Hindawi
Security and Communication Networks
Volume 2021, Article ID 9126833, 17 pages
https://doi.org/10.1155/2021/9126833

Research Article

A Graph Optimization-Based Acoustic SLAM Edge Computing System Offering Centimeter-Level Mapping Services with Reflector Recognition Capability

Zou Zhou,[1] **Guoli Zhang**,[1] **Fei Zheng**,[1] **Tuyang Wang**,[2] **Longjie Chen**,[1] **and Nan Duan**[1]

[1]*Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, Guilin University of Electronic Technology, Guilin 541004, China*
[2]*National Demonstration Center for Experimental Electronic Circuit Education, Guilin University of Electronic Technology, Guilin 541004, China*

Correspondence should be addressed to Fei Zheng; zhengfei@guet.edu.cn and Tuyang Wang; gdwty@foxmail.com

Received 4 September 2021; Revised 4 October 2021; Accepted 2 November 2021; Published 3 December 2021

Academic Editor: Xuyun Zhang

Robots can use echo signals for simultaneous localization and mapping (SLAM) services in unknown environments where its own camera is not available. In current acoustic SLAM solutions, the time of arrival (TOA) in the room impulse response (RIR) needs to be associated with the corresponding reflected wall, which leads to an echo labelling problem (ELP). The position of the wall can be derived from the TOA associated with the wall, but most of the current solutions ignore the effect of the cumulative error in the robot's moving state measurement on the wall position estimation. In addition, the estimated room map contains only the shape information of the room and lacks position information such as the positions of doors and windows. To address the above problems, this paper proposes a graph optimization-based acoustic SLAM edge computing system offering centimeter-level mapping services with reflector recognition capability. In this paper, a robot equipped with a sound source and a four-channel microphone array travels around the room, and it can collect the room impulse response at different positions of the room and extract the RIR cepstrum feature from the room impulse response. The ELP is solved by using the RIR cepstrum to identify reflectors with different absorption coefficients. Then, the similarity of the RIR cepstrum vectors is used for closed-loop detection. Finally, this paper proposes a method to eliminate the cumulative error of robot movement by fusing IMU data and acoustic echo data using graph-optimized edge computation. The experiments show that the acoustic SLAM system in this paper can accurately estimate the trajectory of the robot and the position of doors, windows, and so on in the room map. The average self-localization error of the robot is 2.84 cm, and the mapping error is 4.86 cm, which meet the requirement of centimeter-level map service.

1. Introduction

With the arrival of intelligent society, mobile robots have been widely used in people's life and work, which greatly facilitates people's life and work. In the unknown indoor space, robots realize positioning and navigation services need to know the surrounding environment map and their own position in the map. However, the robot does not know the indoor map information. Simultaneous Localization and Mapping (SLAM) is the service of detecting and sensing the map (contour) of the surrounding environment for a moving subject in an unknown environment, relying only on the mounted sensors, while determining its own position in the map [1]. After decades of continuous development, SLAM technology services have been widely used in mobile robotics [2], virtual/augmented reality [3], autonomous driving [4], and so on. The current mainstream SLAM techniques are classified based on the differences in the sensors used and can be divided into LIDAR SLAM techniques and visual SLAM techniques. Sensors such as LIDAR and cameras have the advantage of accuracy and high resolution but they also have disadvantages: LIDAR is a very

expensive sensor and poses health and safety issues in operation [5]. Cameras, although cost getting lower, require high processing power in low-light environments as well as low signal-to-noise ratios [6]. In addition, the above systems are computationally complex and usually use cloud-based processing, which is costly and involves privacy and security. In contrast, acoustic sensors as the standard for mobile robots can be used in low-light and dark environments [7]. Map reconstruction work can be achieved using the robot's own arithmetic power, which not only has significant cost advantages but also offloads privacy-aware services to MEC (mobile edge computing), avoiding the leakage of private information such as indoor images. Therefore, researchers have begun to explore the implementation of acoustic SLAM.

In indoor environments, the propagation of acoustic signals is obscured and reflected by buildings resulting in multipath effects, which to some extent reflect information about the room arrangement and geometry and can be used to estimate environmental maps. Researchers have started to estimate room shapes from the room impulse response (RIR) of acoustics. Labelling the first-order echoes in the RIR to the walls that generate them is the key for room shape estimation. Since the RIR itself does not specify which reflections come from which walls, there is an echo labelling problem (ELP) [8]. Literatures [9, 10] solved the ELP using the properties of the Euclidean distance matrix (EDM), but the algorithm must traverse the TOA combinations of all echoes, which has a high computational complexity. References [8, 11, 12] improved the computational complexity of their work based on EDM using graph theory, subspace filtering, and greedy iteration, respectively, but the overall computational complexity is still large. There is also a method of solving ELP using elliptic constraints. Literatures [13, 14] solved the series of reflective points of walls based on an elliptic constraint model and finally used the Hough transform for the estimation of each reflective wall. This method needs to arrange many anchor nodes to obtain enough data, which is complex to implement and extremely computationally intensive. Literature [15] proposed an algorithm to reduce the computational complexity based on elliptic constraints for iterative echo marking. The above method uses a stationary distributed microphone array, which requires the sound source and microphone array to be arranged in the room in advance and is not applicable to the practical application scenario of SLAM.

