

TOA based Target Tracking using Optimal Estimation Methods

Shihong Duan^{1,2,3}, Yuanyuan Li^{1,2,3}, Cheng Xu^{1,2,3*}, Jiawang Wan^{1,2,3}, Jie He^{1,2,3}

¹School of Computer and Communication Engineering, University of Science and Technology Beijing

²Beijing Key Laboratory of Knowledge Engineering for Materials Science

³Institute of Artificial Intelligence, University of Science and Technology Beijing
Beijing, China

Abstract—Position is one of the essential attributes of an object. With the development of wireless communication and ranging technology, localization algorithms should also be investigated to improve its accuracy. Ultra Wideband (UWB) has the potentials of low cost and high precision. However, its accuracy is susceptible to environment, such as non-line-of-sight conditions. There is a pressing need for a suitable algorithm to reduce the impact of ranging errors, and ensure the performance of the positioning systems when the ranging error varies.

Index Terms—Time of Arrival (TOA), optimization, localization, location estimation, sensor network

I. INTRODUCTION

Localization technology is widely used in daily life, such as people management in public, search and rescue in dangerous environments, and intelligent tracking of robots. With the development of wireless communication technology, people have higher requirements for localization. Wireless sensor networks require mutual communication between nodes for self-localization. The traditional GPS method can provide high-precision 3D ranging data. The indoor environment is blocked by the wall, and the satellite signal is weakened. Therefore, GPS is not suitable for indoor wireless sensing [1]. Common indoor localization technologies include Bluetooth, Wi-Fi, infrared, UWB, etc. UWB technology has high security, simple design, low cost, and centimeter-level ranging capability, therefore UWB is advantageous and suitable for indoor wireless localization [2], [3].

There are mainly two types of localization algorithms: the range-based localization algorithms and range-free localization algorithms. Range-free localization algorithms have lower requirements on equipment and lower ranging accuracy. For indoor localization system with high accuracy need, range-based localization algorithms is more suitable. Common ranging methods for indoor localization include: TOA/TDOA (time of arrival/time difference of arrival), RSS (received signal strength) and AOA (Angle of arrival) [4]. TDOA requires precise clock synchronization between anchors [5]; the TOA

method doesn't, which avoids synchronization errors [6], [7]. RSS has lower localization accuracy, AOA requires multiple antennas, so the node volume is large [8]. Considering node deployment, cost, etc., this paper will adopt TOA.

TOA measurement data often contain errors, due to the obstacles, weather, target locations and antenna height. Because the ranging error is inevitable, the localization result is not an accurate coordinate, but a constraint range. The purpose of the localization algorithm is to narrow down the possibilities and optimize the localization result so that the result is closer to the real one.

Common localization algorithms based TOA ranging include mathematical analysis, maximum likelihood estimation (MLE), least squares (LS) and genetic algorithms (GA) [9]. As mentioned above, due to the ranging error, the nonlinear equations has non-uniqueness solution. Mathematical analysis refers to averaging the solutions of the equations, but the results are inaccurate. GA is more accurate results when the ranging error is large, reducing the error caused by signal transmission. However, UWB is accurate in ranging, and GA is not suitable. MLE is affected by the probability distribution of measurement error. MLE may cause a large error, if the signal-noise ratio is below a threshold [10].

LS is a classic optimization method and insensitive to the distribution of the ranging error. MLE and LS are same when the ranging error is a normal distribution of 0 mean. Bayesian estimation method maximum posterior probability (MAP) results are better than MLE, but it is hard to get the posterior probability. LS algorithm perform better when the ranging error is small. LS performance varies greatly as the error increases. The optimization problem of localization maybe non-convex, and the SDR method can convert the non-convex problem into a semi-positive programming solution.

