = P_{G} = 1 - P_{FA}$$
(31)

where $f_{\chi^2(2)}(\cdot)$ indicates the probability density function of chi- square distribution with two degrees of freedom. Due to the complexity of the indoor environment, it is impossible for us to know in advance the NLOS situation in this indoor environment. Therefore, we chose a suitable empirical P_G value [14]. When P_G is known, the chi-square distribution table can get value of γ .

We ruled out $z_n(l)$ outside the validation gate and count number that in the validation gate, record it as $N_v(l)$.

3) Data Association and Update: We set up following associated events according to the method of [25]:

 $\theta_q(l)$: { $z_q(l)$ is depend on the LOS BSs with the smallest error covariance, $q = 1, ..., N_V(l)$ }.

 $\theta_0(l)$: {none of the location estimations $z_q(l)$ originates from the LOS BNs}.

The associated probabilities are:

$$\beta_q(l) = \Pr\{\theta_q(l) | Z^l\} = \Pr\{\theta_q(l) | Z(l), N_V(l), Z^{l-1}\}. q = 1, ..., N_V(l)$$
(32)

where $Z(l) = \{Z_q(l)\}_{q=1}^{N_V(l)}$ and Z^l is the cumulative set of the location estimated.

Since Bayesian formula is the basis of association probability calculation, assuming that the innovations $v_q(l), q = 1, ..., N_V(l)$ are mutually independence, then (32) is:

$$\beta_q(l) = \frac{1}{c} f[Z(l) \mid \theta_q(l), N_V(l), Z^{l-1}] \times \Pr\{\theta_q(l) \mid N_V(l), Z^{l-1}\}, q = 0, 1, ..., N_V(l)$$
(33)

where c and $f(\cdot)$ represent normalization factor and joint probability density function for the location estimated, respectively.

We assume that the probability density functions of correct location estimation and incorrect location estimation obey Gaussian distribution and uniform distribution respectively. The probability density functions for correct and incorrect location estimation are as follows:

$$f[z_q(l) | \theta_q(l), N_V(l), Z^{l-1}] = P_G^{-1} N(z_q(l); z(l | l-1), S_q(l))$$
(34)

$$=P_{G}^{-1}N(v_{q}(l);0,S_{q}(l))$$
(35)

$$=P_{G}^{-1}\frac{\exp\{-\frac{1}{2}v_{q}^{T}(l)S_{q}^{-1}(l)v_{q}(l)\}}{2\pi|S_{q}(l)|^{0.5}}$$
(36)

where $|S_q(l)|$ indicates the determinant of matrix $S_q(l)$.

$$f[z_q(l) | \theta_q(l), N_V(l), Z^{l-1}] = V_q^{-1}(l)$$
(37)

where $V_q(l)$ is area of the validation gate (28) of the $N_V(l)$ accepted hypothesis.

The prior probabilities in (33) are:

$$\Pr\{\theta_{q}(l) \mid N_{V}(l), Z^{l-1}\} = \frac{P_{d}P_{G}}{N_{V}(l)}, q = 1, ..., N_{V}(l)$$

$$\Pr\{\theta_{q}(l) \mid N_{V}(l), Z^{l-1}\} = 1 - P_{d}P_{G}, q = 0$$
(38)

Where P_d represents the detection probability.

It can be obtained from (36) and (37), the joint probability density function of correct and incorrect location estimation are:

$$f[Z(l) | \theta_q(l), N_V(l), Z^{l-1}] = \prod_{i=1}^{N_V(l)} \frac{N(v_q(l); 0, S_q(l))}{V_i(l) P_G}, q = 1, \dots, N_V(l)$$
(39)

$$f[Z(l) | \theta_q(l), N_V(l), Z^{l-1}] = \prod_{i=1}^{N_V(l)} V_i^{-1}(l), q = 0$$
(40)

