

A Collaborative and Predictive Localization Algorithm for Wireless Sensor Networks

Yuan Liu, Junjie Chen

Southeast University

Nanjing 210096, Jiangsu - China

[e-mail: liuyseu@gmail.com, inschenjj@seu.edu.cn]

*Corresponding author: Junjie Chen

*Received December 20, 2016; revised March 31, 2017; accepted April 27, 2017;
published July 31, 2017*

Abstract

Accurate locating for the mobile target remains a challenge in various applications of wireless sensor networks (WSNs). Unfortunately, most of the typical localization algorithms perform well only in the WSN with densely distributed sensor nodes. The non-localizable problem is prone to happening when a target moves into the WSN with sparsely distributed sensor nodes. To solve this problem, we propose a collaborative and predictive localization algorithm (CPLA). The Gaussian mixture model (GMM) is introduced to predict the posterior trajectory for a mobile target by training its prior trajectory. In addition, the collaborative and predictive schemes are designed to solve the non-localizable problems in the two-anchor nodes locating, one-anchor node locating and non-anchor node locating situations. Simulation results prove that the CPLA exhibits higher localization accuracy than other tested predictive localization algorithms either in the WSN with sparsely distributed sensor nodes or in the WSN with densely distributed sensor nodes.

Keywords: Wireless sensor networks, localization algorithms, non-localizable problem, Gaussian mixture model, collaborative and predictive schemes

This work was supported by the National Science and Technology Support Project of China under Grant No.2014BAD08B03, the Three New Aquaculture Project of Jiangsu Province under Grant No.Y2016-3, the Science and Technology Special Fund Project of Northern Jiangsu Province under Grant No.BN2014085, and the Agricultural Science and Technology Support Project of Jiangsu Province under Grant No.BE2014312.

1. Introduction

Accurate localization techniques have received considerable attention over the past years, especially as the wireless sensor networks (WSNs) have been applied in various fields [1]. Undoubtedly, the global positioning system (GPS) is the most widely used localization technique. Generally, only a few sensor nodes in WSNs are equipped with the GPS, while the others obtain the location information by references or other devices. The GPS is not suitable for the large-scale applications on account of the high cost and high energy consumption. The Received Signal Strength Indicator (RSSI) is a common non-GPS-based localization technique and has been applied to many practical applications. As the RSSI is susceptible to the shadow fading, multipath effect and non-line-of-sight communication, large measurement errors are produced in the RSSI-based methods [2]. Besides, the RSSI presents low accuracy and poor stability in mobile localization.

A Iteration-localizing algorithm is proposed in [3] to estimate the trajectory of a mobile target. Sensor nodes measure the target through the acoustic signal which is emitted from the target. However, this method is only able to predict the next trajectory point by more than 3 sensor nodes. In addition, it neglects the non-localizable problems. Considering the non-line-of-sight propagation of signals, sensor nodes are always non-uniformly distributed in many actual applications, especially in the indoor environment, irregular scene and dense barrier scenes. Besides, the environmental interferences will lead to the degradation of communication quality. Because most of the localization researches of WSN are implemented with good network connectivity and dense distributed sensor nodes, a mobile target may not be located or tracked in the WSN with sparsely distributed sensor nodes.

The Approximate Point in Triangulation (APIT) is a classic range-free localization algorithm due to the advantages of low communication overhead, high localization accuracy and ease-deployment [4]. Motivated by achieving the accurate localization for a mobile target in the WSN with sparsely distributed sensor nodes, we propose a collaborative and predictive localization algorithm (CPLA) based on the APIT algorithm. We first analyze the non-localizable problems that the mobile target cannot be located or detected in the two-anchor nodes locating, one-anchor node locating and non-anchor node locating situations. Then, the Gaussian mixture model (GMM) is introduced to train the prior trajectory and predict the posterior trajectory for the mobile target. In addition, we design collaborative and predictive schemes to realize accurate localization in the non-localizable situations by combining the APIT algorithm with the GMM. To the best of our knowledge, it is the first time to solve the non-localizable problems by using such collaborative and predictive schemes. The main contributions of this paper are briefly summarized as follows:

- (1) Characterize and analyze the three non-localizable problems for a mobile target in the WSN with sparsely distributed sensor nodes.
- (2) Build the GMM to achieve the accurate and real-time trajectory prediction.
- (3) Design collaborative and predictive schemes to solve the non-localizable problems based on the APIT algorithm and the GMM.

The rest of this paper is organized as follows. Section 2 describes the related work for the mobile localization algorithms. Section 3 analyzes the non-localizable problems in the WSN with sparsely distributed sensor nodes. Section 4 introduces the basic theories and localization schemes for the CPLA. The parameters analysis and simulation results are discussed in Section 5. Section 6 draws the conclusion.

2. Related Work

Extensive studies have been conducted to the performance in high-accuracy and real-time of the localization algorithms. As shown in **Table 1**, the localization algorithms for a mobile target can be categorized into the geometric analysis method, scale transformation method and the probabilistic model.

Table 1. Classification of localization algorithms

Localization mechanism	Localization algorithm
Geometric analysis	[5], [6], [7], [8], [9]
Scale transformation	[15], [16], [17], [18], [19]
Probabilistic model	[10], [11], [12], [13], [14], [20], [21], [22], [27], [28],

Many researches have focus on the mobile localization based on the geometry analysis. An improved APIT algorithm is proposed in [5] that divides the sensing areas into the overlapping sub-regions and non-overlapping regions. Some super anchors are utilized to distinguish which sub-regions the target belongs to by the RSSI. Then, the APIT algorithm is implemented for localization in each sub-region. [6] divides the localization situations into the consistent case and the inconsistent case. The localization computation of the intersected rings are converted into two weighted convex optimization problems. Then, these problems are solved by a distributed alternating projection algorithm. An analytical geometrical localization method is introduced in [7]. Sensor nodes initialize the residence area by two distant anchor nodes. According to one of the distant anchor node, the radius and the half-length of the chord of the target are calculated. In this way, the location of the target is estimated by the radius, the half-length of the chord and the sagitta of an arc. A range-free localization method is also designed in [8] based on the analytical geometry of an arc. The intersection point of two perpendicular bisectors of the chords is regarded as the estimated position of sensor node. A half symmetric lens-based localization algorithm is proposed in [9]. According to the Voronoi diagram and RSSI, the target is located by only two anchor nodes. This method yields a small residence area by the half-symmetric lens.

Monte Carlo Location (MCL) [10] is one of the classical mobile localization algorithm for WSNs. Based on the Bayesian theory, the MCL estimates the possible location of target by the weighted posterior probability distributions. [11] proposes a mobile-Assisted Monte Carlo localization algorithm that only relies on the direct arriver and leaver information from a single mobile-assisted seed. [12] introduces a power mapping algorithm based on the discrete antithetic Markov Monte Carlo method. A mathematical model for the discrete level distance measurement is established to reduce the distance error of anchor nodes and the redundant information for each individual measurement. A refinement of the MCL is given in [13] that consists of three phases, including the sample generating, sample filtering and location estimation phases. The historical locations of anchor nodes and the RSSI are used to derive three constraints. Then, the target is located by the theoretical analysis and the three constraints. [14] presents a Monte Carlo-based algorithm with the area-based and neighbor-based filtering. In the predictive phase, samples are used to describe the posterior probability distribution of the target, and they are optimized by the filters in the filtering phase.