Based on the above-mentioned factors, it is necessary to propose an algorithm, which is insensitive to the probability error distribution and improve localization accuracy. Therefore, this paper proposes an algorithm based on Chebyshev center, which has better performance without probability distribution of ranging error. Chebyshev center considers the real position of the target node, and ensure that the maximum error between the localization result and the real position is minimum. Our work makes the following main contributions: 1) The real position is added to the definition of the objective function;

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC0901303, in part by the National Natural Science Foundation of China (NSFC) under Projects 61671056, 61302065, 61304257, and 61402033, in part by the Beijing Natural Science Foundation under Project 4152036, and in part by the Tianjin Special Program for Science and Technology under Grant 16ZXCXSF00150.

*Corresponding author: xucheng@ieee.org

2)The real position is added to the definition of the objective function; 3)The simulation shows that the localization accuracy is improved compared with other optimal methods, and the performance is better when the ranging error is large.

The rest of this paper is organized as follows: Chapter II, LS and the proposed objective function, and solved by the Chebyshev center. Chapter III, verify it by CRLB theory and simulations. Chapter IV, conclusions.

II. PROBLEM FORMULATION

Ranging value between the base station and target node is used as the constraint condition of the objective function in the multi-target location estimation method. Then get optimized localization results. For minimizing the maximum difference between the localization result and the real position of the target node, we propose an optimal algorithm based on Chebyshev Center, so as to improve the accuracy of the localization system.

Obtain the distance between the base station and the target node, and 1) Define the problem with distance value as the constraint condition; 2) Minimize the maximum error of the optimal estimation result and real position; 3) Establish the objective function based on the Chebyshev Center concept. This chapter will introduce the optimal method from problem definition and constraints.

A. Problem Definition

We consider the multi-target localization problem in 2D. In the localization problem, the node whose position is known and fixed is the base station (Anchor node), whose location unknown is the target node (Tag node). Including N tag nodes with unknown locations, denoted as $P_x = \{1, \dots, n\}$, the coordinates of i th tag node is $x_i = [x_i, y_i]$, $i \in P_x$, M ($M \geq 3$) anchor nodes with known positions, denoted as $P_a = \{n+1, \dots, n+m\}$. The distribution of at least three anchor nodes in 2D is nonlinear, where the coordinates of j th anchor is $a_j = [a_{jx}, a_{jy}]$, $j \in P_a$. The propagation of signals is environmentally sensitive, and the communication distance between nodes will reduce the accuracy of ranging. Define L_x and L_a as the link between the tags, between the anchor and tag respectively. The wireless communication network consists of all node sets $V = \{P_x, P_a\}$ and all link sets $E = \{L_x, L_a\}$, ie the undirected graph $G = \{V, E\}$.

B. Distance Constraint

The ranging-based optimal method is one of the main method of location. Generally, range-free localization algorithms is worse than the range-based localization algorithms. The anchor and tag can send and receive a pulse, and calculate the distance between the nodes based on the time difference of arrival. By using the ranging values of the tags and anchors of nonlinear arrangements in 2D, the estimate values of the target nodes have a constraint. Then attains better solutions through different optimal methods. According to the attribute of nodes, the ranging values are divided into two categories.

1) Uncertainty Constraint

The ranging value between tags be expressed as:

$$\hat{d}_{ij} = d_{ij} + n_{ij} \quad (1)$$

where, $d_{ij} = \|x_i - x_j\|$ is the distance between i th tag and j th tag, $|n_{ij}| \leq \gamma$ is the bound of ranging error. so $\underline{d}_{ij} < \hat{d}_{ij} < \overline{d}_{ij}$, where $\underline{d}_{ij} = \hat{d}_{ij} - \gamma$, $\overline{d}_{ij} = \hat{d}_{ij} + \gamma$.

Since the tag's location is unknown, the constraint interval based on the ranging between tags is uncertain and cannot be used directly to estimate. Target node can be estimated only if the tag communicating with the target node gets a initial estimate.