Finally, we get the association probabilities as follows:

$$\beta_{q}'(l) = \frac{N(v_{q}(l); 0, S_{q}(l))}{P_{G}} (\frac{P_{d}P_{G}}{N_{V}(l)}) \prod_{i=1, i \neq q}^{N_{V}(l)} V_{i}^{-1}(l), q = 1, ..., N_{V}(l)$$

$$\beta_{0}'(l) = (1 - P_{d}P_{G}) \prod_{i=1}^{N_{V}(l)} V_{i}^{-1}(l), q = 0$$
(41)

When we study indoor location, we find that MN direction is also a crucial information. For us, it is a kind of visual data [26]. It can be applied to data association to increase the accuracy of positioning. In order to add direction information, the introduction of pseudo direction is very necessary, because direction information can't be easily obtained in distancebased positioning. The pseudo direction connects the verified measurement position and the predicted position at current time with estimated position at previous time, as shown in Fig. 2. $\hat{z}(l|l-1)$ is the predicted state value of the target trajectory at time l, and $\hat{z}(l-1)$ is the estimated position value at time l-1. We use Δ_q to indicate the direction difference. The expression is as follows:

$$\Delta_q = |\gamma_q - \gamma|, q = 1, \dots, N_V(l) \tag{42}$$

where γ_q represents the included angle between vector orientation and horizontal orientation of the q - th position estimation, γ indicates the angle between vector orientation and horizontal orientation of the state prediction value.

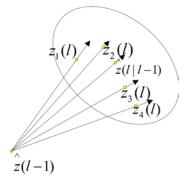


Fig. 2. Illustration of concept of pseudo direction

We use the angle difference between γ and γ_q to decide orientation weight. When Gaussian weight $exp(-\Delta^T \delta_{\phi}^{-1} \Delta/2)$ is applied, it is multiplied by the association probability of PDA. In the exponential the parameter δ_{ϕ} used that is the variance of direction of MN. Define the direction of MN and the speed as follows:

$$\phi = \tan^{-1}(\frac{x(l)}{\cdot}), \tag{43}$$

$$v(l) = \sqrt{\frac{1}{x(l)^2 + y(l)^2}}$$
(44)

We use the linearization method to calculate the first and second moments, and ignore the higher-order terms.

$$\Delta \phi = \frac{\partial \phi}{\partial \dot{x}(l)} \dot{\Delta x}(l) + \frac{\partial \phi}{\partial \dot{y}(l)} \dot{\Delta y}(l) + O(\dot{\Delta x}(l)^2, \dot{\Delta y}(l)^2, \dot{\Delta x}(l) \dot{\Delta y}(l))$$
(45)

By differentiating (45), we get

$$\frac{\partial \phi}{\partial \dot{x}(l)} = \frac{\dot{y}(l)}{v(l)^2}, \quad \frac{\partial \phi}{\partial \dot{y}(l)} = -\frac{\dot{x}(l)}{v(l)^2}$$
(46)

And then,

$$\Delta \phi = \frac{\dot{y}(l)}{v(l)^2} \Delta \dot{x}(l) - \frac{\dot{x}(l)}{v(l)^2} \Delta \dot{y}(l)$$
(47)

Then calculate the expectation and the variance from (47):

$$E\{\Delta\phi\} = \frac{1}{v(l)^{2}} [\dot{y}(l)E\{\Delta\dot{x}(l)\} - \dot{x}(l)E\{\Delta\dot{y}(l)\}] = 0,$$

$$E\{(\Delta\phi)^{2}\} = \delta_{\phi}^{2} = [\dot{y}(l)^{2}P_{\dot{x}(l)} - \dot{2}\dot{x}(l)\dot{y}(l)P_{\dot{x}(l)\dot{y}(l)} + \dot{x}(l)^{2}P_{\dot{y}(l)}] / v(l)^{4}$$
(48)

where the covariance terms are from *P* as follows:

$$P = \begin{pmatrix} P_{x} & P_{xy} & P_{.} & P_{.} \\ & & & xx & & xy \\ P_{xy} & P_{y} & P_{.} & P_{.} \\ P_{xy} & P_{y} & P_{.} & P_{.} \\ P_{.} & P_{.} & P_{.} & P_{.} \\ xx & & yx & x & & xy \\ P_{.} & P_{.} & P_{.} & P_{.} \\ xy & & yy & xy & y \end{pmatrix}$$
(49)