The multidimensional scaling-MAP (MDS-MAP) algorithm [15] has always been widely used for WSNs localization according to the dimensions reduction and data projection techniques. A cluster-based MDS algorithm is introduced in [16]. The clusters are formed by their own coordinate systems by the MDS-MAP. Then, these coordinate systems are merged

into a coordinate system. A weighted MDS method is proposed in [17] based on the time-of-arrival mobile localization. The algorithm achieves the Cramér Rao Bound at moderate noise level before the threshold effect occurs by the Eigen decomposition. [18] divides the sensor nodes into two groups according to the connectivity. Then, the MDS algorithm is carried out in the group with the higher connectivity, and the least square method is used for another group, respectively. [19] gives an improved MDS algorithm based on the hop counts. As the real number hop-count is much accurate, the integer hop-count is transformed into the real number by partitioning the node's one-hop neighbors into three disjoint subsets.

[20] proposes a predictive clustering algorithm that divides the sensor nodes into the static and dynamic clusters. Then, the Kalman, Alpha-Beta, or Particle Filters are used to predict the location. An energy-efficient collaborative algorithm is introduced in [21] based on the neural network aggregation model and Gaussian particle filtering (GPF) estimation. [22] proves a seamless data mining algorithm to explore the temporal movement patterns of the target, which gives two predictive strategies without predefining the segmented time unit.

3. Problem Formulation

In this section, we introduce the localization scheme of the APIT algorithm and discuss the influence of the WSN distribution on the localization performance of the APIT algorithm. Moreover, we give a brief analysis of the non-localizable problems for a mobile target. We define the following metrics by:

- (1) Localization rate: the percentage of the trajectory points which have been successfully located within the residence area.
- (2) Network connectivity: the average number of sensor nodes in the communication range of an anchor node.

3.1 Localization Mechanism of the APIT Algorithm

The APIT is an ideal range-free localization algorithm with simple mode, low-cost and high localization accuracy. When a mobile target enters the sensing area, anchor nodes (i.e., sensor nodes with known locations) detect the mobile target in their communication ranges and transmit the information of labels and locations to the base station. The location of the mobile target is calculated in the base station. In the APIT algorithm, the anchor nodes that detecting the mobile target sequentially form many triangles. The overlap region of these triangles is defined as the residence area as shown in Fig. 1. The centroid position of the residence area is regarded as the estimated location of the mobile target.

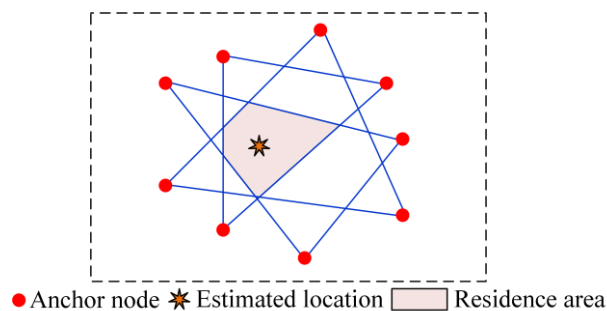


Fig. 1. Localization scheme of the APIT algorithm

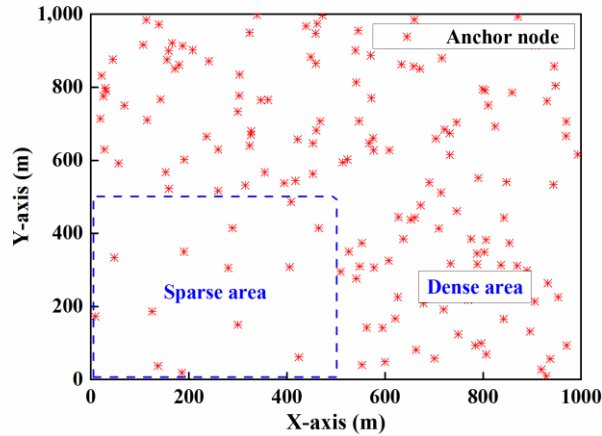


Fig. 2. WSN with irregular nodes distribution

3.2 WSN with Sparsely Distributed Sensor Nodes

In practical applications, the distribution of sensor nodes are always irregular in consideration of the restriction of the terrain and obstacles. Fig. 2 gives a case of the WSN with irregular nodes distribution in a $1000\text{ m} \times 1000\text{ m}$ area. We divide the sensing area into 2 sub-areas, including the sparse area and dense area. In the sparse area, only a few anchor nodes are randomly deployed. Normally, the size of residence area decisively affects localization accuracy [5-9]. The smaller the residence area is, the higher the localization accuracy is. Therefore, the APIT algorithm exhibits high localization accuracy in the dense area than it is in the sparse area.

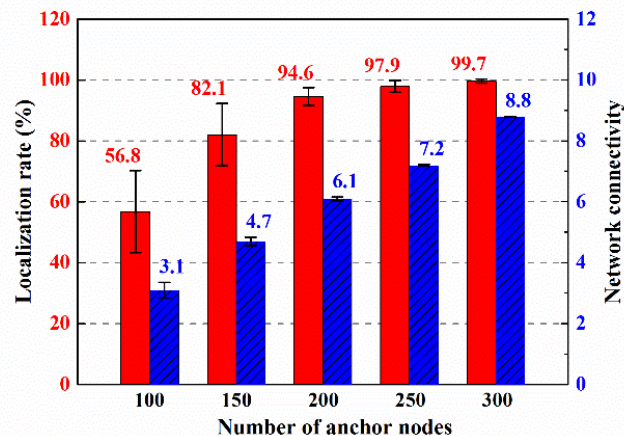


Fig. 3. Impact of the density of anchor nodes on localization performance

Fig. 3 illustrates the effects of the density of anchor nodes on the localization rate and on the network connectivity by the APIT algorithm. The simulation results are expressed by the mean values with the standard deviations. The localization rate increases with the growth of the number of anchor nodes. In addition, the network connectivity is rising as the density of anchor nodes increase. That is, the increase of the number of anchor nodes contributes to the

improvement of localization performance. In addition, the network connectivity has positive correlations with the localization rate. To ensure more than 90% localization rate, the network connectivity should not be less than 6.1 and more than 200 anchor nodes are deployed for localization.

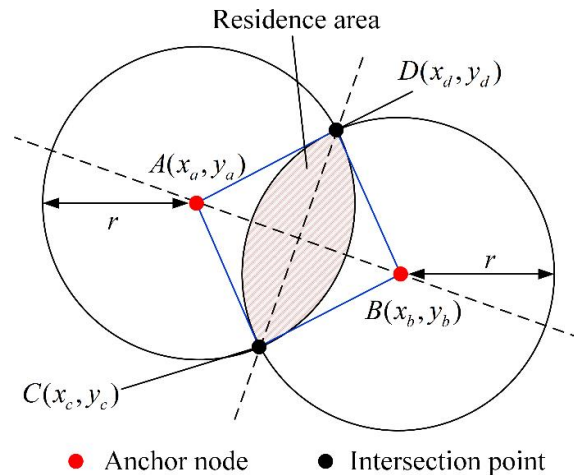


Fig. 4. Localization by two-anchor nodes

3.3 Non-localizable Problems

As discussed above, the localization performance of the APIT algorithm is strongly dependent on the distribution of WSN. When a mobile target moves in the WSN with sparsely distributed sensor nodes, the non-localizable problems are prone to occur. Thus, we discuss three non-localizable situations for the localization of the mobile target, including two-anchor nodes locating, one-anchor node locating and non-anchor node locating situations.

3.3.1 Two-anchor nodes locating situation

Generally, a mobile target is located in the triangle region, which is formed by at least 3 anchor nodes. However, the mobile target may not be located in the WSN with sparsely distributed sensor nodes. **Fig. 4** depicts the situation that the mobile target is outside the triangle region, where only 2 anchor nodes (i.e., the anchor node $A(x_a, y_a)$ and anchor node $B(x_b, y_b)$) are used for localization. The node $C(x_c, y_c)$ and node $D(x_d, y_d)$ are two intersection points of the circle centered in anchor node A and the circle centered in anchor node B with the same radius r . According to the geometric theory, the residence area is symmetrical along with the line l_{AB} . In this situation, the node flip ambiguity problem occurs and it is difficult to distinguish the estimated locations [23].