2) Certainty Constraint

The ranging value between tag and anchor is expressed as:

$$\hat{d}_{ik} = d_{ik} + n_{ik} \quad (2)$$

where, $d_{ik} = \|x_i - x_k\|$ is the distance between i th tag and k th anchor, $|n_{ik}| \leq \gamma$ is the bound of ranging error. so $\underline{d}_{ik} < \hat{d}_{ik} < \overline{d}_{ik}$, where $\underline{d}_{ik} = \hat{d}_{ik} - \gamma$, $\overline{d}_{ik} = \hat{d}_{ik} + \gamma$.

3) Distance Constrains Condition

According to the certainty constraint, the distance constrain interval of target m_i is:

$$C = \{m_i : \underline{d}_{ij} < \|m_i - m_j\| < \overline{d}_{ij}, \forall (i, j) \in L_x, \underline{d}_{ik} < \|m_i - a_k\| < \overline{d}_{ik}, \forall (i, k) \in L_a\} \quad (3)$$

III. ESTABLISHMENT OF OBJECTIVE FUNCTION

The objective function is an expression related to variables and reflect the purpose of an algorithm, generally pursuing the localization estimate value nearest to the result. Combining the constraints proposed above, solve the objective function can find the most suitable point in the constraint region. There are many methods for solution of the objective function, but their effects vary a lot. The following objective function could be used to minimize maximum error between result and real position.

A. Objective Function Based on LS

LS is a classical optimal method. The objective function based on LS commitment to the smallest overall error in the current wireless sensor network [11]–[13], so the objective function is:

$$\min_{m_i} \sum_{i=1}^n \sum_{j=1}^m (\|m_i - a_j\| - \hat{d}_{ij})^2 \quad (4)$$

where, n and m is the number of Tags and Anchors respectively. This objective function requires that the distance between result and anchor is the closest to ranging value. LS method depend on ranging value, but have to allow for the ranging error, which will also bring errors to the localization result. LS has limitations:

1) The LS objective function's priority is that the total error in the entire wireless sensor network is the smallest, which results in a large error between some localization results and the true position.

- 2) Due to the limitation of the localization mode, LS objective function is non-convex. That makes solving more difficult. Optimal algorithm may fall into local optimum.

LS objective function barely relies on the ranging and needs the ranging error distribution. Therefore, it is urgently needed to establish a objective function to solve the problem of the least squares objective function.

B. Objective Function Based on Real position

We propose an objective function that considers the true position of the target node, and relaxes the convex function to simplify the program. The following describes the construction of the objective function.

The main purpose is minimizing the possibly maximum error of the localization results. First, arbitrarily choose a point in the constraint interval as the result (the optimal method will consider that any position in the constraint region may be the real position). Correspondingly find a "real position" with maximum error as result, which means that the error of current result is the largest. Then concentrate on minimizing this maximum error, that is, Minmax objective function.

1) Propose Objective Function

If the result of i th tag is \hat{x} , the maximum error is:

$$\max_{m_i} \sum_{i=1}^n \|m_i - x_i\|^2 \quad (5)$$

where, \hat{x} is the worst result in the constraint interval. Then find the appropriate \hat{x} to minimize the maximum error.

$$\begin{aligned} \min_{m_i} \max_{m \in C} Tr((m - \hat{x})(m - \hat{x})^T) \\ s.t. \\ \underline{d}_{ij} < \|m_i - m_j\| < \overline{d}_{ij}, \forall (i, j) \in L_x \\ \underline{d}_{ik} < \|m_i - a_k\| < \overline{d}_{ik}, \forall (i, k) \in L_a \end{aligned} \quad (6)$$

where, $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in R^{2n}$ is the real position, $\mathbf{m} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n]^T \in R^{2n}$ is the possible result, $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_n]^T \in R^{2n}$ is the estimate result of optimal methods.

