WILEY | Hindawi

Research Article

Security-Oriented Indoor Robots Tracking: An Object Recognition Viewpoint

Yaoqi Yang [ID],[1] **Xianglin Wei** [ID],[2] **Renhui Xu,**[1] **Laixian Peng** [ID],[1] **Yunliang Liao,**[1] **and Lin Ge**[1]

[1]*College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China*
[2]*63rd Research Institute, National University of Defense Technology, Nanjing 210007, China*

Correspondence should be addressed to Laixian Peng; lxpeng@hotmail.com

Indoor robots, in particular AI-enhanced robots, are enabling a wide range of beneficial applications. However, great cyber or physical damages could be resulted if the robots' vulnerabilities are exploited for malicious purposes. Therefore, a continuous active tracking of multiple robots' positions is necessary. From the perspective of wireless communication, indoor robots are treated as radio sources. Existing radio tracking methods are sensitive to indoor multipath effects and error-prone with great cost. In this backdrop, this paper presents an indoor radio sources tracking algorithm. Firstly, an RSSI (received signal strength indicator) map is constructed based on the interpolation theory. Secondly, a YOLO v3 (You Only Look Once Version 3) detector is applied on the map to identify and locate multiple radio sources. Combining a source's locations at different times, we can reconstruct its moving path and track its movement. Experimental results have shown that in the typical parameter settings, our algorithm's average positioning error is lower than 0.39 m, and the average identification precision is larger than 93.18% in case of 6 radio sources.

1. Introduction

1.1. Motivation and Background. Indoor robots are becoming increasingly popular in the market in view of their beneficial applications, ranging from navigating and sweeping to healthy-caring. In combination with artificial intelligence (AI), they are anticipated to drastically change people's daily lives. Although these with advantages, indoor robots still face severe cyber and physical threats due to their hardware and software vulnerabilities [1]. In extreme cases, even the simplest cleaning robots could be manipulated to launch aggressive physical attacks, such as assassinating revealed in [2] and eavesdropping on private conversations in [3–9]. Therefore, it is essential to track the nonstationary robots for security besides traditional efficiency concerns.

By literature review, we identify three types of alternative techniques that could be utilized for indoor robots' tracking. In the first place, a robot could be tracked using indoor navigation technique, where multiple anchors that transmit signals are needed for the robot to derive its own position based on its received signals. Note that to perform the tracking task, we need the robot to report its position [10]. This process could be easily spoofed and spoiled by selfish or malicious robots. In the second place, based on SLAM- (simultaneous localization and mapping-) based technique, a robot may obtain and report its position through diverse sensors [11]. Obviously, this is only feasible for the 'honest' robots. In the third place, image recognition or video recognition could be adopted to identify and distinguish different robots if cameras could be installed on the ceiling [12, 13]. However, this alternative is very sensitive to light condition and obstacles. In summary, existing efforts could not fulfill the requirements for malfunctioned robots tracking task due to two challenges. On one hand, they usually require the cooperation from hijacked or malicious robots. On the other hand, they are sensitive to indoor multipath effects (as for radio-based techniques) or illumination conditions (as for image-based solutions).

1.2. Related Work. For ease of presentation, the state of art efforts in indoor localization is shown in Table 1. As shown in Table 1, the efforts in the indoor localization are classified by their method, environment settings, and the one-time object identification consideration. It can be seen that most methods do not take the dynamic environment into the environment, while some traits in the dynamic environment would change, such as wireless channel. Besides, Table 1 is self-explainable, and no existing efforts consider the identification issue for the localized objects in one time. In order to make up this research gap, a one-time indoor object localization and identification scheme is proposed in this paper.

1.3. Main Work and Contribution. In this backdrop, this paper aims to develop an active but nonintrusive indoor robots tracking algorithm. Our tracking process contains two steps, i.e., identification step and localization step, and each indoor robot is treated as a wireless radio source. In the identification step, a YOLO v3 network [16] is deployed to identify and distinguish different robots' signals based on the constructed RSSI maps. Then, each robot's position is derived based on the identification results in the localization step. Our main contributions are threefold:

(1) An RSSI map construction scheme is proposed in the identification step, which has a satisfactory localization resolution with low cost by deploying a small number of monitors.

(2) After establishing RSSI maps, a network based on YOLO v3 is trained and applied for multiple robots recognition; then, the bounding boxes information of YOLO v3 [16] is further utilized on the recognition results to refine the localization accuracy. To our best knowledge, this is the first paper that treats indoor robot tracking from the viewpoint of object recognition.

(3) A series of experiments are conducted on a collected dataset to evaluate effectiveness of our algorithm. Results have shown that in typical indoor environments, our proposal's positioning error is less than 0.39 m with a recognition precision higher than 93.18% in case of 6 radio sources.

The remaining part of this paper is organized as follows. Section 2 introduces the system model. In Section 3, identification and localization algorithm for indoor radio sources is detailed. Evaluation experiments are conducted to verify the correctness and effectiveness of the proposed algorithm in Section 4. Finally, a brief conclusion is provided in Section 5.

2. System Model

The scenario we considered here is that several indoor robots work on a floor, named the task area, of a building; and there are several classrooms, meeting rooms, and corridors on this floor. To realize efficient control and navigation, a robot usually has a communication module and could be treated as an indoor radio source. Therefore, we could track different robots if we can recognize and distinguish their equipped radios. For the ease of description, each radio source refers to a robot in the following analysis.