In addition to the static deployment of sources and microphones scheme in the above work, there is another scheme that embeds acoustic sensors on a mobile robot, which is more in line with the practical needs of SLAM and is more relevant for research. Whether the robot is equipped with an acoustic source can be classified as active acoustic SLAM and passive acoustic SLAM. References [16, 17] used robots equipped with multichannel microphones for their own localization as well as localization of acoustic sources by sensing the ambient sound sources around them. However, suitable sound sources that can be detected are not always available in the actual environment. Another option is to use a robot equipped with both sound sources and microphones for simultaneous localization and mapping. In literatures [18–20], a mobile robot with a juxtaposition of an acoustic source and a single-channel microphone was used to collect first-order echoes. Due to the weak spatial perception of the single-channel microphone, the robot moved at least three times for a reflective wall estimation, and the movement error may affect the accuracy of the reflective wall estimation. Multichannel microphones have better spatial perception capability than single microphones and can obtain more information about the interior geometry for the same number of measurements. Literature [21] achieved room shape estimation using a robot equipped with an acoustic source and a four-channel microphone by estimating the location of the first-order image source by clustering, but the robot needs to collect a large amount of RIR data at each movement. All the above acoustic SLAM schemes can achieve the estimation of room shape but ignore the effect of cumulative errors in the measurement of the robot's moving state, that is, on the reconstruction of the room contour. The room map has only room shape information and lacks the location information of doors and windows because it cannot distinguish between different reflective materials.

In this paper, a robot active sounding scheme equipped with both sound sources and microphone arrays is used for simultaneous localization and mapping. To address the ELP problem mentioned above and the cumulative error of robot movement measurement affecting the accuracy of map building, this paper proposes a graph-optimized acoustic SLAM edge computing system based on graph optimization that can identify room detail information. The main contributions of this paper are as follows:

(1) A robot system prototype is designed and implemented that can be used for acoustic SLAM

(2) A first-order echo labelling algorithm based on RIR cepstrum is proposed, which solves ELP by distinguishing different reflective materials

(3) A graph optimization-based method is proposed for correcting the pose estimated by the trapezoidal constraint

The sections of this paper are organized as follows. Section 1 introduces the background and current research status of acoustic SLAM, and Section 2 describes the problem setup and the architecture of the graph optimization-based acoustic SLAM system. In Section 3, a prototype of our designed robotic system for acoustic SLAM is presented. Section 4 introduces the related acoustic SLAM methods. Section 5 shows the simulations and experiments, and Section 5 concludes.

2. Problem Description and System Structure

In an indoor environment, the sound signal received by a microphone consists of the direct sound from the source and the reflected sound that is reflected by the walls. In the image source model [22, 23], the reflected sound from the actual sound source was replaced by the direct sound from the image source, as shown in Figure 1(a). For a first-order echo

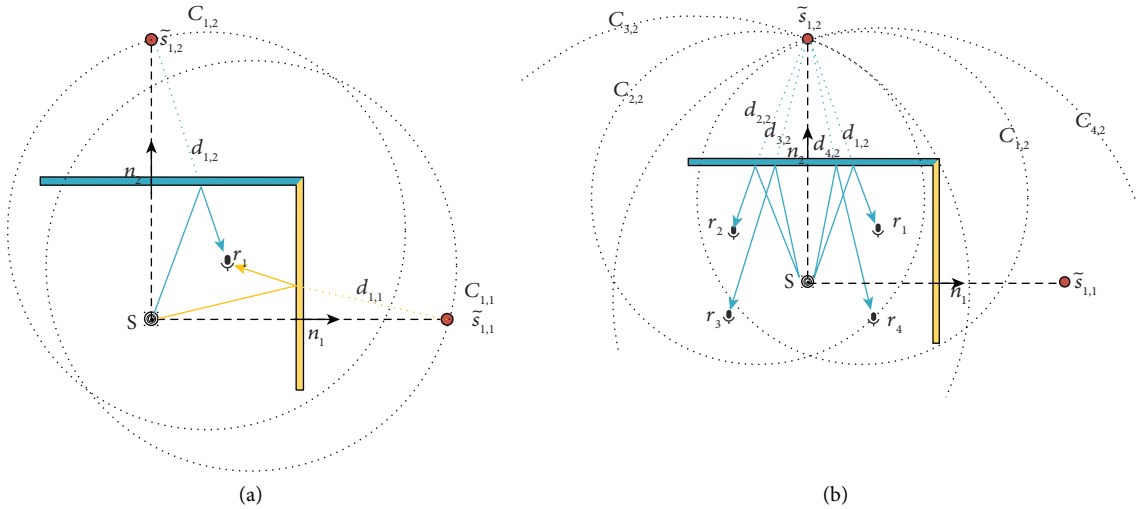


FIGURE 1: Schematic diagram of the first-order image sound source: (a) single source with a single microphone and (b) single source with a multichannel microphone.