3.3.2 One-anchor node locating situation

Despite of the two-anchor nodes locating problem, the mobile target may move into the area where only one anchor node is available for localization. As shown in **Fig. 5**, a mobile target is detected by the anchor node $A(x_a, y_a)$. In this situation, the residence area covers most of the communication scope of the anchor node A ; the node $C(x_c, y_c)$, node $D(x_d, y_d)$ and node $E(x_e, y_e)$ are three of the potential locations for the mobile target, such as. Therefore, the

mobile target cannot be located in one-anchor node locating situation.

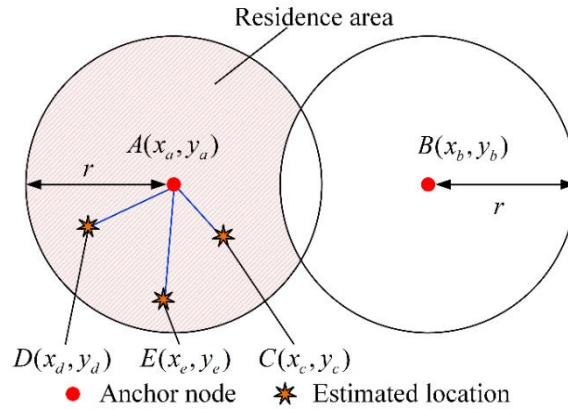


Fig. 5. Localization by one-anchor node

3.3.3 Non-anchor node locating situation

Normally, the communication scopes of the RF devices are susceptible to the environmental interferences. The interferences cause the signal fading and form the blind zone where the mobile target cannot be detected. **Fig. 6** describes the situation that the actual communication scopes of the anchor node $A(x_a, y_a)$ and anchor node $B(x_b, y_b)$ are less than the ideal communication ranges (i.e., the maximum communication scope) with $r' < r$ and $r'' < r$. If a mobile target moves into the blind zone, the mobile target (i.e., the location of node $C(x_c, y_c)$) will be missing in the sensing area. Hence, the APIT algorithm presents poor localization performance w environmental interferences.

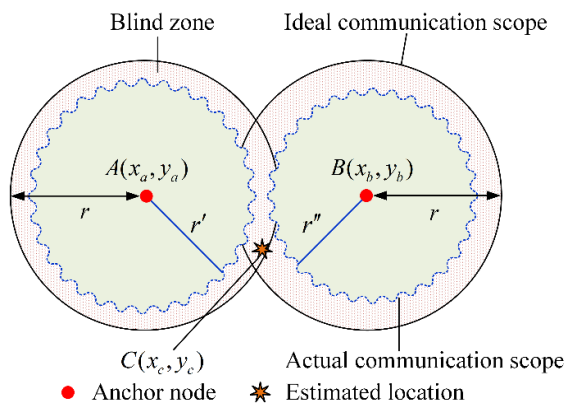


Fig. 6. Localization by non-anchor nodes

4. Proposed Localization Algorithm

The basic theories and localization mechanism of the CPLA are introduced in this section. We first establish the localization model for the mobile target. Then, we design the Gaussian mixture models (GMM) and introduce how to predict the posterior trajectory of the mobile target. Moreover, we formulate the collaborative and predictive schemes to solve the non-localizable problems for the WSN with sparsely distributed anchor nodes.

4.1 Description of Localization Model

Let $\mathbf{T} = \{z_1, z_2, \dots, z_N\}$ be the trajectory for the mobile target, where $z_i = (x_i, y_i)$ is the i -th ($0 \leq i \leq N$) discrete trajectory point. The trajectory is modeled by the two-dimensional plane as

$$\begin{aligned} \mathbf{T} &= \{\mathbf{T}_x, \mathbf{T}_y\} \\ &= \{(z_{1x}, z_{2x}, \dots, z_{Nx})^T, (z_{1y}, z_{2y}, \dots, z_{Ny})^T\} \end{aligned} \quad (1)$$

where \mathbf{T}_x is the trajectory projection vector in the X-axis, \mathbf{T}_y is the trajectory projection vector in the Y-axis and z_{ix} is the projection of the i -th ($0 \leq i \leq N$) trajectory point onto the X-axis.

Assuming that $\mathbf{Z}_{tra} = [z(x, y)]$ denotes the training dataset of the trajectory and accords with the Gaussian distribution [24]. Then, \mathbf{Z}_{tra} is described by a random variable dataset as

$$\mathbf{Z}_{tra} = \{f(z_1), f(z_2), \dots, f(z_N)\} \quad (2)$$

where $f(z) \sim p(\mu(z), \Sigma(z))$. $\mu(z)$ is the mean with $\mu(z) = E[f(z)]$ and $\Sigma(z)$ is the covariance matrix with $\Sigma(z) = E[(f(z) - \mu(z))(f(z) - \mu(z))]$.

4.2 Description of the GMM

Theoretically, any probability distribution can be smoothly approximated by the Gaussian mixture model (GMM) [25, 26]. Because the GMM is the linear combination of K independent Gaussian distributions (or K components), the probability density function (PDF) of a Gaussian distribution is modelled as

$$p_j(z_j; \mu_j, \Sigma_j) = \frac{1}{\sqrt{(2\pi)^{|\Sigma_j|}}} \exp\left[-\frac{1}{2}(z_j - \mu_j)^T \Sigma_j^{-1} (z_j - \mu_j)\right] \quad (3)$$

and the PDF of the GMM is modelled as

$$p(z; \mu, \Sigma) = \sum_{j=1}^K \alpha_j p_j(z_j; \mu_j, \Sigma_j) \quad (4)$$

where α_j is the weight of the j -th Gaussian distribution with $\sum_{j=1}^K \alpha_j = 1$ ($\alpha_j > 0$). We learn from (4) that the PDF of the GMM is determined by the weight and the PDF of K components. We assume that $\theta_j = (\alpha_j, \mu_j, \Sigma_j)$ denotes the parameters of the j -th Gaussian distribution and $\Theta = (\theta_1, \theta_2, \dots, \theta_K)^T$ denotes the parameters of K Gaussian distributions.

Let z be the dataset with $z = \{z_1, z_2, \dots, z_N\}$, where $\{z_1, z_2, \dots, z_N\}$ are independent and identically distributed (i.i.d.) [24]. We express the PDF of z as

$$\begin{aligned} p(z; \Theta) &= p(z_1, z_2, \dots, z_N; \Theta) \\ &= \prod_{i=1}^N p(z_i; \Theta) \end{aligned} \quad (5)$$

To let the probability in (5) be the maximum value, we search for the optimal parameter as

$$\Theta^* = \arg \max p(z; \Theta) \quad (6)$$

As $p(z; \Theta)$ is a little value, the result of the product in (5) will be too little to be correctly expressed. Hence, the PDF of z is represented by the log-likelihood function as

$$\begin{aligned} \log p(z; \Theta) &= \log \prod_{i=1}^N \sum_{j=1}^M \alpha_j p_j(z_j; \mu_j, \Sigma_j) \\ &= \prod_{i=1}^N \log \sum_{j=1}^M \alpha_j p_j(z_j; \mu_j, \Sigma_j) \end{aligned} \quad (7)$$

Obviously, it is quite difficult to find the optimal parameter in (7). We apply the expectation maximization (EM) method to solve the problem by the following steps:

(1) **E-step**. Estimate the posterior probability β of each Gaussian distribution for the dataset by

$$\beta_j = E(\alpha_j | z_j; \theta_j) = \frac{\alpha_j p_j(z_j; \theta_j)}{\sum_{j=1}^M \alpha_j p_j(z_j; \theta_j)}. \quad (8)$$

(2) **M-step**. When the posterior probability of each Gaussian distribution is obtained, update the parameters $\theta_j = (\alpha_j, \mu_j, \Sigma_j)$ by

$$\begin{cases} \alpha_j = \frac{\sum_{i=1}^N \beta_{ij}}{N} \\ \mu_j = \frac{\sum_{i=1}^N \beta_{ij} z_{ij}}{\sum_{i=1}^N \beta_{ij}} \\ \Sigma_j = \frac{\sum_{i=1}^N \beta_{ij} (z_i - \mu_j)(z_i - \mu_j)^T}{\sum_{i=1}^N \beta_{ij}} \end{cases}. \quad (9)$$

Repeat the **E-step** and **M-step**, and update the posterior probability and the key parameters to meet the conditions

$$|l(z | \Theta) - l(z | \Theta')| < \varepsilon. \quad (10)$$

Generally, we set $\varepsilon = 10^{-5}$. In this way, we establish the optimal GMM based on the historical trajectory dataset.

4.3 Trajectory Prediction

According to the GMM modelled above, we achieve the trajectory prediction of the mobile target through the Gaussian mixture regression (GMR) function. The observation function for the trajectory vectors is expressed as

$$z = g(t) + p(0, \sigma_N^2) \quad (11)$$

where $p(0, \sigma_N^2)$ is the Gaussian noise. Assume that \mathbf{Z}_{tes} denotes the test dataset with $\mathbf{Z}_{tes} = [\dot{z}(\dot{x}, \dot{y})]$. As the \mathbf{Z}_{tra} and \mathbf{Z}_{tes} accord with the Gaussian distribution, the joint Gaussian distribution of the training dataset and test dataset is described as

$$\begin{bmatrix} z \\ \dot{z} \end{bmatrix} \sim p\left(\mu, \begin{bmatrix} \Sigma & \dot{\Sigma}^T \\ \dot{\Sigma} & \ddot{\Sigma} \end{bmatrix}\right) \quad (12)$$

where Σ is the covariance matrix of z , $\dot{\Sigma}$ is the covariance matrix of z and \dot{z} , and $\ddot{\Sigma}$ is the covariance matrix of \dot{z} . Then, the conditional PDF is described as

$$p(\dot{z} | z) = \sum_{j=1}^M \alpha_j p(\dot{z} | z; \hat{\mu}_j, \hat{\Sigma}_j) \quad (13)$$

where $\hat{\mu} = E[\dot{z} | z] = \dot{\mu} + \dot{\Sigma} \Sigma^{-1} (z - \mu)$ and $\hat{\Sigma} = \text{Var}[\dot{z} | z] = \ddot{\Sigma} - \dot{\Sigma} \Sigma^{-1} \dot{\Sigma}^T$. The predicted trajectory points of \dot{z} are calculated by

$$\hat{\dot{z}} = \sum_{j=1}^M \alpha_j \hat{\mu}_j. \quad (14)$$

4.4 Collaborative and Predictive Schemes

To solve the non-localizable problems, we propose the CPLA to obtain the revised location of

the mobile target. The revised location is expressed by combining the estimated locations with the predictive locations. We assume that L_e denotes the estimated location calculated by the APIT algorithm, L_p denotes the predictive location calculated by the GMM in (14) and L denotes the revised location.

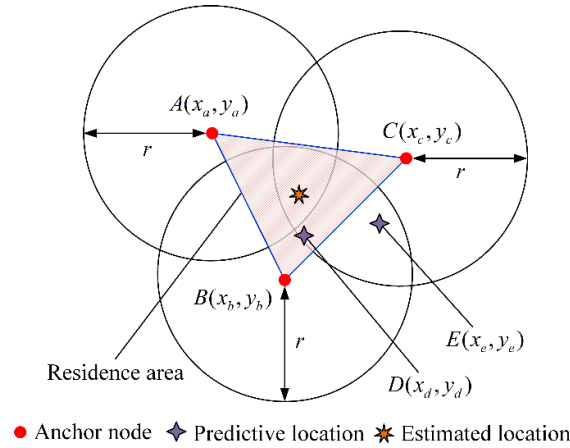


Fig. 7. Location estimation by multiple anchor nodes

4.4.1 Improved schemes for multiple anchor nodes locating situation

The distribution of WSN greatly affects localization performance, so the APIT algorithm presents low localization accuracy in the WSN with sparsely distributed sensor nodes. If the mobile target is inside the triangles formed by anchor nodes, we define the revised location by

$$L_t = w_t * L_{pt} + (1 - w_t) * L_{et} \tag{15}$$

where $w_t (0 \leq w_t \leq 1)$ is the reliability factor of L_p at time $t (t > 0)$. The reliability factor is expressed as

$$w_t = \begin{cases} w_{t-1} + \varphi, & L_{pt-1} \in R_{t-1} \\ w_{t-1} - \varphi, & L_{pt-1} \notin R_{t-1} \end{cases} \tag{16}$$

where $\varphi (\varphi > 0)$ is the regulatory factor of w and R is the residence area. In **Fig. 7**, if L_p (the location of node $D(x_d, y_d)$) is inside the residence area at time $t - 1$, we increase the reliability factor w_t in (16); if L_p (the location of node $E(x_e, y_e)$) is outside the residence area at time $t - 1$, we decrease the reliability factor w_t . Hence, the reliability factor is adaptively adjusted by the previous predictive performance.

4.4.2 Improved schemes for two-anchor nodes locating situation

The node flip ambiguity problem is very likely to happen in the WSN with sparsely distributed sensor nodes. If the mobile target is detected by only two anchor nodes, the revised location can be expressed by the predicted location and 2 intersection points. In **Fig. 8**, the node $C(x_c, y_c)$ and node $D(x_d, y_d)$ are two intersection points of the circle centered in node $A(x_a, y_a)$ and the circle centered in node $B(x_b, y_b)$ with the same radius r .

We express the revised location by

$$L = \chi_{CE} * L_{CE} + \chi_{DE} * L_{DE} \tag{17}$$

where L_{CE} and L_{DE} are two components of the revised location, χ_{CE} is the weight of L_{CE} and χ_{DE} is the weight of L_{DE} . The location component L_{CE} is expressed as

$$L_{CE} = \text{avg}\{L'_1, L'_2, \dots, L'_h\} \tag{18}$$

where $\{L'_1, L'_2, \dots, L'_h\}$ denote the $h(h > 0)$ potential locations which are randomly distributed on the line segment l_{CE} . These potential locations are calculated by the geometric theory as

$$\begin{cases} y_i = k_{CE}(x_i - x_c) + y_c \\ k_{CE} = (y_c - y_e)/(x_c - x_e) \end{cases} \tag{19}$$

where $(x_i, y_i)(i=1,2,\dots,h)$ are the potential locations on the line segment l_{CE} , x_i is randomly generated with $x_i \in (x_c, x_e)$ and k_{CE} is the slope of line segment l_{CE} . The locations of node C and node D are given by calculating the intersection points of two circles as

$$\begin{cases} (x_j - x_a)^2 + (y_j - y_a)^2 = r^2 \\ (x_j - x_b)^2 + (y_j - y_b)^2 = r^2 \end{cases} \tag{20}$$

where $(x_j, y_j)(j=1,2)$ are the locations of the intersection points and r is the communication range of the anchor nodes. Similarly, we obtain another location component L_{DE} in (17) by using (18), (19) and (20).