2.1. Task Area Model. A Cartesian coordinate system is established for the task area to assist radio source localization or tracking, in which an origin, two perpendicular axes, and their positive directions are determined based on the floor plane. In our setting, the origin locates at the northeast corner of the floor, while east and north directions are the positive directions of the horizontal and vertical axes, respectively. The set of to-be-tracked radio sources, i.e., indoor robots, is $\mathcal{S} = \{S_1, S_2, \ldots, S_N\}$, where S_i denotes the i-th radio source, and N is the number of radio sources. All radio sources' positions at time t are $\mathcal{L} = \{(x_1(t), y_1(t)), (x_2(t), y_2(t)), \ldots, (x_N(t), y_N(t))\}$, where $x_i(t)$ and $y_i(t)$ are the i-th radio source's abscissa and ordinate, respectively.

A number of monitors are deployed for tracking purpose through collecting the RSSI values. To determine their respective positions, the task area is divided into $(m-1) \times (n-1)$ rectangular areas with the same size, and the monitor is placed at each vertex of all rectangles. In this way, a total of $m \times n$ monitors are placed. Denote the set of monitors as $\mathcal{D} = \{D_{1,1}, D_{1,2}, \ldots, D_{m,n}\}$, and $D_{i,j}$ refers to the monitor deployed in the i-th column and the j-th row. Each monitor periodically collects its received RSSI value. $D_{i,j}$'s monitored RSSI value at time t is denoted as $Z(t)_{i,j}$. Then, all monitors' collected RSSI values at time t (i.e., 1 of Figure 1(a)) could be recorded as

$$\mathfrak{R}_1(t) = [X|Y|Z(t)], \tag{1}$$

where X, Y, and $Z(t)$ are all matrices with m rows and n columns, $X_{i,j}$ and $Y_{i,j}$ are $D_{i,j}$'s abscissa and ordinate, respectively, and $Z(t)_{i,j}$ represents the maximum RSSI value from the collected N RSSI values by $D_{i,j}$ at time t, $1 \le i \le m$ and $1 \le j \le n$.

2.2. Radio Propagation Model. It is a difficult and time-consuming task to build the radio propagation model for the task area due to the challenges brought by multipath effects. Moreover, the change of the indoor environment may easily bias the established empirical model [14]. To solve this tricky problem, based on the collected RSSI values at the monitors without building any analytical model, an RSSI map is built to extract the radiation trait, which is the spatial distribution of the power within the environment [13] (i.e., the image selected by the bounding box in Figure 1(c)). To be specific, the collected raw dataset $\mathfrak{R}_1(t)$ could be transformed to an RSSI map, in which an RSSI value is represented by a rectangular shape, as 1–2 in Figure 1(a). Let the set of extracted radiation traits be $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$, where P_i is the i-th radio source's radiation trait.

2.3. Tracking Model. To track multiple radio sources or robots, we have to continuously derive their respective positions at different time snaps. Therefore, in each time

TABLE 1: A comparison of efforts of indoor localization.

| Reference	Method	Environment	One-time identification
Liang et al. [12]	Multifingerprint	Time-varying	No consideration
Rahman et al. [13]	Pedestrian dead reckoning	Static	No consideration
Yu et al. [14]	Maximum likelihood estimation	Static	No consideration
Poulose and Han [15]	Bayesian decision theory	Static	No consideration
Zhu et al. [16]	Augmented reality	Static	No consideration
Fazelinia et al. [17]	Fingerprinting with AM signal	Static	No consideration
This work	YOLO v3 approach	Time-varying	Yes
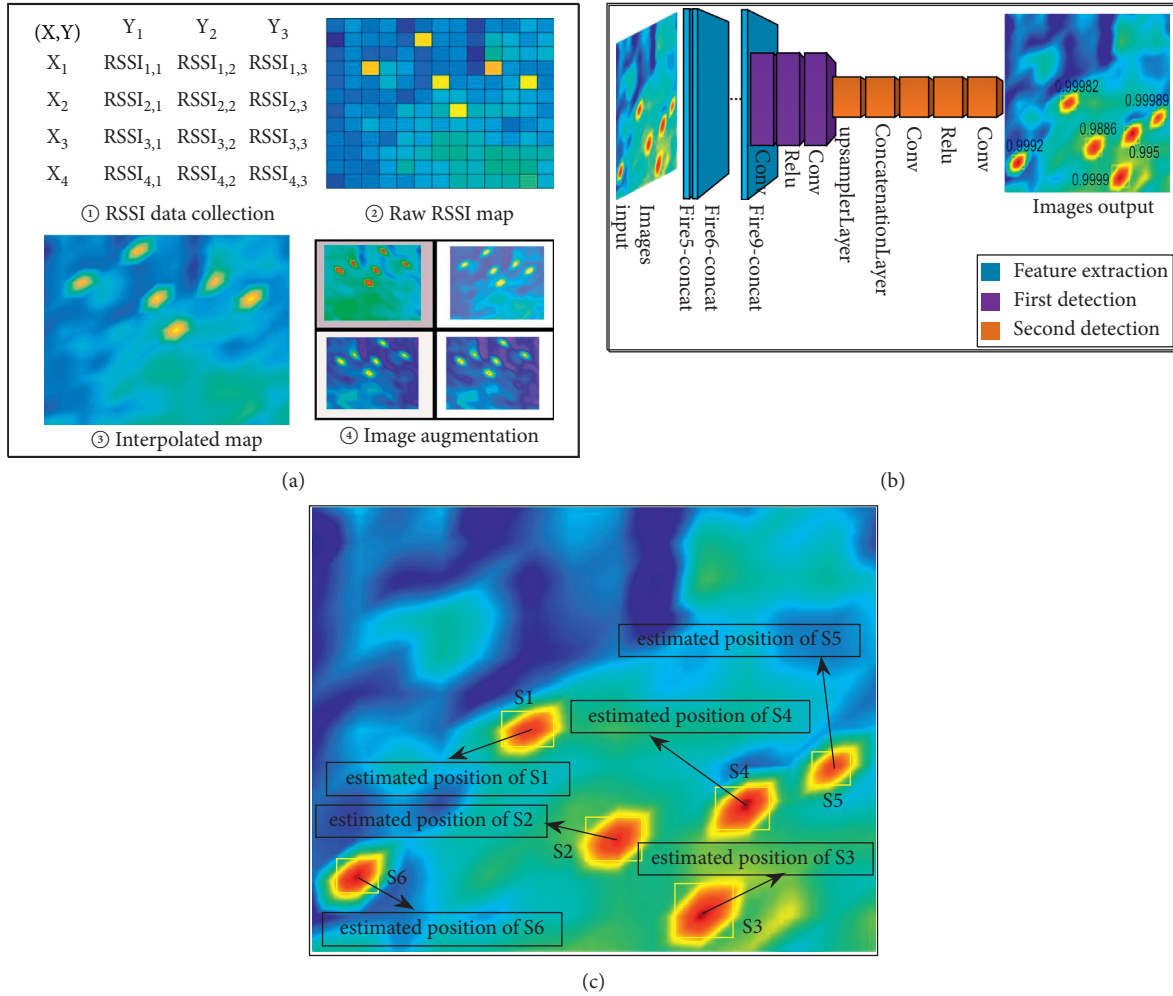
FIGURE 1: The flowchart of the YOLO v3-based radio source tracking algorithm. (a) Images preparation. (b) Radio sources recognition. (c) Radio sources localization.