The uncertainty constraint is:

$$\|m_i - m_j\| = e_{(2i-1)(2j-1)}^T \mathbf{m} \mathbf{m}^T e_{(2i-1)(2j-1)}^T + e_{(2i)(2j)}^T \mathbf{m} \mathbf{m}^T e_{(2i)(2j)}^T \quad (7)$$

where, $e_{(i)(j)} \in R^{2n}$ is a column vector, the i th term is 1, the j th term is -1, the i th term is 1, and other term is 0;

The certainty constraint is:

$$\|m_i - a_k\| = a_k a_k^T - 2a_k \mathbf{m}^T e_{2i-1} - 2a_k \mathbf{m}^T e_{2i} + e_{2i-1}^T \mathbf{m} \mathbf{m}^T e_{2i-1} + e_{2i}^T \mathbf{m} \mathbf{m}^T e_{2i} \quad (8)$$

where, $e_i \in R^{2n}$ is a column vector, the i th term is 1, and other term is 0;

2) Relaxation

Since the non-convex constraint interval is not suitable for solving, we will relax and transform it into a convex optimal problem. This paper does not consider the uncertainty constraints between target nodes, and only relaxes the deterministic constraints.

It can be obtained directly from (3).

$$\underline{d}_{ik}^2 < \|m - a_k\|^2 < \overline{d}_{ik}^2 \quad (9)$$

Which is:

$$\underline{d}_{ik}^2 < \mathbf{m} \mathbf{m}^T - 2a_k \mathbf{m} + a_k a_k^T < \overline{d}_{ik}^2 \quad (10)$$

Let $\Delta = \mathbf{m} \mathbf{m}^T$, (10) can be written as:

$$\begin{aligned} \Delta - 2a_k \mathbf{m} + a_k a_k^T - \overline{d}_{ik}^2 &< 0 \\ \Delta - 2a_k \mathbf{m} + a_k a_k^T - \underline{d}_{ik}^2 &> 0 \end{aligned} \quad (11)$$

Because $\Delta = \mathbf{m} \mathbf{m}^T$ is not affine, the optimal problem is non-convex. We make relaxation $\Delta \geq \mathbf{m} \mathbf{m}^T$.

The problem (6) written as:

$$\begin{aligned} \min_{\hat{\mathbf{x}}} \max_{(\mathbf{m}, \Delta) \in \mathcal{K}} Tr(\Delta - 2\hat{\mathbf{x}} \mathbf{m}^T + \hat{\mathbf{x}} \hat{\mathbf{x}}^T) \\ s.t. \\ \Delta - 2a_k \mathbf{m} + a_k a_k^T - \overline{d}_{ik}^2 &< 0 \\ \Delta - 2a_k \mathbf{m} + a_k a_k^T - \underline{d}_{ik}^2 &> 0 \\ \Delta &\geq \mathbf{m} \mathbf{m}^T \end{aligned} \quad (12)$$

C. Solving Method for Function

1) Solving LS function

The optimal problem based LS is:

$$\begin{aligned} \min_{m_i} \sum_{i=1}^n \sum_{j=1}^m (\|m_i - a_j\| - \hat{d}_{ij})^2 \\ s.t. \\ \underline{d}_{ij} < \|m_i - m_j\| < \overline{d}_{ij}, \forall (i, j) \in L_x \\ \underline{d}_{ik} < \|m_i - a_k\| < \overline{d}_{ik}, \forall (i, k) \in L_a \end{aligned} \quad (13)$$

LS method has a non-convex objective function, and cannot be solved by the convex optimal tool. Fmincon optimal toolbox in Matlab is specifically used for solving nonlinear constrained objective functions. Fmincon needs the initial value. So setting a uniform initial value for different target nodes in the solution. Consequently reduce the precision of localization. Use the iterative method to solve the initial value problem. In the follow-up study, we obtain the initial value of fmincon with some other optimal methods.

2) Solving Minmax function

The minmax function's target is to find a point \hat{x} satisfies that the furthest distance from the arbitrary point m within the constraint interval is smallest. Because the real location is unknown, point m may be the real location of the target node. Obviously the position of the point \hat{x} is related to the shape of the constraint interval.