described by the unit normal n_k and an arbitrary wall point p_k and the kth reflected wall, the first-order image source $\tilde{s}_{1,k}$ of the real sound source s with microphone r_j is calculated according to $\tilde{s}_{1,k} = r_j + 2\langle p_k - r_j, n_k\rangle n_k$. The sound propagation process was described in terms of the room impulse response (RIR), which consists of a series of Dirac pulses δ:

$$h_j(t) = \sum_i \alpha_i \delta(t - \tau_i) + \varepsilon(t), \tag{1}$$

where $\varepsilon(t)$ is the noise, α_i is the amplitude of the received pulse, and its magnitude depends on the absorption coefficient of the wall and the distance from the image source to the microphone [18]. τ_i is the arrival time of the corresponding pulse, which is proportional to the distance from the image source to the microphone r_j. The room impulse response can be represented as a dataset $\text{data}_j = \{(\alpha_i, \tau_i), i = 1, 2, \ldots, n\}$. ELP finds the data (α_i, τ_i) associated with the first-order image source $\tilde{s}_{1,k}$ from the dataset data_j. The label_i was defined as the label corresponding to data (α_i, τ_i). The data association process for a first-order image source $\tilde{s}_{1,k}$ can be represented by the function $L(\alpha_i, \tau_i)$:

$$L(\alpha_i, \tau_i) = \begin{cases} 1, & \text{label}_i = k, \\ 0, & \text{else}. \end{cases} \tag{2}$$

The distance from the first-order image source $\tilde{s}_{1,k}$ to the microphone r_j can be known from the marked first-order echo data. For a single-channel microphone, as shown in Figure 1(a), the position of the image source $\tilde{s}_{1,k}$ is on a circle $C_{1,k}$ with the position of microphone r_1 as the center and $d_{j,k}$ as the radius, and the exact position of the image source $\tilde{s}_{1,k}$ on the circle cannot be determined due to the lack of spatial information. For multichannel microphones, as shown in Figure 1(b), the position of the image source $\tilde{s}_{1,k}$ is at the intersection point between circles $C_{j,k}$. In 2D space, it is known from the TOA localization algorithm [24] that the uniqueness of the image source s location can be guaranteed when the number of microphones is greater than 3. The midpoint of the line connecting the real source s and the image source $\tilde{s}_{1,k}$ is the location of the reflecting wall.

In this paper, an omnidirectional sound source was defined, and an omnidirectional 4-channel microphone array are installed on the robot, and the location of the microphone array in relation to the sound source is shown in Figure 2(a). The robot travels around the room in a circle, and for each step the robot takes, the sound source generates a pulse while the microphone array records an echo. The room was defined as a 2D polygon for the sake of descriptive simplicity, and the approach in this paper can be easily extended to 3D.

The robot can estimate the position of each wall relative to itself in the room using echo information, as shown in Figure 2(b). Based on the above description, the difficulty of first-order image source position estimation is solving the echo labelling problem (ELP) [8]. In this paper, taking advantage of the strong spatial perception of multichannel microphones, an RIR cepstrum feature that can distinguish reflective walls with different absorption coefficients was proposed, and based on this feature, this paper proposes a solution for first-order echo labelling based on the RIR cepstrum.

The robot travels around the room in a circle and uses the echo information from different locations to estimate the distance from the wall to itself for the shape estimation of the indoor room, as shown in Figure 2(b). Since there is a cumulative error in the robot position estimated by IMU data, this can lead to inaccurate estimation of the wall position. To solve this problem, a graph optimization-based acoustic SLAM method was proposed, which uses graph optimization to fuse the robot pose estimated by the sound echo signal with the pose estimated by IMU to eliminate the cumulative error, and the system block diagram of this method is shown in Figure 3. Referring to other graph optimization structures [25–27], the graph optimization system in this paper is also mainly divided into two parts: front end and back end. The front end establishes the graph

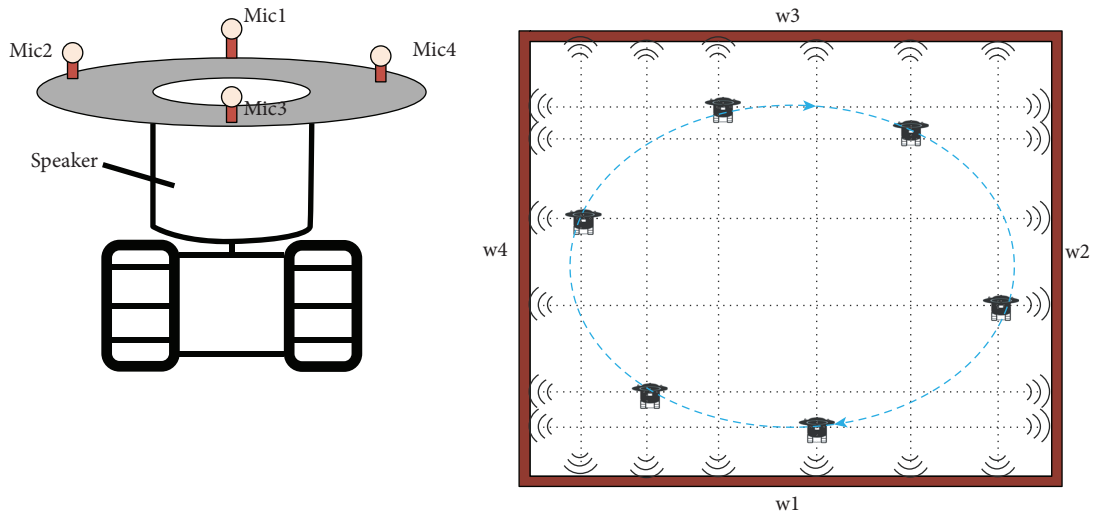


Figure 2: (a) Layout of the sound source and microphone on the robot. (b) Schematic diagram of the robot moving in the room.