snap t, we have to firstly recognize different radio sources based on the established RSSI map and $\mathscr{P}$ and then derive each radio source's position.

For the recognition purpose, we need to derive a bounding box for each radio source through treating the RSSI map as an image. Then, N bounding boxes could be obtained for N radio sources; the set of bounding boxes is denoted as $\varepsilon = \{E_1, E_2, \ldots, E_N\}$, where E_i represents the bounding box for the i-th radio source and contains four elements: the abscissa and ordinates of the bounding box's top left vertex and the length and the width of the bounding box in order.

The next move is to get all recognized radio sources' positions based on ε in combination with the coordinate system embedded in the RSSI map and thus its derived image. Let $\Lambda = \{(\hat{x}_1(t), \hat{y}_1(t)), (\hat{x}_2(t), \hat{y}_2(t)), \ldots, (\hat{x}_N(t), \hat{y}_N(t))\}$ be the set of N radio sources' estimated positions, where $\hat{x}_i(t)$ and $\hat{y}_i(t)$ are the abscissa and ordinate of the i-th radio source at time t.

2.4. Problem Formulation. To facilitate the problem formulation, we define $\Omega_{\mathfrak{R}_1(t)}$ as the identification function which can derive $\mathscr{S}$ based on $\mathfrak{R}_1(t)$; $\Psi_{\mathfrak{R}_1(t)}$ as the bounding function to derive ε based on $\mathfrak{R}_1(t)$; and Θ_ε as the localization function to get the estimated positions Λ based on ε:

$$\Omega_{\mathfrak{R}_1(t)}: \left\{ \left(\mathfrak{R}_1(t) \right) \longrightarrow (\mathscr{S}) \right\}, \tag{2}$$

$$\Psi_{\mathfrak{R}_1(t)}: \left\{ \left(\mathfrak{R}_1(t) \right) \longrightarrow (\varepsilon) \right\}, \tag{3}$$

$$\Theta_\varepsilon: \{ \varepsilon \longrightarrow \Lambda \}. \tag{4}$$

When N radio sources exist in the task area, and M radios are identified successfully from the N radio sources at time t, this paper aims to jointly minimize the identification error in the identification stage (stage 1) and the positioning error in the localization stage (stage 2):

$$\begin{cases} \min |M - N| & \text{(stage 1)} \\ \min \left\{ \sum_{j=1}^{M} \left[\left(x_j(t) - \hat{x}_j(t) \right)^2 + \left(y_j(t) - \hat{y}_j(t) \right)^2 \right] \right\} & \text{(stage 2)} \end{cases} \tag{5}$$

3. YOLO v3-Based Radio Source Tracking

3.1. Basic Idea. After obtaining the interpolated RSSI map and its derived image at time t, we have to first identify each radio source and determine its bounding box. From the perspective of object recognition in an image, the recognition task could be completed if we can capture the characteristics of each radio source's radiation trait, which represents a radio source's radiation range and intensity in an RSSI map [17]. The process of capturing a radio source's radiation trait equals to deriving $\Omega_{\mathfrak{R}_1(t)}$ and $\Psi_{\mathfrak{R}_1(t)}$. However, there are many ways to solve (2) and (3). For example, the empirical model for indoor object detection and localization based on RSSI is widely utilized because of its simplicity and low-cost. But its positioning accuracy is unsatisfactory since the measured RSSI values are sensitive to the indoor multipath effects. Moreover, it is impractical to distinguish different radio sources based solely on RSSI data.

In this backdrop, this paper adopts deep neural networks for tackling image segmentation, i.e., radio source recognition, in view of their great success in image recognition areas. To be specific, YOLO v3 is adopted for this purpose due to its unique capability to capture the different radio sources' radiation traits and their differences [16]. We could identify each radio source and obtain its bounding box at the same time after conducting YOLO v3 on the image of an RSSI map. In other words, YOLO v3 is adopted to solve both (2) and (3) simultaneously. Compared with empirical model-based solutions, deep neural networks-based methods are robust to indoor multipath effects.