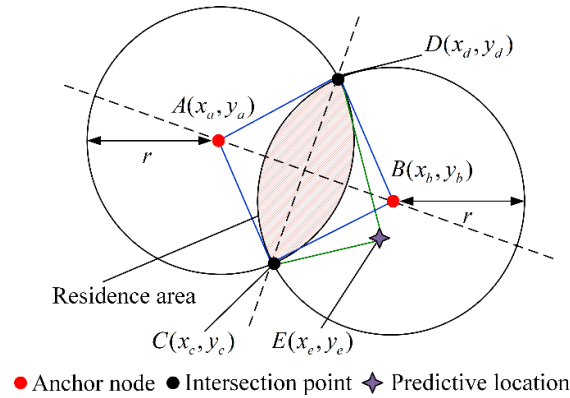


Fig. 8. Location estimation by two-anchor nodes

The weights of the location components in (17) are expressed as

$$\begin{cases} \chi_{CE} = \frac{d_{DE}}{d_{CE} + d_{DE}} \\ \chi_{DE} = \frac{d_{CE}}{d_{CE} + d_{DE}} \end{cases} \tag{21}$$

where d_{CE} is the Euclidean distance between the node C and node $E(x_e, y_e)$.

4.4.3 Improved schemes for one-anchor node locating situation

If the mobile target is detected by only one anchor node, the residence area covers most of the communication scope of the anchor node. Thus, the mobile target is impossible to be located. To solve the non-localizable problem, we propose the improved scheme to reduce the residence area and improve the localization accuracy. **Fig. 9** and **Fig. 10** show the situations

that the mobile target is detected by the anchor node $A(x_a, y_a)$. In **Fig. 9**, the predictive location (i.e., the location of node $B(x_b, y_b)$) is inside the communication scope, whereas it is outside the communication scope in **Fig. 10**. We calculate the two intersection points of the line l_{AB} and the circle centered in the anchor node A with radius r by

$$\begin{cases} (x_k - x_a)^2 + (y_k - y_a)^2 = r^2 \\ y_k = k_{AB}(x_k - x_b) + y_b \\ k_{AB} = (y_a - y_b)/(x_a - x_b) \end{cases} \quad (22)$$

where $(x_k, y_k)(k=1,2)$ denote the locations of the intersection points and k_{AB} denotes the slope of the line l_{AB} . Assume that $C(x_c, y_c)$ is the point which is closer to the predictive location between the two intersection points. Hence, the revised location is expressed by the centroid position of the residence area as

$$L = \begin{cases} \text{centroid}\{(x_a, y_a), (x_c, y_c)\}, & B \in R' \\ \text{centroid}\{(x_b, y_b), (x_c, y_c)\}, & B \notin R' \end{cases} \quad (23)$$

where R' denotes the communication scope of the anchor node A .

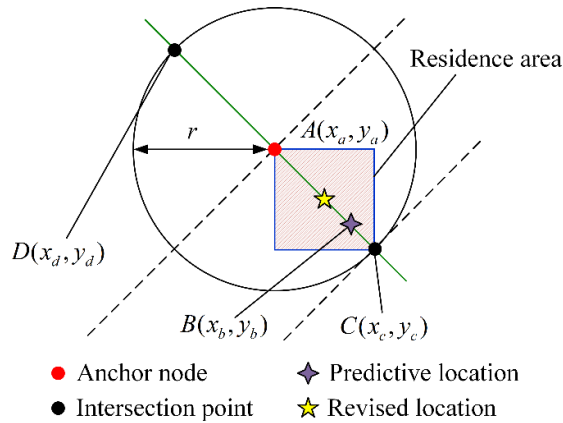


Fig. 9. Predictive location inside the communication scope of anchor node

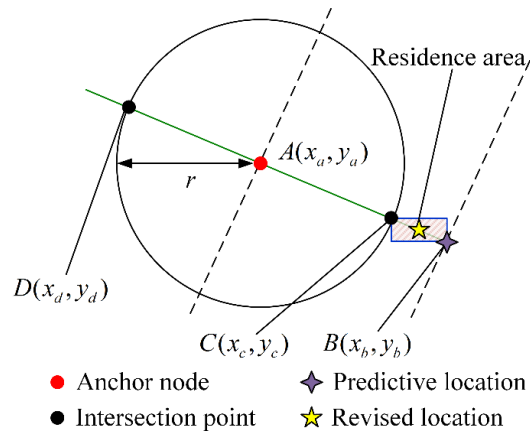


Fig. 10. Predictive location outside the communication scope of anchor node

As shown in **Fig. 9** and **Fig. 10**, if the predictive location is inside the communication scope

of the anchor node A , the revised location will be the centroid position for the rectangular residence area formed by the node A and node C . If the predictive location is outside the communication scope of the anchor node A , the revised location will be the centroid position for the rectangular residence area formed by the node B and node C .

4.4.4 Improved schemes for non-anchor node locating situation

If the sensing area is affected by the environmental interferences, the mobile target may move into the blind zone as shown in Fig. 6. In this case, the mobile target cannot be detected by any anchor node. We estimate the revised location by the predictive location calculated in (14). Therefore, the revised location is represented by

$$L = L_p. \quad (24)$$

5. Parameters Analysis and Simulation Discussion

In this section, we study the effect of various influencing parameters of the CPLA and carry out the simulation experiments to evaluate the localization performance of the tested algorithms. We first analyze the time complexity and space complexity of the CPLA. Then, the key parameters on the localization performance of the CPLA are discussed, including the Gaussian components, sizes of training dataset and test dataset, reliability factor and regulatory factor. Moreover, we demonstrate the impacts of varying the communication range and the density of anchor nodes on localization accuracy of the CPLA compared to the APIT, single Gaussian model-based (SGM-based) [27], Kalman filtering-based (KF-based) [28] and MA-MCL algorithms [11]. In this paper, we evaluate the localization performance of the APIT algorithm by calculating the localization accuracy for the trajectory points that can be located by no less than 3 anchor nodes. The single Gaussian model is applied to the trajectory prediction in the SGM-based algorithm and the Kalman filtering method is used for the trajectory prediction in the KF-based algorithm.

The simulation parameters are listed in Table 2. By default, we randomly deploy 200 anchor nodes in the $1000 \text{ m} \times 1000 \text{ m}$ area. The anchor nodes have the same properties and the communication range is 100 m. 50 historical trajectory points that have been estimated by the APIT algorithm are inputted as the training dataset, and 1 trajectory point is obtained as the test dataset by the GMR in (14) with 3 Gaussian components. The simulation experiments are carried out on the PC platform with Inter i5 CPU, 8GHz memory and Windows 10 operating system. Each experiment is randomly simulated 50 times by the MATLAB 2014 software and the plotted point is represented by the average value and the standard deviation value.

Table 2. Simulation parameters

Symbol	Description	Value
n	Number of anchor nodes	100 - 300
r	Communication range	80 m - 120 m
K	Gaussian components	2 - 8
N	Size of training dataset	10, 20, 30, 50, 100
M	Size of test dataset	1, 3, 5, 10, 20

5.1 Analysis of Algorithm Complexity

5.1.1 Time complexity

The parameters are described in Table 2. Based on the triangulation method, the time

complexity of the APIT algorithm is $O((n-2)*(n-2)*(n-2))+O(1)$, and it is reduced to $O(n^3)$. In the CPLA algorithm, the time complexity for the GMM is $O(K)*O(N)*O(M)$. The time complexity of the collaborative and predictive schemes for solving the non-localizable problems is $O(M)$. The other statements take the time complexity as $O(M)$. Because the number of anchor nodes is much larger than that of the GMM components and that of the test dataset, the time complexity of the CPLA is reduced to $O(n^3)$. That is, the time complexities of the APIT algorithm and the CPLA have the same order of magnitude.