In order to solve the minmax problem, we introduce a graphics concept, Chebyshev circle, which is defined as a circle with the smallest radius enclose the entire interval. Similar to the circumscribed circles, but the Chebyshev circle cannot intersect with all vertices of the interval. The result \hat{x} is the center of the Chebyshev circle corresponding to the constraint interval. When the constraint interval is non-convex, the Chebyshev center may fall outside the constraint interval, which obviously reduces the accuracy.

The definition of Chebyshev center is:

$$\mathbf{x}_{cheby} = \arg \min_{\hat{\mathbf{x}}} \max_{\mathbf{m} \in C} Tr((\mathbf{m} - \hat{\mathbf{x}})(\mathbf{m} - \hat{\mathbf{x}})^T) \quad (14)$$

Any point \mathbf{m} in the interval C must included in the Chebyshev circle:

$$\min_{\mathbf{x}_{cheby}} \{r_{cheby} : \|\mathbf{y} - \mathbf{x}_{cheby}\| \leq r_{cheby}\} \quad (15)$$

where, r_{cheby} is the radius of Chebyshev circle.

Constraint (3) can be written as:

$$C = \{\mathbf{x}_{cheby} + \mu \mid \|\mu\| \leq r_{cheby}\} \quad (16)$$

where, μ is 2D vector.

The (17) transformed to:

$$\begin{aligned} & \min r_{cheby} \\ & s.t. \\ & \mathbf{m} = \mathbf{x}_{cheby} + \mu, \|\mu\| \leq r_{cheby} \\ & \Delta - 2a_k m + a_k a_k^T - \underline{d}_{ik}^2 < 0 \\ & \Delta - 2a_k m + a_k a_k^T - \underline{d}_{ik}^2 > 0 \\ & \Delta \geq \mathbf{m} \mathbf{m}^T \end{aligned} \quad (17)$$

IV. RESULTS AND DISSCUTION

The experimental site is a square, and four base stations are settled in each corners of the experimental area. The positions of the base stations are $[10, 10]$, $[10, 30]$, $[30, 10]$, $[30, 30]$, respectively. The tags will be randomly distributed throughout overall square.

A. Theoretical Results (CRLB)

In order to theoretically verify the accuracy of the proposed localization method, consider of the Cramér–Rao bound (CRB). CRB is the lowest theoretically achievable value of algorithm, so it is also known as CRLB. The merits and drawbacks and the improvement space of the algorithm can be shown by comparing with CRLB.

The CRLB has need for joint density function of all parameters, in this paper, which is the probability density function of ranging error:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (18)$$

where, σ^2 is the variance of ranging error, μ is expectation. (Because CRLB is for unbiased estimation, $\mu = 0$) The

distribution of ranging error is hard to obtain, compare with simulation result only.

Fig.1 shows CRLB uniformly of 121 sampled-data with the ranging error variance to 0.1. There is higher localization accuracy between each two base stations, and the minimum error was 0.27; diagonally part outside the base station shows poor performance and the localization error is up to 0.32. Obtain the theoretical minimum error, and then compare it with the simulation experiment.

The Fisher matrix is:

$$J = \begin{bmatrix} X & \theta \\ \theta & Y \end{bmatrix} \quad (19)$$

where,

$$\begin{aligned} X &= \sum_{n=1}^N \left\{ \frac{(x_1 - a_{1,n})^2}{\sigma^2 \|x - a_n\|} \right\} \\ \theta &= \sum_{n=1}^N \left\{ \frac{(x_1 - a_{1,n})(x_2 - a_{2,n})}{\sigma^2 \|x - a_n\|} \right\} \\ Y &= \sum_{n=1}^N \left\{ \frac{(x_2 - a_{2,n})^2}{\sigma^2 \|x - a_n\|} \right\} \end{aligned} \quad (20)$$

where, N is the number of anchors, $x = (x_1, x_2)$ is estimate result, $a_n = (a_{1,n}, a_{2,n})$ is n th anchor.