To tackle the localization problem shown in (4), a straightforward idea is treating the center of each radio source's bounding box as its position. However, this will introduce extra localization errors if the bounding box derived by YOLO is biased. Taking this into consideration, the position of the pixel point with the largest RSSI value within the i-th radio source's bounding box E_i is chosen as the radio source's location. If there are multiple pixel points with the same maximum RSSI value, the center position is adopted as the localization result.

3.2. Indoor Radio Sources Identification and Localization. As shown in Figure 1, our indoor radio sources tracking method contains three steps: images preparation, radio sources recognition, and radio sources localization. In the first step, raw RSSI values at the monitors are collected to build RSSI maps, which will further be transformed into images using the interpolation theory. Then, a YOLO v3 detector is trained offline on the images for identifying and distinguishing different radio sources, and the trained YOLO v3 network is utilized for online radio sources' recognition. Finally, all radio sources' positions are determined based on their respective bounding boxes and the largest RSSI values of the pixels within the boxes.

3.2.1. Images Preparation. Adapting $\mathfrak{R}_1(t)$ directly to construct the RSSI map will result in two defects. On one hand, the positioning granularity of the RSSI map is determined by the monitors' deployment density. Sparsely deployed monitors will lead to low tracking accuracy while high-density deployment would introduce high deployment cost. On the other hand, biased or even error monitored data are common due to the impacts of malfunctioned monitors or indoor multipath effect. Therefore, to achieve low-cost monitoring while promoting the tracking accuracy, the 2-th Bernstein Bezier interpolation theory [18] is utilized to refine the raw RSSI map, and $\mathfrak{R}_1(t)$ is expanded to $\mathfrak{R}_2(t)$.

As shown in Figure 2, $D_{i,j}$, $D_{i,j+1}$, $D_{i+1,j}$, and $D_{i+1,j+1}$ are four deployed monitors, and their positions are $(x_{i,j}(t), y_{i,j}(t))$, $(x_{i,j+1}(t), y_{i,j+1}(t))$, $(x_{i+1,j}(t), y_{i+1,j}(t))$, and $(x_{i+1,j+1}(t), y_{i+1,j+1}(t))$, respectively. I_1, and I_2 are the center of gravity of the triangle $\triangle D_{i,j} D_{i,j+1} D_{i+1,j}$ and triangle $\triangle D_{i+1,j} I_1 D_{i+1,j+1}$, and their positions are $(x_{I_1}(t), y_{I_1}(t))$ and $(x_{I_2}(t), y_{I_2}(t))$. The purpose of applying 2-th Bernstein Bezier interpolation theory [14] is to derive the RSSI values at I_1 and I_2 without deploying extra monitors.

Next, according to the 2-th Bernstein Bezier polynomial theory, we have the following interpolation formula:

$$z_{I_h} = \sum_{\substack{0 \leq f,g \leq 2 \\ f+g=2}} \frac{2!}{f!g!} b_{i,j} x_{I_h}{}^f y_{I_h}{}^g \ (h = \{1, 2\}). \tag{6}$$

Here, z_{I_h} is the RSSI value of the interpolation point; x_{I_h} and y_{I_h} are the interpolated point's abscissa and ordinate; and $b_{i,j}$ is called the Bezier ordinates of z_{I_h} [14]. Let the dataset in the task area after applying the 2-th Bernstein Bezier interpolation be $\mathfrak{R}_2(t)$:

$$\mathfrak{R}_2(t) = \left[X' | Y' | Z'(t) \right]_{p \times 3q}, \tag{7}$$

where X', Y', and $Z'(t)$ are all matrices with p rows and q columns, and they contain the abscissas, ordinates, and RSSI values of the points after interpolation, respectively. $b_{i,j}$ is called the Bezier ordinates of z_{I_h} and needs to be determined.

Then, $\mathfrak{R}_2(t)$ is adopted to establish the refined RSSI map (i.e., the interpolated RSSI map), as 1–3 of Figure 1(a) shows. Finally, image augmentation method presented in reference [19] is applied to expand the number of images in the training dataset, as step 4 in Figure 1(a) shows.

3.2.2. Radio Sources Recognition. As shown in the radio sources recognition step in Figure 1, a YOLO v3 network contains three parts: feature extraction network, first detection head, and second detection head, in which 70 net layers consist the neutral networking, and 78 connection tables and 58 learnable tables are used to connect different layers [16].

In the offline training stage, a number of images obtained in the images preparation step are adopted as the training dataset, and each image is labelled with all the radio sources' class and bounding box information. Then, the outputs of the network are the images where the radio sources are recognized with their respective bounding boxes. To refine the weights in the network, back propagation method is adopted and cross entropy is the loss function [16]. The learning rate, the number of training epochs, the number of warm-up periods, and the regularization are set to be 0.001, 3500, 1000, and 0.0005, respectively. In the online recognition stage, the penalty, the confidence, and the overlap thresholds are all set to be 0.5.

3.2.3. Radio Sources Localization. To tackle the potential error introduced by biased bounding box derived by YOLO v3 network, the position of the pixel with the largest RSSI values within a recognized radio source's bounding box is chosen as the source's location, as shown in the radio source localization step shown in Figure 1. Therefore, for an identified radio source, (4) can be converted to

$$(\hat{x}_i(t), \hat{y}_i(t)) = \frac{1}{w} \left(\sum_{k=1}^{w} x_k(t) \sum_{k=1}^{w} y_k(t) \right) \quad w = 1, 2, \cdots, \tag{8}$$

where $(x_k(t), y_k(t)) \in E_k$, $z_k(t) = \max\{Z(t)\}$, and w is the number of the selected pixels in the i-th bounding box.