After adding the direction, the association probabilities of the DPDA becomes:

$$\beta_{q}^{'}(l) = \frac{N(v_{q}(l); 0, S_{q}(l))}{P_{G}} (\frac{P_{d}P_{G}}{N_{V}(l)}) \prod_{i=1, i \neq q}^{N_{V}(l)} V_{i}^{-1}(l) \times \exp(-\frac{\Delta_{q}^{-T} \delta_{\phi}^{-1} \Delta_{q}}{2}), q = 1, ..., N_{V}(l)$$

$$\beta_{0}^{'}(l) = (1 - P_{d}P_{G}) \prod_{i=1}^{N_{V}(l)} V_{i}^{-1}(l), q = 0$$
(50)

Due to the fact that the base of the exponential function is greater than 1, it is an increasing function. Therefore, a larger $-\frac{\Delta q^T \delta \phi^{-1} \Delta q}{2}$ represents a higher probability of allocation, while a larger $-\frac{\Delta q^T \delta \phi^{-1} \Delta q}{2}$ represents a smaller direction difference. So smaller directional differences will be assigned a higher probability.

Finally, the association probabilities are modeled as:

$$\beta_{q}(l) = \frac{\beta_{q}(l)}{\beta_{0}(l) + \sum_{q=1}^{N_{v}(l)} \beta_{q}(l)}$$
(51)

$$\beta_{0}(l) = \frac{\beta_{0}(l)}{\beta_{0}(l) + \sum_{q=1}^{N_{V}(l)} \beta_{q}(l)}$$
(52)

The innovation covariance and Kalman gain are respectively expressed by the following equation:

$$S(l) = BP(l | l-1)B^{T} + \sigma_{L}^{2}I_{2}$$
(53)

$$K(l) = P(l | l-1)B^{T}S^{-1}(l)$$
(54)

We make use of the association probabilities calculated above to get final state estimate $\hat{X}(l|l)$ and covariance matrix P(l|l).

$$\hat{X}(l \mid l) = \hat{X}(l \mid l-1) + K(l) \sum_{q=1}^{N_v(l)} \beta_q(l) \nu_q(l)$$
(55)

$$P(l|l) = \beta_0(l)P(l|l-1) + (1 - \beta_0(l))P_c(l|l) + P(l)$$
(56)

$$P_{c}(l \mid l) = (I_{4} - K(l)B(l))P(l \mid l - 1)$$
(57)

$$\tilde{P}(l) = K(l) \left[\sum_{q=1}^{N_v(l)} \beta_q(l) v_q(l) v_q^T(l) - v(l) v^T(l) \right] K^T(l)$$
(58)

$$v(l) = \sum_{q=1}^{N_v(l)} \beta_q(l) v_q(l)$$
(59)

4. Simulation and Experimental Results

4.1 Simulation

In this part, we conduct simulation experiments and the simulation results are analyzed. The trajectory and scene of the simulation experiment are shown in **Fig. 3**. Positioning 100 points and sampling interval is 0.5s. We set the initial state and covariance matrix as $X(0) = [1m \ 19.99m \ 1 \ m/s \ 0.1 \ m/s]^T$ and $P(0) = I_4$ respectively. The P_G is 0.99, P_d is 0.9. We randomly generate a number between 0 and 1 to simulate the NLOS environment. We compare the random number with the NLOS probability threshold to judge whether BN and the MN are in NLOS state. If the digit is less than NLOS probability threshold, this BN is in NLOS state. The NLOS errors comply with Gaussian distribution in simulation experiment. We compare IPF-DPDA with the MJPDA [27], RIMM [13], IMM-EKF [9], REKF [13] EKF [7]. 1000 Monte Carlo runs are used to get simulation results. And the good or bad of algorithm is estimated through the root mean square error (RMSE) and cumulative distribution function (CDF) of the average localization error.