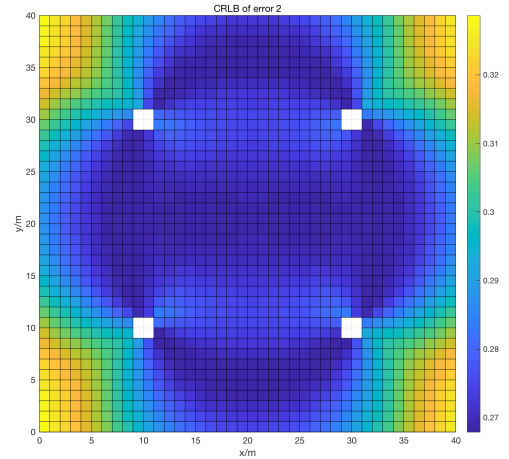


Fig. 1. CRLB distribution under variance 1.

B. Simulation Results

The control group are as follows: minmax method based on Chebyshev Center (che), adopt CVX toolbox; optimal method based on LS, adopt fmincon. We selected Active set (act) and Sequence Quadratic Program (sqp) in the algorithm of fmincon.

1) Comparison with CRLB

With the same configuration, with variance of the ranging error is 0.1, sampling points are uniformly taken. The mean square error(MSE) of the proposed optimal method is: che 0.4237, sqp 0.6115, act 0.6008, and the average value of

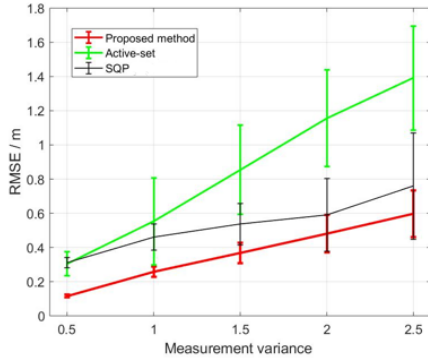


Fig. 2. Performance comparison of methods under different variances.

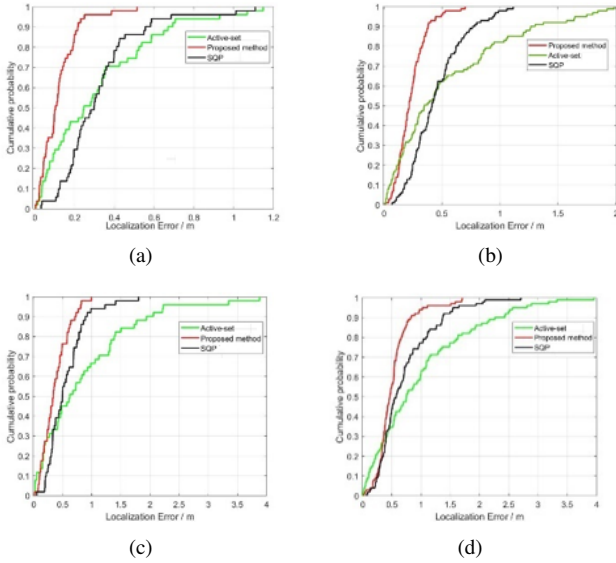


Fig. 3. CDF of 100 random targets under variances of 0.5, 1.0, 1.5, 2.0, 2.5.

CRLB is 0.2862. The comparison shows that the results of proposed method are closer to the CRLB.

2) Comparison of Methods under Different Variances

In order to verify the effect of different optimal methods under longer distance communication between the base station and the target node, expand experimental area. The experimental scenario is set to $100 \times 100 m^2$ square area. The positions of the base stations are $[0, 0]$, $[0, 100]$, $[100, 0]$, $[100, 100]$ respectively. The tags will be randomly distributed throughout overall square area.

Multiple experiments were performed on the three sets of methods with different ranging error variances. From Fig.2 we can see that under the same error, proposed method has the lowest RMSE and the act algorithm has highest. The RMSE of three methods increase as the variance increases, greatly act algorithm grow faster. Che maintains good performance when the error is large.