3.2.4. Algorithm Description. The indoor radio sources tracking algorithm is illustrated in Algorithm 1. Step 1 determines the dataset to construct the RSSI map after interpolation. In steps 2 and 3, the YOLO v3 network is trained, next deriving the image to be recognized in step 5, and the trained YOLO v3 detector is utilized to obtain the identification and localization results in steps 6–9.

4. Simulation Results

4.1. Experimental Settings. All experiments are conducted in Room 701, Communication Hall, Army Engineering University. The floor plan of the room is shown in Figure 3; the size of room is $11.2 \, \text{m} \times 10.4 \, \text{m}$; and the vertical and horizontal distances between two neighboring monitors are both $0.8 \, \text{m}$ [17]. The size of established RSSI map is 700×525 pixels and the output image's size is normalized as 227×227 pixels to accelerate the training process. In addition, the software for collecting RSSI data is WiFi NetSpot. The tracking period T lasts for 30 minutes, and it is divided into a number of time intervals with each interval t_0 being 30 seconds. 6 radio sources are investigated, i.e., MECHERVO (S1), Huawei MatePad (S2), Thinkpad T580 (S3), Xiaomi mix2 (S4), HUAWEI P40 (S5), and Thinkpad T480 (S6). These portable devices are carried by Turtlebot Robots. We open all the collected raw RSSI values to the research community (https://github.com/tracking-data/tracking-project/releases/tag/v1.0).

4.2. Performance Metrics

4.2.1. Average Identification Precision. Assume a times of experiments are conducted in total. In the j-th experiment, the i-th radio source emerges k_3 times, and it is detected k_1 times while being recognized correctly by Algorithm 1 for k_2 times. Then, the identification rate of the i-th radio source in the j-th experiment will be $V_i(j) = k_2/k_1$ and the recall rate $R_i(j) = k_2/k_3$ is derived to measure the false alarm performance. Then, after a times of experiments, the average identification precision V_i^{ave} and the average recall rate R_i^{ave} of the i-th radio source will be

$$V_i^{\text{ave}} = \frac{1}{a} \sum_{j=1}^{a} V_i(j),$$

$$R_i^{\text{ave}} = \frac{1}{a} \sum_{j=1}^{a} R_i(j). \tag{9}$$

4.2.2. Average Positioning Error. For the i-th identified radio source, its location $(\hat{x}_i(j), \hat{y}_i(j))$ can be estimated by Algorithm 1 in the j-th experiment, and its real position $(x_i(j), y_i(j))$ is known in advance. With time period $[0, T]$ $(T = n \times t_0)$, denote the average positioning error for the i-th radio source defined as

$$\mu_i = \frac{1}{n} \sum_{j=0}^{n} \left[(\hat{x}_i(j) - x_i(j))^2 + (\hat{y}_i(j) - y_i(j))^2 \right]. \tag{10}$$

4.3. Results and Analysis

4.3.1. Radio Sources Recognition Results. A straightforward presentation of the radio source recognition results in one single timeslot is shown in the images output layer in Figure 1(b), where 6 radio sources (ranging from S1 to S6)
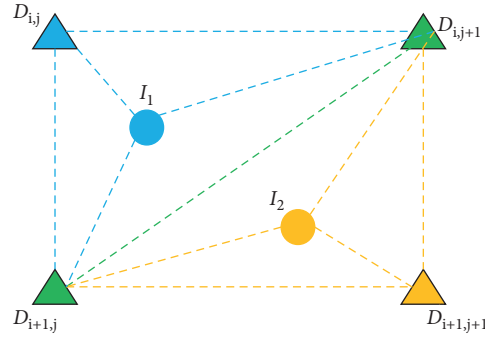
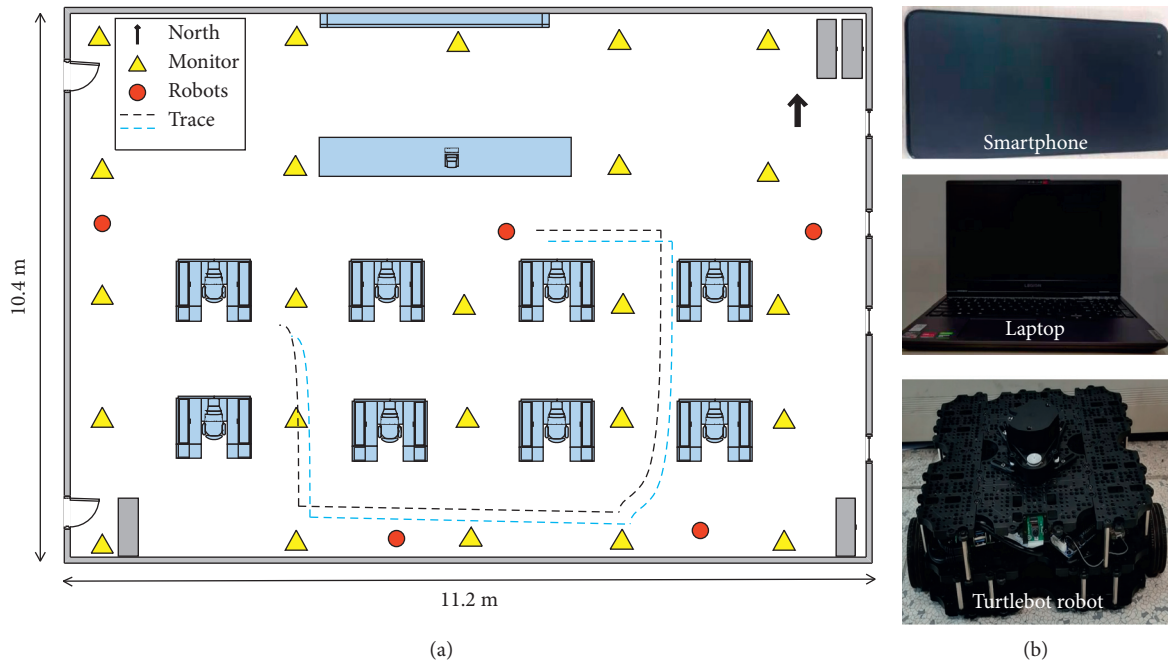FIGURE 2: The principle of the 2-th Bernstein Bezier interpolation.