$$RMSE = \sqrt{\frac{1}{MC} \frac{1}{L} \sum_{j=1}^{MC} \sum_{l=1}^{L} \left((\hat{x}_{j}(l) - x_{j}(l))^{2} + (\hat{y}_{j}(l) - y_{j}(l))^{2} \right)}$$
(60)

Where L = 100 and MC = 1000 represent the number of moving steps and the number of Monte Carlo simulation runs, respectively. $(\stackrel{\wedge}{x_j}(l), \stackrel{\wedge}{y_j}(l))$ is the position estimation and $(x_j(l), y_j(l))$ is the true location of MN during the j - th Monte Carlo run.

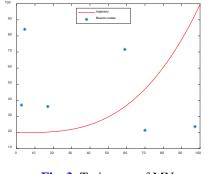


Fig. 3. Trajectory of MN

Gaussian Distribution: We suppose that NLOS errors and measurement noise obeyed the Gaussian distribution $N(\mu_{NLOS}, \sigma_{NLOS}^2)$ and $N(0, \sigma_L^2)$, respectively. The specific parameter settings are shown are displayed in Table 1. When we do simulation experiments, one parameter changes while the other parameters remain unchanged, for the sake of observing the effect of this parameter on location accuracy of algorithm.

Parameter	Symbol	Values
Quantity of BN	М	6
NLOS probability	P _{NLOS}	0.5
Gauge noise	$N(0, \sigma_L^2)$	$N(0, 1^2)$
NLOS error	$N(\mu_{NLOS}, \sigma_{NLOS}^2)$	$N(5, 6^2)$

Table 1. Default Parameters of Gaussian Distribution

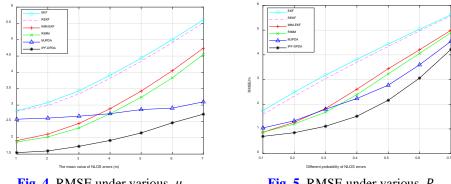


Fig. 4. RMSE under various μ_{NLOS}

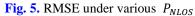
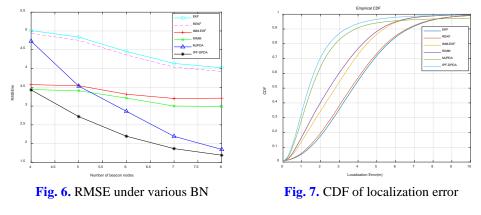


Fig. 4 mirrors the transform when the μ_{NLOS} raises from 1 to 7. The black curve represents my algorithm, it can be obviously seen that the location accuracy of IPF-DPDA is better. In more details, average positioning error of IPF-DPDA, MJPDA, RIMM, IMM-EKF, REKF and EKF are 2.017m, 2.776m, 2.936m, 3.081m 3.963m, and 4.043m respectively, and IPF-DPDA improves average location accuracy by 27.34%, 31.30%, 34.53%, 49.10% and 50.11% separately.

Fig. 5 distinctly reflects the average localization errors an ascending tendency as P_{NLOS} increases. The average location accuracy of IPF-DPDA is 1.944m, while MJPDA, RIMM, IMM-EKF, REKF and EKF are 2.473m, 2.617m, 2.744m, 3.649m, 3.761m respectively. The IPF-DPDA raises average location accuracy by 21.39%, 25.72%, 29.15%,



47.85% and 48.31% respectively.

With the BN raises in **Fig. 6**, the location accuracy of all algorithms has been improved. However, RIMM, IMM-EKF, REKF and EKF improve slowly, while IPF-DPDA and MJPDA improve faster than the previous four algorithms. In more details, the location accuracy of IPF-DPDA, MJPDA, RIMM, IMM-EKF, REKF and EKF are 2.381m, 3.032m, 3.21m, 3.368m, 4.401m, and 4.491m respectively. The IPF-DPDA raises average location accuracy by 21.47%, 25.83%, 29.31%, 45.90% and 46.98% respectively.