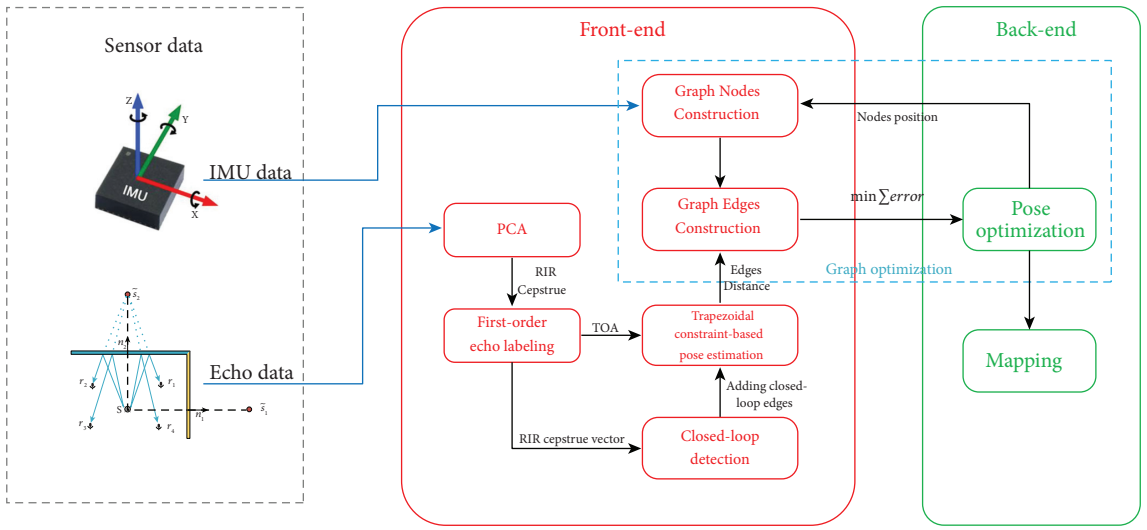


Figure 3: Framework of an acoustic SLAM system based on graph optimization.

vertices and the positional constraint relations between graph vertices based on the IMU sensor and acoustic sensor data, and the back end optimizes the positional graph based on the closed-loop constraints added by the closed-loop detection and the constraint relations between graph vertices to finally obtain globally consistent robot trajectories and indoor room maps.

3. Acoustic SLAM Method

According to the system framework of acoustic SLAM, the acoustic SLAM method can be divided into two parts: echo labelling and pose correction. Echo labelling extracts the first-order echo signal from the echo signal and then estimates the position of the first-order image sound source based on the first-order echo signal. The pose correction is to eliminate the accumulated errors during the robot movement globally using graph optimization methods.

3.1. Echo Labelling Based on the RIR Cepstrum Feature. The ELP problem is often solved using the Euclidean distance matrix approach, which requires traversing all possible combinations of echoes with high complexity. In this section, an echo labelling method based on RIR cepstrum is proposed, which can achieve fast and accurate echo labelling.

3.1.1. RIR Cepstrum. Multichannel microphones are more spatially aware than single-channel microphones, and a spatial cepstrum feature is proposed in literature [28], which

can represent the relative position of the sound source in the room. Inspired by this, the room impulse response cepstrum feature was proposed, which can be used for first-order echo labelling and loopback detection.

Suppose the robot is equipped with M omnidirectional microphones and an omnidirectional sound source s. As shown in Figure 4, the robot moves N steps from x_1 to x_N, and each time it moves, the robot's sound source generates a pulse, while the microphone acquires the room impulse response at the current position. The robot can acquire M room impulse responses $h_{i,j}(n)$, $j = 1, 2, \ldots, M$ at x_i. According to equation (1), $h_{i,j}(n)$ consists of a series of pulses, and the average energy feature $r_{i,j,k}$ of the kth pulse is extracted:

$$r_{i,j,k} = \sqrt{\frac{1}{2L+1} \sum_{n=\tau_{i,j,k}-L}^{\tau_{i,j,k}+L} h_{i,j}(n)^2}, \tag{3}$$

where $\tau_{i,j,k}$ is the TOA value of the kth pulse in $h_{i,j}(n)$ and $2L+1$ is the width of the rectangular window.

The time delay feature of the kth pulse in $h_{i,j}(n)$ is

$$t_{i,j,k} = \tau_{i,j,k} \cdot c, \tag{4}$$

where c is the speed of sound propagation in the air.