5.1.2 Space complexity

In the CPLA, the trajectory prediction by the GMM takes up extra space. Determined by the GMM components and the size of test dataset, the space complexity of the GMM is $O(K*M)$. For other statements, the space complexity is $O(1)$. Therefore, the space complexity of the CPLA is reduced to $O(K*M)$.

5.2 Impact of the Gaussian Components on Localization Accuracy

As the Gaussian components significantly influence the fitting degree of the GMM, we explore the correlation between the number of the Gaussian distributions and the localization error of the CPLA. We define the localization error as

$$error = \frac{\sum_{i=1}^M \sqrt{(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2}}{Mr} \quad (26)$$

where (\bar{x}_i, \bar{y}_i) and (x_i, y_i) denote the real location and the revised location of the i -th trajectory point, respectively.

Fig. 11 shows the impact of the number of the Gaussian distributions on localization error. Clearly, the localization error is decreased as the number of the Gaussian distributions varies from 2 to 8. Although the localization error is negatively correlated with the Gaussian components, its decreasing trend becomes weak when the number of the Gaussian distributions keeps rising. Besides, too many Gaussian distributions lead to the extremely high computational complexity. Hence, the Gaussian components should be set to an appropriate value in view of the localization accuracy and computational complexity.

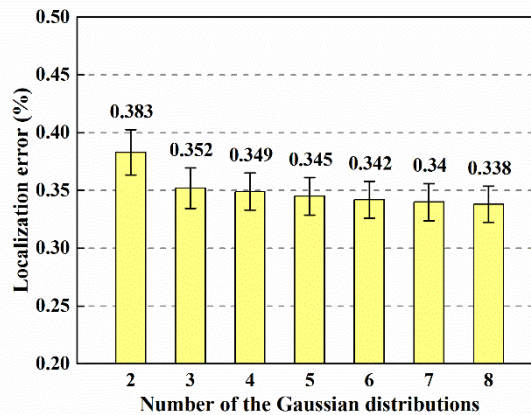


Fig. 11. Impact of the number of the Gaussian distributions on localization accuracy

5.3 Impact of the Size of Training Dataset on Localization Accuracy

Considering that the size of training dataset also affects the fitting degree for the GMM, we study the impact of varying the size of training dataset on localization accuracy of the CPLA. **Fig. 12** shows the simulation results with the size of training dataset varying from 10 to 100. We observe that the localization error is obviously decreased with the increase of the size of training dataset. Therefore, the localization accuracy of the CPLA is positively correlated with the size of training dataset.

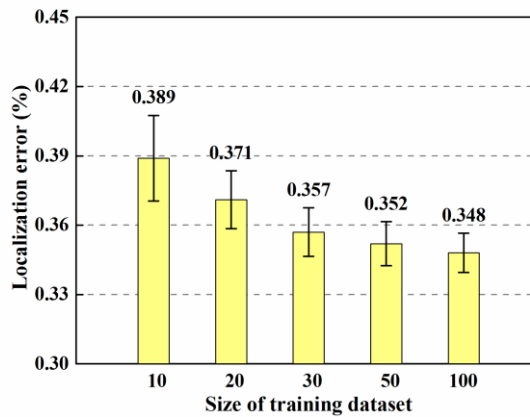


Fig. 12. Impact of the size of training dataset on localization accuracy

5.4 Impact of the Size of Test Dataset on Localization Accuracy

As plotted in (11), the test dataset can be a single trajectory point or a trajectory vector. Accordingly, the GMM is capable of obtaining the predicted trajectory points in continuous sampling time by (14). We explore the effect of the size of test dataset on localization accuracy of the CPLA. **Fig. 13** demonstrates that the increase of the size of test dataset leads to the decrease of the localization accuracy. As the size of test dataset determines the number of trajectory points predicted by the GMM each time, the predictive efficiency of the CPLA is significantly promoted by increasing the size of test dataset. Thus, we can regulate the computation performance of the CPLA by adjusting the size of test dataset.

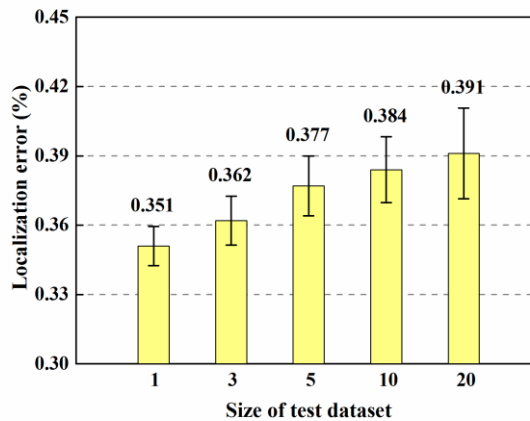


Fig. 13. Impact of the size of test dataset on localization accuracy

5.5 Impact of the Key Factors on Localization Accuracy

In (15) and (16), we learn that the reliability factor and regulatory factor have the decisive influence on the accuracy of the revised location. We discuss the impact of the initial values for the reliability factor and regulatory factor on localization error of the CPLA. The initial values of the reliability factor varies from 0.3 to 0.6, and for each reliability factor, the regulatory factor varies from 0.01 to 0.02. As shown in Fig. 14, the larger the reliability factor is, the smaller the localization error is. For each reliability factor, the localization error for the regulatory factor at 0.02 is greater than it is at 0.01.

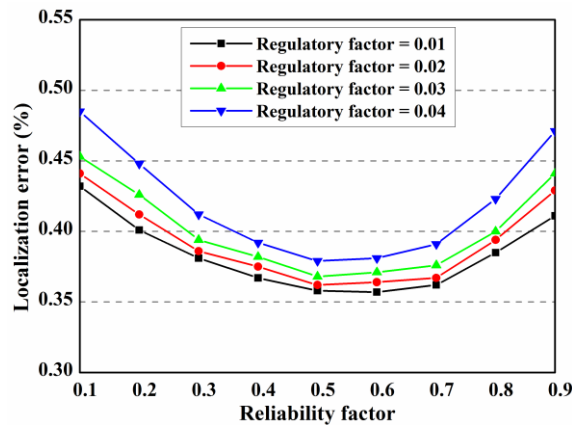


Fig. 14. Impact of the key factors on localization error

5.6 Trajectory Prediction

Fig. 15 and Fig. 16 show the network distribution and connectivity of anchor nodes. Fig. 17 presents an example of the trajectory prediction for a mobile target by the CPLA. Note that the real trajectory is plotted as the red line, the estimated trajectory by the CPLA is plotted as the blue line and the non-localizable points are expressed as the green rings. The localization error in Fig. 17 of the CPLA is 0.349 and its corresponding standard deviation is 0.017. Fig. 18 captures the localization error for 150 trajectory points from Fig. 17. The number of the iteration for the GMM in CPLA is 27.