In **Fig. 7**, CDF reflects the positioning accuracy of my algorithm, the 90% localization error of IPF-DPDA, MJPDA, RIMM, IMM-EKF, REKF and EKF are less than 3.357 m, 3.863m, 5.192m, 5.375m, 6.601m and 6.693m separately.

4.2 Experiment

Because the simulations experiment can't fully explain the localization accuracy of the IPF-DPDA, we designed a true experiment to verify it. The ultra-wideband (UWB) location has low transmission power and can realize accurate information exchange between equipment, so as to realize high-precision positioning. The UWB equipment used is displayed in **Fig. 8** and **9** in this paper. **Fig. 8** shows the MN and power supply, **Fig. 9** shows the BN.



The experiment is carried out in our laboratory, as shown in **Fig. 10**. The length of laboratory is 12.6 meters and the width is 6.6 meters. There are seven beacon nodes in the experimental scene, and the person moves along the trajectory. The blue and green positions are MN and BN respectively. Place MN and BNs 1.2m above the floor, for the sake of avoiding UWB signal reflecting from the floor. The initial state of MN is set to $X(0) = [1.8m \ 6m \ 0m/s \ -0.6 \ m/s]^T$. The seven BNs are located at $(x_1 = 1.2m, y_1 = 3.6m)$, $(x_2 = 3.3m, y_2 = 0.6m)$, $(x_3 = 2.4m, y_3 = 4.6m)$, $(x_4 = 5.4m, y_4 = 4.2m)$, $(x_5 = 9.1m, y_5 = 1.2m)$, $(x_6 = 9.6m, y_6 = 3.6m)$, and $(x_7 = 11.4m, y_7 = 5.2m)$

respectively. We sampled a total of 30 points, with a 1-second interval between each sampling point. Obtain 20 values at each sampling point and take the average as the current distance measurement value. Other parameters remain unchanged, the same as the simulation.



Fig. 10. (a) Real Scene of the Laboratory (b) Plan of Laboratory

The error distribution of each sampling point and the CDF of positioning error are displayed in Fig. 11 and Fig. 12 separately.

As can be observed from **Fig. 11**, the black curve represents my algorithm, the location accuracy of IPF-DPDA is the highest more than half of the sampling points. The positioning accuracy of IPF-DPDA, MJPDA, RIMM, IMM-EKF, REKF and EKF are 0.6647m, 0.7693m, 0.8532m, 0.8442m, 0.8313m, and 0.8317m respectively on average.

From **Fig. 12**, the CDF of each algorithm is shown when experiment in the laboratory. The light blue line is my algorithm, which is obviously better than other algorithms. In more detail, the 90% localization error of IPF-DPDA, MJPDA, RIMM, IMM-EKF, REKF and EKF are less than 1.105 m, 1.541, 1.577m. 1.709m, 1.837m and 1.701m separately.

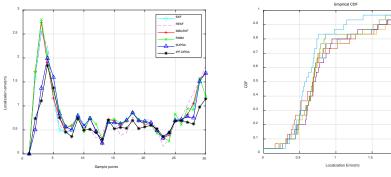
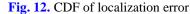


Fig. 11. Sampling point positioning error



5. Conclusion

A location method based on improved particle filter and directional probability data association is proposed to track MN. Firstly, an improved particle filter is applied to reduce error of measuring distance. Then the hypothesis test is applied to detect whether measurements are in LOS situations or in NLOS situations for *N* different groups. If there are measurements in the validation gate, the corresponding association probabilities are applied to weight retained position estimate to gain the final location estimation. If the number of inside validation gate is 0, update with predicted state estimate. Simulation and experimental results display IPF-DPDA performance better than MJPDA, RIMM, IMM-EKF, REKF and EKF in

all aspects. The next work we need to do is to enhance the robustness of IPF-PDDA algorithm when NLOS error probability is large.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant No.62273083 and No.61803077; Natural Science Foundation of Hebei Province under Grant No. F2020501012.