Log operations are performed on the above two features separately to obtain the log energy vector $p_{i,k}$ and the log time delay vector $q_{i,k}$:

$$\begin{aligned} p_{i,k} &= \left(\log\left(r_{i,1,k}\right) \ \log\left(r_{i,2,k}\right) \ \cdots \ \log\left(r_{i,M,k}\right) \right)^T, \\ q_{i,k} &= \left(\log\left(t_{i,1,k}\right) \ \log\left(t_{i,2,k}\right) \ \cdots \ \log\left(t_{i,M,k}\right) \right)^T. \end{aligned} \tag{5}$$

The robot is obtained from $x_1 \longrightarrow x_N, N > M$; the matrix of the average amplitude logarithm of the impulse response about the room A_k and the matrix of the logarithm of the arrival distance B_k can be obtained as follows:

$$\begin{aligned} A_k &= \left(p_{1,k} \ \ p_{2,k} \ \cdots \ p_{n,k} \right)^T, \\ B_k &= \left(q_{1,k} \ \ q_{2,k} \ \cdots \ q_{n,k} \right)^T. \end{aligned} \tag{6}$$

As in the method for extracting the spatial cepstrum [29], we also use PCA instead of DFT or DCT. R_{A_k} can be obtained from A_k.

$$R_{A_k} = A_k A_k^T. \tag{7}$$

Since R_{A_k} is a symmetric matrix, R_{A_k} the eigenvalue decomposition can be expressed as follows:

$$R_{A_k} = E_{A_k} D_{A_k} E_{A_k}^T, \tag{8}$$

where E_{A_k} is the eigenvector matrix, D_{A_k} is the diagonal matrix, and the diagonal elements are the eigenvalues, in descending order.

After PCA dimensionality reduction, the data d_{A_k} can be expressed as

$$d_{A_k} = E_{A_k}\left(A_k - \overline{A_k}\right). \tag{9}$$

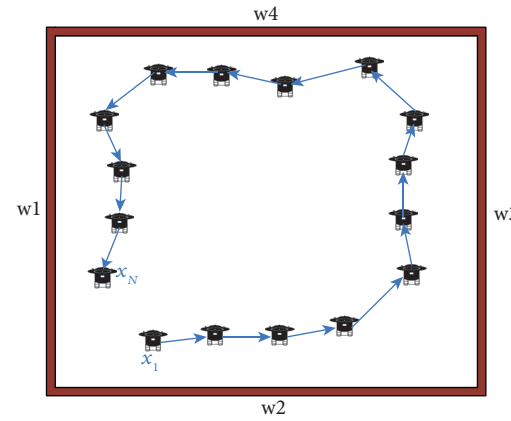


Figure 4: Schematic diagram of robot movement.

Similarly, for the matrix B_k, PCA dimensionality reduction is performed.

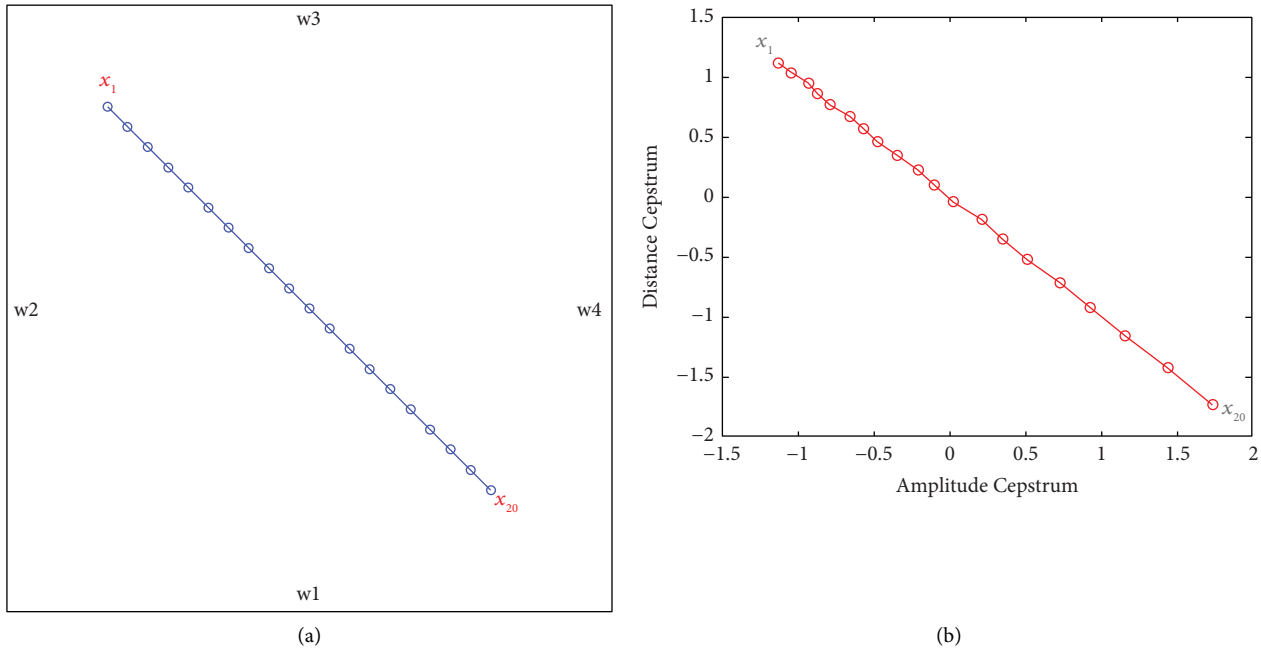
$$d_{B_k} = E_{B_k}\left(B_k - \overline{B_k}\right). \tag{10}$$

The principal component components are selected from d_{A_k} and d_{B_k}, and they are the components with the largest eigenvalues d_{a_k} and d_{b_k}, which form the room impulse response cepstrum d_{h_k}.

$$d_{h_k} = \left[d_{a_k} \ \ d_{b_k} \right], \tag{11}$$

where d_{h_k} is the matrix, d_{h_k} is defined as the room impulse response cepstrum, d_{a_k} is the amplitude cepstrum, and d_{b_k} is the distance cepstrum.