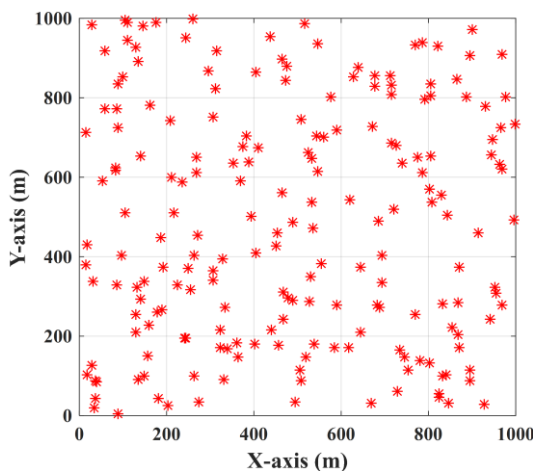


Fig. 15. The distribution of anchor nodes

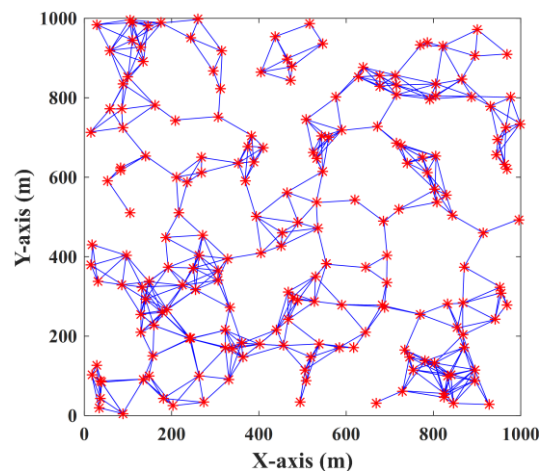


Fig. 16. The connectivity of anchor nodes

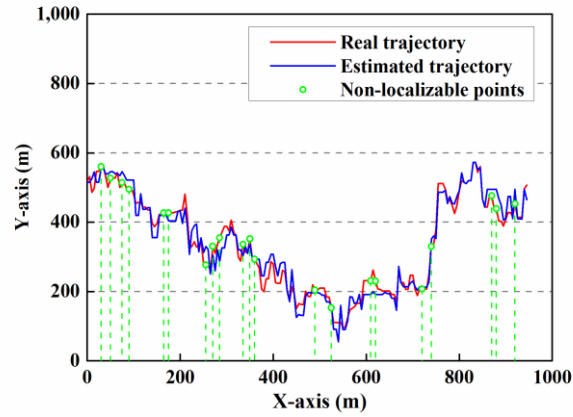


Fig. 17. Trajectory prediction for the mobile target

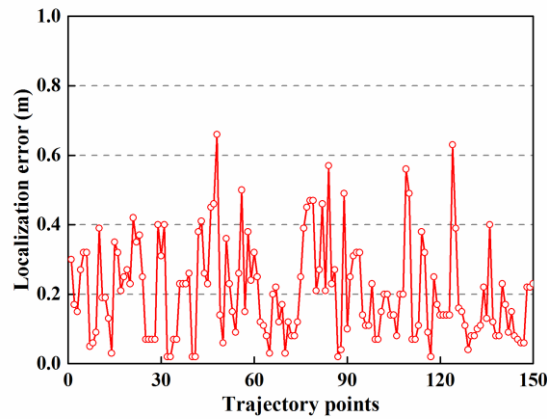


Fig. 18. Localization error for each trajectory point

The localization performance of the APIT, SGM-based, KF-based and CPLA is demonstrated in **Table 3** in the same experimental condition. We observe that the localization accuracy of the CPLA is 10.3% higher than that of the APIT algorithm, and is 3.4% higher than that of the SGM-based algorithm, and is 4.6% higher than that of the KF-based algorithm and is 2.6% higher than that of the MA-MCL algorithm. Clearly, the CPLA performs better localization performance compared to the other tested algorithms. The runtime of the tested algorithms are given in **Table 4**. The APIT algorithm takes the smallest runtime among the other tested algorithms, whereas the runtime of the KF-based algorithm is the longest among that of all tested algorithms.

Table 3. Tested algorithm performance

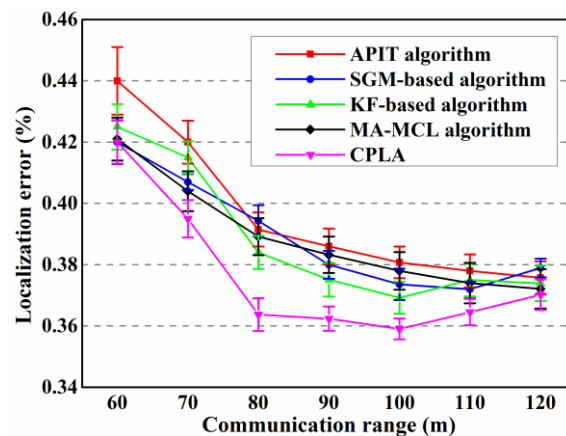
Algorithms	Localization error	Standard deviation
APIT	0.385	0.053
SGM-based	0.361	0.041
KF-based	0.365	0.029
MA-MCL	0.358	0.034
CPLA	0.349	0.017

Table 4. Runtime of the tested algorithm

Algorithms	Runtime (s)
APIT	2.28
SGM-based	2.72
KF-based	3.65
MA-MCL	3.08
CPLA	3.16

5.7 Impact of the Communication Range on Localization Accuracy

In view of the environmental interferences and the heterogeneous hardware of devices, the actual communication scopes of anchor nodes differ from the theoretical communication ranges. Therefore, we study the impact on the localization accuracy of the tested algorithms by changing the communication range. Fig. 19 depicts the simulation results with the communication range varying from 60 m to 120 m. Clearly, the localization error of the tested algorithms is decreasing when the communication range of anchor nodes changes from 60 m to 100 m. However, we see that the localization error of the SGM-based, KF-based, MA-MCL algorithms and the CPLA is increasing when the communication range continues to rise from 100 m to 120 m. The increase of the communication range improves the localization rate, but it weakens the predictive performance for the tested algorithms in non-localizable situations. Hence, the communication range has the positive correlation on the localization accuracy of the tested algorithms to some extent. In any case, the CPLA shows higher localization accuracy than other tested algorithms despite of the variations in the communication range of anchor nodes.

**Fig. 19.** Impact of the communication range on localization accuracy

5.8 Impact of the Number of Anchor Nodes on Localization Accuracy

We evaluate the localization accuracy by changing the density of anchor nodes of the tested algorithms. As Fig. 20 shows, the increase of the number of anchor nodes brings about the decrease of the localization error of the tested algorithms. Obviously, the localization error of the SGM-based, KF-based, MA-MCL and CPLA is lower than that of the APIT algorithm. That is to say, our proposed collaborative and predictive schemes take effects on solving the non-localizable problems. In addition, the CPLA presents the highest localization accuracy among the APIT, SGM-based, KF-based and MA-MCL algorithms either in the WSN with sparsely distributed sensor nodes or in the WSN with densely distributed sensor nodes.

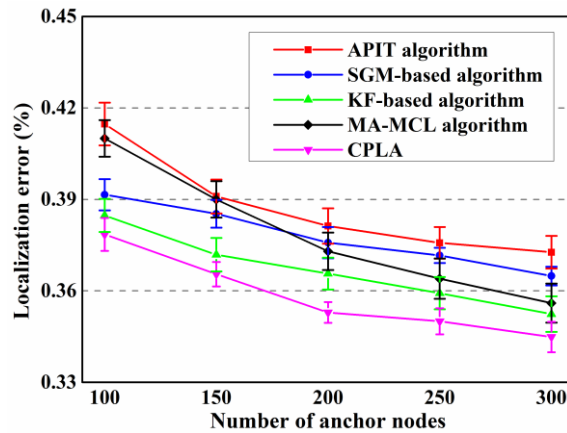


Fig. 20. Impact of the density of anchor nodes on localization accuracy

6. Conclusion

In this paper, we proposed a range-free localization algorithm for a mobile target in the WSN with sparsely distributed sensor nodes. To solve the non-localizable problems, we designed the collaborative and predictive schemes for the two-anchor nodes locating, one-anchor node locating and non-anchor node locating situations. In the CPLA, the location of a mobile target was calculated by the estimated locations in the APIT algorithm and the predictive locations in the GMM. We studied the effects of various influencing parameters that affect the localization performance, such as the Gaussian components, sizes of training dataset and test dataset, reliability factor and regulatory factor. Moreover, we explored the impact of varying the density and the communication range of anchor nodes on the localization accuracy of the CPLA compared to the APIT, SGM-based, KF-based and MA-MCL algorithms. Simulation results demonstrated that the CPLA performs higher localization accuracy than other tested predictive localization algorithms either in the WSN with sparsely distributed sensor nodes or in the WSN with densely distributed sensor nodes. So far, many interesting and challenging problems remain unsolved, including the methods from the single mobile target to the multiple mobile targets and from the two-dimensional plane to the three-dimensional space.