(a)

(b)

FIGURE 3: The experimental environment. (a) Floor plan. (b) Robot.

are identified with 0.9986, 0.9999, 0.995, 0.9999, 0.9998, and 0.9999 confidence scores, respectively.

Figure 4 shows the 6 radios' identification precision and localization errors. As shown in Figure 4(a), the average identification precisions of the 6 radios are 0.9648, 0.9562, 0.9469, 0.9318, 0.9404, and 0.9440, respectively. The differences between different radios' identification precision lie in the fact that different radio sources have different radiation traits. Generally speaking, the more obvious, i.e., the larger the transmitting power, a radio's radiation trait, the higher its identification precision; and the radiation traits are subject to lots of factors such as the transmitting power, the usage degree, and the position. Figure 4(b) presents the CDF of 6 radios' average positioning error. As shown in Figure 4(b), each radio's average positioning error is less than 0.39 m with a probability higher than 90%.

Figure 5 shows the 6 radio sources' real and estimated traces in 60 timeslots. As can be seen, the difference between

the real and the estimated traces is less than 0.4 m, which is better than 1.54 m with fingerprinting approach in [20]. Besides, the higher the identification precision of a radio source is, the less its positioning error will be. A video is made based on the tracking results and is made publicly available (https://github.com/tracking-data/tracking-project/releases/tag/v1.0).

4.3.2. Generality of the Trained Network. To evaluate the generality of the trained network, we have tested the trained network in Room 705 and Room 725 located in the same building with different number of robots from $S1$ to $S4$, and the testing results are shown in Table 2. From Table 2, we know that the trained network can still achieve high recognition (>90%) and localization (≤0.65m) accuracy in different rooms. Moreover, it is validated that our algorithm still works independent of the number of emerged radios.
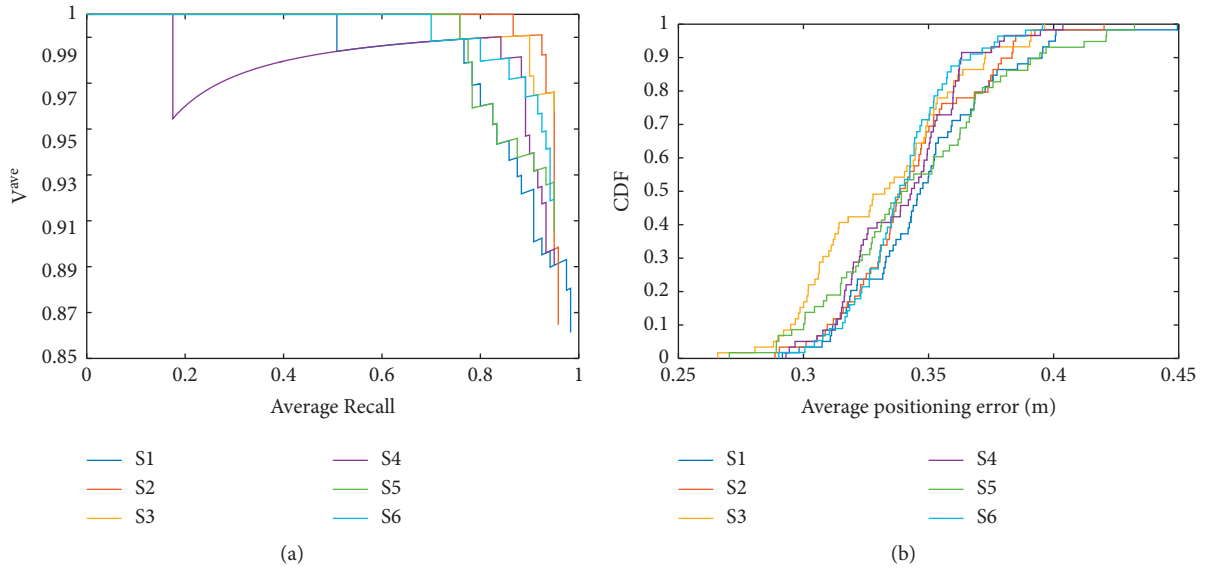
(a)

(b)