References

- Zheng Haifeng, W. Guo, and N. Xiong, "A Kernel-Based Compressive Sensing Approach for Mobile Data Gathering in Wireless Sensor Network Systems," *IEEE Trans Systems Man & Cybernetics*, vol. 48, no. 12, pp. 2315-2327, Dec. 2018. <u>Article (CrossRef Link)</u>
- [2] P. Thilagavathi, J. M. L. Manickam, "An Enhanced RSSI based Tree Climbing mechanism for well-planned path localization in WSN using the virtual force of Mobile Anchor Node," *Journal* of Ambient Intelligence and Humanized Computing, vol. 07, Jul. 2020. <u>Article (CrossRef Link)</u>
- [3] S. Bottigliero, D. Milanesio, M. Saccani and R. Maggiora, "A Low-Cost Indoor Real-Time Locating System Based on TDOA Estimation of UWB Pulse Sequences," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021. <u>Article (CrossRef Link)</u>
- [4] C. Geng, T. E. Abrudan, V. -M. Kolmonen and H. Huang, "Experimental Study on Probabilistic ToA and AoA Joint Localization in Real Indoor Environments," in *Proc. of IEEE International Conference on Communications*, pp. 1-6, 2021. <u>Article (CrossRef Link)</u>
- [5] U.S. Environmental Protection Agency :: U.S. EPA, "Report to Congress on indoor air quality. Volume 2: Assessment and control of indoor air pollution," U.S. Environ. Protect. Agency, Washington, DC, USA, Tech. Rep. EP/400/1-89/001C, 1989.
- [6] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, Feb. 2002. <u>Article (CrossRef Link)</u>
- [7] W. Cui, B. Li, L. Zhang, "Robust Mobile Location Estimation in NLOS Environment Using GMM, IMM, and EKF," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3490-3500, Sep. 2019. <u>Article (CrossRef Link)</u>
- [8] J. Liao and B. Chen, "Robust Mobile Location Estimator with NLOS Mitigation using Interacting Multiple Model Algorithm," *IEEE Trans Wireless Communications*, vol. 5, no. 11, pp. 3002-3006, Nov. 2006. <u>Article (CrossRef Link)</u>
- [9] B. S. Chen, C. Y. Yang, F. K. Liao, and J. F. Liao, "Mobile location estimator in a rough wireless environment using extended Kalman based IMM and data fusion," *IEEE Trans Veh Technol*, vol. 58, no. 3, pp. 1157–1169, 2009. <u>Article (CrossRef Link)</u>
- [10] C. Y. Yang, B. S. Chen, and F. K. Liao, "Mobile location estimation using fuzzy-based IMM and data fusion," *IEEE Trans Mobile Comput*, vol. 9, no. 10, pp. 1424–1436, 2010. <u>Article (CrossRef Link)</u>
- [11] U. Hammes, E. Wolsztynski and A. M. Zoubir, "Semi-parametric geolocation estimation in NLOS environments," in *Proc. of 16th European Signal Processing Conference*, Lausanne, Switzerland, pp. 1-5, 2008. <u>Article (CrossRef Link)</u>
- [12] Huber, P.J, "Robust Statistics," in *International Encyclopedia of Statistical Science*, 2nd ed., Berlin, Germany: Springer, 2014, pp. 1248–1251. [Online]. Available: https://link.springer.com.
- [13] U. Hammes, and A. M. Zoubir, "Robust MT Tracking Based on M-Estimation and Interacting Multiple Model Algorithm," *IEEE Trans. Signal Process*, vol. 59, no. 7, pp. 3398-3409, Jul. 2011. <u>Article (CrossRef Link)</u>