References

- [1] G. Han, J. Jiang, C. Zhang, "A survey on mobile anchor node assisted localization in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2220-2243, March 2016. [Article \(CrossRef Link\)](#).
- [2] G. Wang, K. Yang, "A new approach to sensor node localization using RSS measurements in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 5, pp. 1389-1395, May 2011. [Article \(CrossRef Link\)](#).
- [3] V. Tran-Quang, T. Ngo-Quynh, M. Jo, "A Lateration-localizing algorithm for energy-efficient target tracking in wireless sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 34, no. 1-4, pp. 191-220, 2016. [Article \(CrossRef Link\)](#).
- [4] T. He, C. Huang, B. M. Blum et al., "Range-free localization schemes for large scale sensor networks," in *Proc. of 9th Annual International Conference on Mobile Computing and Networking*, pp. 81-95, September 14-19, 2003. [Article \(CrossRef Link\)](#).

- [5] S. M. Hosseinirad, J. Pourdeilami, M. Niazi et al., "On improving APIT algorithm for better localization in WSN," *Journal of Ai & Data Mining*, vol. 2, no. 2, pp. 97-104, July 2013. [Article \(CrossRef Link\)](#).
- [6] Y. Zhang, Y. Lou, Y. Hong et al., "Distributed projection-based algorithms for source localization in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3131-3142, June 2015. [Article \(CrossRef Link\)](#).
- [7] M. Singh, P. M. Khilar, "An analytical geometric range free localization scheme based on mobile beacon points in wireless sensor network," *Wireless Networks*, pp. 1-14, November 2015. [Article \(CrossRef Link\)](#).
- [8] M. Singh, P. M. Khilar, "Mobile beacon based range free localization method for wireless sensor networks," *Wireless Networks*, pp. 1-16, February 2016. [Article \(CrossRef Link\)](#).
- [9] N. Lasla, M. F. Younis, A. Ouadjaout et al., "An effective area-based localization algorithm for wireless networks," *IEEE Transactions on Computers*, vol. 64, no. 8, pp. 2013-2118, August 2015. [Article \(CrossRef Link\)](#).
- [10] A. Baggio, K. Langendoen, "Monte Carlo localization for mobile wireless sensor networks," *Ad Hoc Networks*, vol. 6, no. 5, pp. 718-733, January 2008. [Article \(CrossRef Link\)](#).
- [11] G. Teng, K. Zheng, W. Dong, "MA-MCL: mobile-assisted Monte Carlo localization for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2011, no. 4, pp. 235-245, June 2011. [Article \(CrossRef Link\)](#).
- [12] B. M. Vasim, A. V. Ramprasad, "Discrete antithetic Markov Monte Carlo based power mapping localization algorithm for WSN," in *Proc. of IEEE International Conference on Advanced Communication Control and Computing Technologies*, pp. 56-62, August 23-25, 2012. [Article \(CrossRef Link\)](#).
- [13] J. F. Huang, G. Y. Chang, G. H. Chen, "A historical-beacon-aided localization algorithm for mobile sensor networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 6, pp. 1109-1122, August 2014. [Article \(CrossRef Link\)](#).
- [14] Q. Niu, T. Huan, P. Chen, "NMCT: a novel Monte Carlo-based tracking algorithm using potential proximity information," *International Journal of Distributed Sensor Networks*, vol. 3, no. 51, pp. 1-10, February 2016. [Article \(CrossRef Link\)](#).
- [15] Y. Shang, W. Ruml, Y. Zhang et al., "Localization from mere connectivity," in *Proc. of Acm International Symposium on Mobile Ad Hoc Networking & Computing*, pp. 201-212, June 1-3, 2003. [Article \(CrossRef Link\)](#).
- [16] M. Shona, M. Job, H. Choo, "An interactive cluster-based MDS localization scheme for multimedia information in wireless sensor networks," *Computer Communications*, vol. 35, no. 15, pp. 1921-1929, September 2012. [Article \(CrossRef Link\)](#).
- [17] H. W. Wei, Q. Wan, Z. X. Chen et al., "A novel weighted multidimensional scaling analysis for time-of-arrival-based mobile location," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3018-3022, July 2008. [Article \(CrossRef Link\)](#).
- [18] A. Amar, Y. Wang, G. Leus, "Extending the classical multidimensional scaling algorithm given partial pairwise distance measurements," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 473-476, June 2010. [Article \(CrossRef Link\)](#).
- [19] D. Ma, M. J. Er, B. Wang, H. B. Lim, "Range-free wireless sensor networks localization based on hop-count quantization," *Telecommunication Systems*, vol. 50, no. 3, pp. 199-213, July 2012. [Article \(CrossRef Link\)](#).
- [20] S. Éfren L, R. W. Pazzi, E. F. Nakamura, "A prediction-based clustering algorithm for tracking targets in quantized areas for wireless sensor networks," *Wireless Networks*, vol. 21, no. 7, pp. 1-16, October 2015. [Article \(CrossRef Link\)](#).
- [21] L. Shang, K. Zhao, Z. Cai et al., "An energy-efficient collaborative target tracking framework in distributed wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2014, no. 2, pp. 1-18, January 2014. [Article \(CrossRef Link\)](#).
- [22] K. W. Lin, M. H. Hsieh, V. S. Tseng, "A novel prediction-based strategy for object tracking in sensor networks by mining seamless temporal movement patterns," *Expert Systems with Applications*, vol. 37, no. 4, pp. 2799-2807, April 2014. [Article \(CrossRef Link\)](#).

- [23] W. Liu, E. Dong, Y. Song, "Robustness analysis for node multilateration localization in wireless sensor networks," *Wireless Networks*, vol. 21, no. 5, pp.1473-1483, July 2015.
[Article \(CrossRef Link\)](#).
- [24] B. Li, C. Wei, B. Wang, "A robust wireless sensor network localization algorithm in mixed LOS/NLOS scenario," *Sensors*, vol. 15, no. 9, pp. 23536-23553, September 2015.
[Article \(CrossRef Link\)](#).
- [25] B. N. Vo, W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091-4104, December 2010.
[Article \(CrossRef Link\)](#).
- [26] D. Gu, "Distributed EM algorithm for Gaussian mixtures in sensor networks," *IEEE Transactions on Neural Networks*, vol. 19, no. 7, pp. 1154-1166, August 2008. [Article \(CrossRef Link\)](#).
- [27] C. Taeryon, Q. S. Jian, W. Bo, "A Gaussian process regression approach to a single-index model," *Journal of Nonparametric Statistics*, vol. 23, no. 1, pp. 21-36, March 2011.
[Article \(CrossRef Link\)](#).
- [28] Z. Chen, H. Zou, H. Jiang et al., "Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715-732, January 2015.
[Article \(CrossRef Link\)](#).



Yuan Liu is currently pursuing the Doctoral degree with School of Instrument Science and Engineering, Southeast University, China. His research interests include wireless sensor networks localization and intelligent information processing.



Junjie Chen is currently a Professor and Ph.D. Supervisor with Southeast University. His research interests include theory and applications of wireless sensor networks, technology and applications for system integrations of Internet of Things, theory and method of ubiquitous computing, intelligent sensing and control technology, control technology and applications of protected agriculture and intelligent agricultural machinery equipment.