Figure 4: The tracking performance of Algorithm 1. (a) Identification performance. (b) Localization performance.
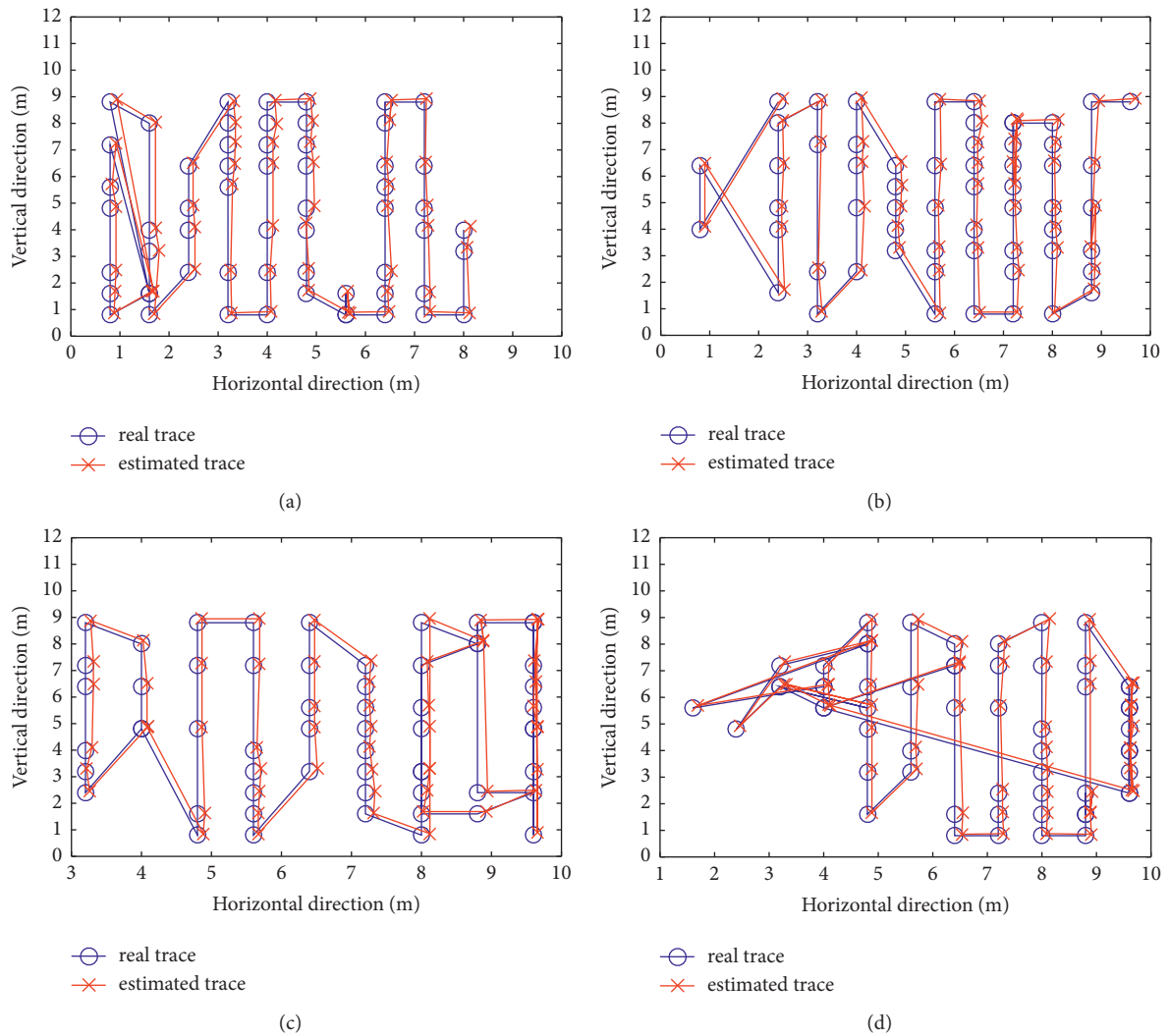

(a)
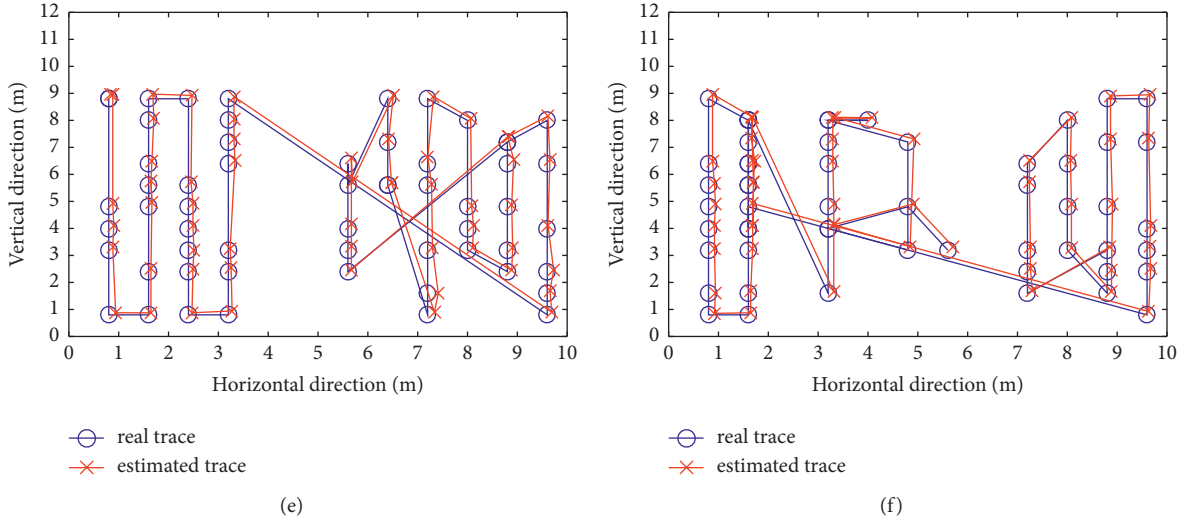
(b)

(c)

(d)

Figure 5: Continued.

(e)

(f)

FIGURE 5: The tracking traces of 6 radio sources. (a) The trace of $S1$. (b) The trace of $S2$. (c) The trace of $S3$. (d) The trace of $S4$. (e) The trace of $S5$. (f) The trace of $S6$.