The amplitude cepstrum corresponds to the average amplitude of the pulses observed by the microphone array. When the pulses observed by the microphone array are consistent, i.e., the first-order echoes come from the same reflecting wall, the magnitude of the amplitude cepstrum is inversely proportional to the distance of the robot from the reflecting wall. Similarly, when the pulses observed by the microphone array are consistent, the magnitude of the distance cepstrum is proportional to the distance of the robot from the reflecting wall. As shown in Figure 5(a), in a 6 m * 6 m rectangular room, the robot moves from x_1 to x_{20}, and to ensure the consistent observation of the microphone matrix, the reflection coefficients of walls w1–w4 to 0.8, 0, 0, and 0 were defined. At this time, the robot can only receive the echo from wall w1. We select the second pulse in the room impulse response and extract the RIR cepstrum d_{h_2} according to the above method and represent the RIR cepstrum in a two-dimensional Cartesian coordinate system, as shown in Figure 5(b). We can observe that when the pulses observed by the microphone matrix are consistent, the cepstrum of the room impulse response d_{h_k} of d_{a_k} and d_{b_k} approximates a linear relationship. When the value of the RIR cepstrum of the microphone array pulse combination is in the vicinity of the straight line corresponding to the reflective wall and then combined with the size of the microphone array, it can be determined whether the microphone array is observed consistently.



(a) (b)

FIGURE 5: (a) Schematic diagram of the robot's trajectory in a 6 m ∗ 6 m room; (b) the RIR cepstrum d_{h_2} mapping schematic in the two-dimensional coordinate system.

3.1.2. Echo Labelling Based on the RIR Cepstrum Feature. Based on the feature that the one-dimensional component of the RIR cepstrum approximately satisfies a linear relationship with the two-dimensional component when the RIR cepstrum is observed consistently, a method for first-order echo labelling was proposed. According to the image source model, it is known that the 2nd and 3rd pulses received by the microphone must be the first-order echoes reflected from the wall. The robot follows the trajectory in Figure 4 from x_1 moving to x_N. Since the robot moves against the wall and the spacing between the microphones is small, it can be guaranteed that the 2nd and 3rd pulses observed by the microphones are mostly from the same virtual source, i.e., the observations are consistent. Assuming that the acoustic reflection coefficients of each of the four walls of the room in Figure 4 are different, the robot takes the 2nd and 3rd pulses from $x_1 \longrightarrow x_N$. The second and third pulses received by the microphone are taken to find the cepstrum of the room impulse response; the matrix of the logarithm of the room impulse response amplitude at this time $A_{2,3}$ and the matrix of the logarithm of the arrival distance $B_{2,3}$ are as follows:

$$A_{2,3} = \begin{bmatrix} A_2 \\ A_3 \end{bmatrix}, \quad B_{2,3} = \begin{bmatrix} B_2 \\ B_3 \end{bmatrix}. \tag{12}$$

Following the method in the previous section, PCA operations are performed on $A_{2,3}$ and $B_{2,3}$, to obtain the feature matrices $E_{A_{2,3}}$ and $E_{B_{2,3}}$, respectively. The RIR cepstrum after PCA dimensionality reduction is

$$d_{h_{2,3}} = \left[d_{a_{2,3}} \;\; d_{b_{2,3}} \right]. \tag{13}$$

According to the characteristics of the RIR cepstrum feature, $d_{h_{2,3}}$ can be fitted as a straight line l_i corresponding to the reflecting wall i with different reflection coefficients.

$$a_i x + b_i y + c_i = 0. \tag{14}$$

When the kth pulse observed by the microphone matrix is greater than 3, it is not possible to determine whether the kth pulse observed at this time is a first-order echo. The first-order echo candidates from wall i can be obtained from the TOA values of the first and second echoes and the relationship between the microphone positions. These candidates can be combined to obtain a new combination of room pulses, which corresponds to the RIR cepstrum d_τ as follows:

$$d_\tau = \left[d_{a,\tau} \;\; d_{b,\tau} \right]. \tag{15}$$

If the new combination is a first-order echo from wall i, the corresponding room impulse response cepstrum d_τ should be near the straight line l_i, and the distance from d_τ to the straight line l_i satisfies the following equation:

$$D_{i,\tau} = \frac{\left| a_i d_{a,\tau} + b_i d_{b,\tau} + c_i \right|}{\sqrt[2]{a_i^2 + b_i^2}} < \Delta\varepsilon, \tag{16}$$

where $\Delta\varepsilon$ is the Euclidean distance threshold. Following the above method, the first-order echoes of different walls can be distinguished, and thus, the location of the image source can be estimated.