Fig.3 is cumulative distribution function graph (CDF) of three methods. Histogram offer less information, in order to

have a comprehensive understanding of the distribution of localization error, we selected the CDF. The experimental settings of the four images are the exactly same, but localization errors. It can be seen from the range of the abscissa in fig.8(a)-(d) that the main regions of the error distribution grow as the error increases. Fig.8(a) illustrate that that the main distribution range of che is $[0-0.2]$, up to 0.5; and the error of act and sqp method is evenly distributed in $[0-0.6]$, some exceeding 1. It is easy to see that 1) che method not only has smaller RMSE, but also each error is concentrated in a small interval. The error distributions of the other method are more scattered. 2) both act and sqp method have excessive error (more than twice the RMSE). Che method focus on minimizing the maximum error, there is no excessive error.

V. CONCLUSIONS

In this paper, a localization optimal method with maximum error minimization is proposed, and the objective function is solved by the relaxation and optimal toolbox CVX. In the simulation, the comparison with the LS is carried out in many aspects. It is verified that the accuracy of the proposed method is higher, and the localization error is lower when the ranging error is larger. Subsequent experiments can be considered from: 1) solving the initial value of the LS function, 2) localization between the target nodes, and so on.

REFERENCES

- [1] Yu P . A review: wireless sensor networks localization[J]. Journal of Electronic Measurement and Instrument, 2011, 25(5):389-399.
- [2] Cardinali R , Nardis L D , Lombardo P , et al. Lower bounds for ranging accuracy with multi band OFDM and direct sequence UWB signals[C]// IEEE International Conference on Ultra-wideband. IEEE, 2005.
- [3] Juan L , Zanyi H , Yuxi Z , et al. Indoor cooperation localization algorithm based grid in wireless networks[J]. Journal of Huazhong University of Science and Technology, 2015.
- [4] Du Y , Wei P , Zhang H . Semidefinite programming approach for TDOA/GROA based source localization[J]. Journal of Systems Engineering and Electronics, 2015, 26(4):680-687.
- [5] Lee J Y , Scholtz R A . Ranging in a dense multipath environment using an UWB radio link[J]. Proc. IEEE J. Sel. Areas Commun, 2002, 20(9):1677-1683.
- [6] Gezici S , Poor H V . Position Estimation via Ultra-Wide-Band Signals[J]. Proceedings of the IEEE, 2009, 97(2):386-403..
- [7] Deligiannis N, Louvros S, Kotsopoulos S. Optimizing location positioning using hybrid TOA-AOA techniques in mobile cellular networks[C]// International Conference on Mobile Multimedia Communications. 2007.
- [8] Niculescu D , Nath B . Ad hoc positioning system (APS) using AoA[C]// INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies. IEEE, 2003.
- [9] Shen G , Zetik R , Thoma R S . Performance comparison of TOA and TDOA based location estimation algorithms in LOS environment[C]// Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on. IEEE, 2008.
- [10] Mallat A , Gezici S , Dardari D , et al. Statistics of the MLE and Approximate Upper and Lower Bounds—Part I: Application to TOA Estimation[J]. IEEE Transactions on Signal Processing, 2014, 62(21):5663-5676.
- [11] Shi X , Mao G , Anderson B D O , et al. Robust Localization Using Range Measurements with Unknown and Bounded Errors[J]. IEEE Transactions on Wireless Communications, 2017:1-1.
- [12] Bonnet V , Claudia Mazzà, Fraisse P , et al. A least-squares identification algorithm for estimating squat exercise mechanics using a single inertial measurement unit[J]. Journal of Biomechanics, 2012, 45(8):0-1477.
- [13] Aghamolki H G , Miao Z , Fan L . SOCP Convex Relaxation-Based Simultaneous State Estimation and Bad Data Identification[J]. 2018.