- [14] U. Hammes and A. M. Zoubir, "Robust mobile terminal tracking in NLOS environments based on data association," *IEEE Trans Signal Process*, vol. 58, no. 11, pp. 5872–5882, Nov. 2010. <u>Article (CrossRef Link)</u>
- [15] M. Zhou, Y. Li, M. J. Tahir, X. Geng, Y. Wang and W. He, "Integrated Statistical Test of Signal Distributions and Access Point Contributions for Wi-Fi Indoor Localization," *IEEE Transactions* on Vehicular Technology, vol. 70, no. 5, pp. 5057-5070, May. 2021. <u>Article (CrossRef Link)</u>
- [16] W. Njima, M. Chafii, A. Chorti, R. M. Shubair and H. V. Poor, "Indoor Localization Using Data Augmentation via Selective Generative Adversarial Networks," *IEEE Access*, vol. 9, pp. 98337-98347, 2021. <u>Article (CrossRef Link)</u>
- [17] S. Haigh, J. Kulon and A. Partlow, "A Robust Algorithm for Classification and Rejection of NLOS Signals in Narrowband Ultrasonic Localization Systems," *IEEE Trans Instrumentation and Measurement*, vol. 68, no. 3, pp. 646-655, Mar. 2019. <u>Article (CrossRef Link)</u>
- [18] C. Zhou, J. Liu, M. Sheng, Y. Zheng and J. Li, "Exploiting Fingerprint Correlation for Fingerprint-Based Indoor Localization: A Deep Learning Based Approach," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5762-5774, June. 2021. <u>Article (CrossRef Link)</u>
- [19] Z. Deng, X. Zheng, C. Zhang, H. Wang, L. Yin and W. Liu, "A TDOA and PDR Fusion Method for 5G Indoor Localization Based on Virtual Base Stations in Unknown Areas," *IEEE Access*, vol. 8, pp. 225123-225133, 2020. <u>Article (CrossRef Link)</u>
- [20] S. M. Chang, Y.M. Li, X. J. Yang, H. Wang, W. F. Hu and Y. Q. Wu, "A Novel Localization Method Based on RSS-AOA Combined Measurements by Using Polarized Identity," *IEEE Sensors Journal*, vol. 19, no. 4, pp. 1463-1470, Feb. 2019. <u>Article (CrossRef Link)</u>
- [21] Deng Z, Wang H, Zheng X, Yin L, "Base Station Selection for Hybrid TDOA/RTT/DOA Positioning in Mixed LOS/NLOS Environment," *Sensors*, vol. 20, no.15, Aug. 2020. <u>Article (CrossRef Link)</u>
- [22] X. Liu et al., "Kalman Filter-Based Data Fusion of Wi-Fi RTT and PDR for Indoor Localization," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 8479-8490, Mar. 2021. <u>Article (CrossRef Link)</u>
- [23] S. Tomic, M. Beko, "A Robust NLOS Bias Mitigation Technique for RSS-TOA-Based Target Localization," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 64-68, Jan. 2019. Article (CrossRef Link)
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, U.K: Cambridge Univ Press, 2004.
- [25] T. KIRUBARAJAN and Y. BAR-SHALOM, "Probabilistic data association techniques for target tracking in clutter," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 536-557, March. 2004. <u>Article (CrossRef Link)</u>
- [26] M. Winter and G. Favier, "A neural network for data association," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, vol.2, pp. 1041-1044, 1999. <u>Article (CrossRef Link)</u>
- [27] L. Cheng, Y. Li, M. Xue and Y. Wang, "An Indoor Localization Algorithm Based on Modified Joint Probabilistic Data Association for Wireless Sensor Network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 63-72, Jan. 2021. <u>Article (CrossRef Link)</u>



Long Cheng received B.S. degree in automatic control from Qingdao University of Science and Technology, Qingdao, China, in 2008, the M.S. degree in pattern recognition and intelligent system from Northeastern University, Shenyang, China, in 2010. He received Ph.D. degree in pattern recognition and intelligent system from Northeastern University, Shenyang, China, in 2013. Now, he is an associate professor in Northeastern University. His research interests include wireless sensor networks and distributed system.



Jiayin Guan obtained a Bachelor's degree in Electronic Information Engineering from Shenyang Aerospace University in Shenyang, Liaoning Province, China in 2021. He is currently pursuing a master's degree at Northeastern University in Qinhuangdao, China. His research interest is localization technology based on wireless sensor networks.