Input: Data $\mathfrak{R}_1^T = \{\mathfrak{R}_1(0), \mathfrak{R}_1(t_0), \ldots, \mathfrak{R}_1(T)\}$ recorded by deployed monitors every interval t_0 within $[0, T]$, $T = n \cdot t_0$
Output: Identification results (S_i, V_i), localization results $(\widehat{x}_i(t), \widehat{y}_i(t))$ $(1 \leq i \leq M, 0 \leq t \leq T)$
    Offline stage
(1) Derive $\mathfrak{R}_2^T$ based on $\mathfrak{R}_1^T$ according to (6) and (7)
(2) Configure network parameters (i.e., training epochs)
(3) Determine $\Omega_{\mathfrak{R}_2}$ and $\Psi_{\mathfrak{R}_2}$ by training the network
(4) Return the trained YOLO v3 network
    Online stage
(5) Derive the RSSI map according to $\mathfrak{R}_2^T$
(6) Input the established RSSI map to the network
(7) Derive identification results as (S_i, V_i) based on step 6
(8) Derive localization results as $(\widehat{x}_i(t), \widehat{y}_i(t))$ based on (8)
(9) Return (S_i, V_i), $(\widehat{x}_i(t), \widehat{y}_i(t))$

ALGORITHM 1: Indoor radio sources tracking algorithm.

TABLE 2: Generality evaluation results.

Experiment scene	Radio source	V_i^{ave}	μ_i (m)
	S1	0.9217	0.49
	S2	0.9029	0.63
Room 705	S3	0.9311	0.52
	S4	0.9168	0.65
	S1	0.9065	0.59
Room 725	S2	0.9436	0.57
	S3	0.9179	0.61

5. Conclusion

In this paper, we proposed an algorithm to identify and localize indoor robots (radio sources) in the real time. Experiments show that the proposed algorithm can not only do well in the indoor radio sources identification with 93.18% average identification precision but also is good at localizing them with 0.39 m average positioning error under typical parameters settings. In the future, it would be more interesting to extend the proposed algorithm to enable an incremental robots tracking with variable number of unknown robots.

Data Availability

All data are available within this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] M. Thomas, "2020 global threat intelligence report," 2020, https://at.nttdata.com.

[2] M. Brundage, S. Avin, J. Clark, and H. Toner, "The malicious use of artificial intelligence: forecasting, prevention, and mitigation," 2018, https://www.eff.org.

[3] S. Sami, D. Yimin, X. T. Sean Rui, R. Nirupam, and H. Jun, "Spying with your robot vacuum cleaner: eavesdropping via lidar sensors," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pp. 354–367, Virtual Event, Japan, November 2020.

[4] J. Song, Q. Zhong, W. Wang, C. Su, Z. Tan, and Y. Liu, "FPDP: flexible privacy-preserving data publishing scheme for smart agriculture," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17430–17438, 2021.

[5] L. Zhang, M. Peng, W. Wang, Z. Jin, Y. Su, and H. Chen, "Secure and efficient data storage and sharing scheme for blockchain-based mobile-edge computing," *Transactions on Emerging Telecommunications Technologies*, Article ID e4315, 2021.

[6] W. Wang, H. Huang, L. Zhang, and C. Su, "Secure and efficient mutual authentication protocol for smart grid under blockchain," *Peer-to-Peer Networking and Applications*, vol. 14, pp. 1–13, 2020.

[7] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for IIoT devices," *IEEE Transactions on Industrial Informatics*, vol. 2021, Article ID 3084753, 2021.

[8] L. Zhang, Z. Zhang, W. Wang, Z. Jin, Y. Su, and H. Chen, "Research on a covert communication model realized by using smart contracts in blockchain environment," *IEEE Systems Journal*, vol. 2021, Article ID 3057333, 2021.

[9] Y. Zou, L. Zhang, W. Wang, Z. Jin, Y. Su, and H. Chen, "Resource allocation and trust computing for blockchain-enabled edge computing system," *Computers & Security*, vol. 105, Article ID 102249, 2021.

[10] D. Dardari, P. Closas, and P. M. Djuric, "Indoor tracking: theory, methods, and technologies," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1263–1278, 2015.

[11] R. Liu, S. H. Marakkalage, M. Padmal et al., "Collaborative SLAM based on WiFi fingerprint similarity and motion information," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1826–1840, 2020.

[12] X. Liang, H. Wang, Y.-H. Liu, B. You, Z. Liu, and W. Chen, "Calibration-free image-based trajectory tracking control of mobile robots with an overhead camera," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 2, pp. 933–946, 2020.

[13] M. M. Rahman, V. Moghtadaiee, and A. G. Dempster, "Design of fingerprinting technique for indoor localization using AM radio signals," in *Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–7, Sapporo, Japan, September 2017.

[14] L. Yu, Y. W. Leung, X. Chu, and J. K. Y. Ng, "Multi-Fingerprint for wireless localization in time-varying indoor environment,," in *Proceedings of the GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, December 2020.

[15] A. Poulose and D. S. Han, "Indoor localization using PDR with Wi-Fi weighted path loss algorithm," in *Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 689–693, Jeju Island, Korea (South), October 2019.

[16] J. Zhu, Q. Chen, and J. Zhang, "Localization optimization algorithm of maximum likelihood estimation based on received signal strength," in *Proceedings of the 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, pp. 830–834, Guangzhou, China, May 2017.

[17] M. Fazelinia, M. R. Daliri, and S. Ebadollahi, "Wi-Fi RSS-based indoor localization using reduced features second order discriminant function," in *Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE)*, pp. 921–924, Yazd, Iran, April 2019.

[18] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: a survey," *Journal of Electronic Imaging*, vol. 11, no. 2, 2002.

[19] H. Rizk, M. Torki, and M. Youssef, "CellinDeep: robust and accurate cellular-based indoor localization via deep learning," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2305–2312, 2019.

[20] R. Ma, Q. Guo, C. Hu, and J. Xue, "An improved WiFi indoor positioning algorithm by weighted fusion," *Sensors*, vol. 15, no. 9, pp. 21824–21843, 2015.