3.2. Pose Correction Based on Graph Optimization. The pose correction method consists of three parts: closed-loop

detection, pose estimation, and pose correction. In this paper, the pose at different locations of the robot is used as nodes of the graph. The constraint relationship between graph nodes is established using closed-loop detection and pose estimation. Finally, the graph optimization method is used to correct the robot's pose.

3.2.1. Closed-Loop Detection. Closed-loop detection determines whether the robot has reached the previous position, and it is extremely important for back-end optimization. After the above echo labelling method, the robot at the x_i location can obtain the RIR cepstrum consistent with the k-sided wall observation, and these cepstrums can be combined into a 2k-dimensional RIR cepstrum vector X_i as follows:

$$X_i = [\, x_{i1} \;\; y_{i1} \;\; x_{i2} \;\; y_{i2} \;\; \dots \;\; x_{ik} \;\; y_{ik} \;\;], \tag{17}$$

where $[\, x_{i,k} \;\; y_{i,k} \,]$ is the RIR cepstrum from wall k.

Similarly, the robot can obtain RIR cepstrum vector X_j at x_j as follows:

$$X_j = \left[\, x_{j1} \;\; y_{j1} \;\; x_{j2} \;\; y_{j2} \;\; \dots \;\; x_{jk} \;\; y_{jk} \;\; \right]. \tag{18}$$

The vectors X_i and X_j can be used to express the Euclidean distance between them to express the similarity of their spaces. When x_i and x_j are in the same position or close to each other, the Euclidean distance between the two vectors should satisfy the following formula:

$$\text{distance}_{ij} = \left| X_j - X_i \right| < \delta, \tag{19}$$

where δ is the Euclidean distance threshold.

3.2.2. Trapezoidal Constraint-Based Pose Estimation. In indoor space, the actual position of the robot and the position movement of the image source satisfy the isosceles trapezoidal constraint [18, 29], as shown in Figure 6. Based on this constraint, a robot positional estimation method was proposed.

In world coordinates, let the robot's pose at x_{i-1} be $X_{i-1} = (x_{i-1}, y_{i-1}, \theta_{i-1})$ and the robot's pose at x_i be $X_i = (x_i, y_i, \theta_i)$. The change in pose dX_i of the robot moving from x_{i-1} to x_i is

$$dX_i = X_i - X_{i-1} = (x_i - x_{i-1}, y_i - y_{i-1}, \theta_i - \theta_{i-1}), \tag{20}$$

where dX_i is expressed in polar coordinates as

$$dX_i = (r \cos \alpha, r \sin \alpha, \theta). \tag{21}$$

In equation (21), r is the displacement variable, α is the angle of the displacement direction to the X-axis of the world coordinate system, and θ is the rotation angle of the robot coordinate system.

The coordinates of the image source at x_{i-1} and x_i are expressed in polar coordinates in the robot coordinate system, respectively, as follows:

$$\begin{cases} \widetilde{x}_{i-1,1} = \left(r_{i-1,1}, \theta_{i-1,1}\right), \\ \widetilde{x}_{i-1,2} = \left(r_{i-1,2}, \theta_{i-1,2}\right), \\ \widetilde{x}_{i-1,3} = \left(r_{i-1,3}, \theta_{i-1,3}\right), \\ \widetilde{x}_{i-1,4} = \left(r_{i-1,4}, \theta_{i-1,4}\right), \end{cases} \longrightarrow \begin{cases} \widetilde{x}_{i,1} = \left(r_{i,1}, \theta_{i,1}\right), \\ \widetilde{x}_{i,2} = \left(r_{i,2}, \theta_{i,2}\right), \\ \widetilde{x}_{i,3} = \left(r_{i,3}, \theta_{i,3}\right), \\ \widetilde{x}_{i,4} = \left(r_{i,4}, \theta_{i,4}\right). \end{cases} \tag{22}$$

The robot rotation angle θ is related only to the angle of the polar coordinates of the image source.

$$\theta = \theta_{i,1} - \theta_{i-1,1} = \theta_{i,2} - \theta_{i-1,2} = \theta_{i,3} - \theta_{i-1,3} = \theta_{i,4} - \theta_{i-1,4}. \tag{23}$$

The estimated value of the robot rotation angle is

$$\widehat{\theta} = \frac{1}{4} \sum_{w=1}^{w=4} \theta_{i,w} - \theta_{i-1,w}. \tag{24}$$

Robots from x_{i-1} move to x_i, and the length of the image acoustic source polar coordinates changes as follows:

$$\begin{cases} r_{i,1} = r_{i-1,1} + 2r \, \sin\left(\alpha + \alpha_{i-1,1}\right), \\ r_{i,2} = r_{i-1,2} + 2r \, \sin\left(\alpha + \alpha_{i-1,2}\right), \\ r_{i,3} = r_{i-1,3} + 2r \, \sin\left(\alpha + \alpha_{i-1,3}\right), \\ r_{i,4} = r_{i-1,4} + 2r \, \sin\left(\alpha + \alpha_{i-1,4}\right), \end{cases} \tag{25}$$

where $\alpha_{i-1,n} = \theta_{i-1,1} - \theta_{i-1,n}, \quad n = 1, 2, 3, 4.$

Let $s = (r, \alpha)$, the following vector function can be obtained:

$$\begin{aligned} d(s) &= \left[\left\|r_{i,1} - r_{i-1,1}\right\|, \left\|r_{i,2} - r_{i-1,2}\right\|, \left\|r_{i,3} - r_{i-1,3}\right\|, \left\|r_{i,4} - r_{i-1,4}\right\|\right], \\ &= \left[2r \, \sin\left(\alpha + \alpha_{i-1,1}\right), 2r \, \sin\left(\alpha + \alpha_{i-1,2}\right),\right. \\ &\quad \left. 2r \, \sin\left(\alpha + \alpha_{i-1,3}\right), 2r \, \sin\left(\alpha + \alpha_{i-1,4}\right)\right]. \end{aligned} \tag{26}$$

The estimated value of the acoustic sensor $\widehat{d}$ has a random error ζ, and the variance of the error is σ^2.

$$\widehat{d} = d(s) + \varsigma. \tag{27}$$

Weighted error function (s) is as follows:

$$\varepsilon(s) = (\widehat{d} - d(s))^T \Phi_n^{-1} (\widehat{d} - d(s)), \tag{28}$$

where $\Phi_n = \sigma^2 I$ and I is the unit matrix.

The objective optimization function is obtained by minimizing the weighted error function $\varepsilon(s)$.

$$\widehat{s} = \arg\min_s \varepsilon(s). \tag{29}$$

The Levenberg–Marquardt algorithm is used to solve equation (29).

Linearize the error function $\varepsilon(s)$ by linearly expanding $d(s)$ through a first-order Taylor series:

$$d(s + \Delta s) = d(s) + J(s)\Delta s, \tag{30}$$

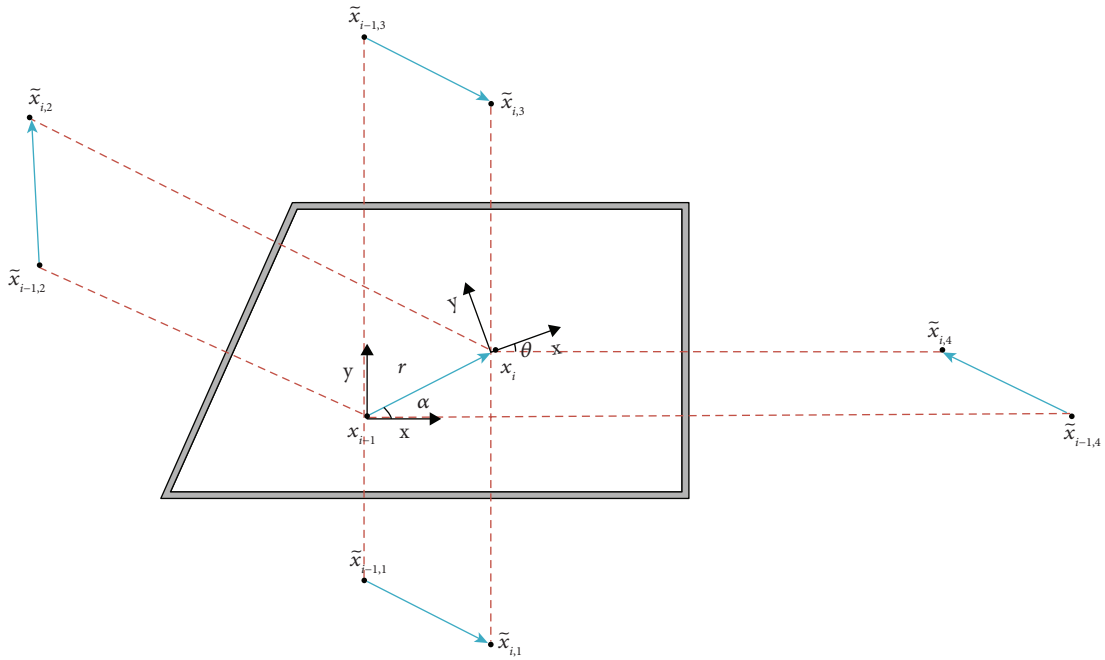


Figure 6: Isosceles trapezoidal constraint diagram.

where $J(s)$ is the Jacobi matrix, which is expressed as follows:

$$J(s) = \begin{bmatrix} 2r\sin\left(\alpha + \alpha_{i-1,1}\right) & 2r\cos\left(\alpha + \alpha_{i-1,1}\right) \\ 2r\sin\left(\alpha + \alpha_{i-1,2}\right) & 2r\cos\left(\alpha + \alpha_{i-1,2}\right) \\ 2r\sin\left(\alpha + \alpha_{i-1,3}\right) & 2r\cos\left(\alpha + \alpha_{i-1,3}\right) \\ 2r\sin\left(\alpha + \alpha_{i-1,4}\right) & 2r\cos\left(\alpha + \alpha_{i-1,4}\right) \end{bmatrix}. \quad (31)$$

Equation (30) is brought into equation (28) to solve for the extreme value of the weighted error function $\varepsilon(s)$. According to the Levenberg–Marquardt method:

$$(H + \lambda I)\Delta s = g, \quad (32)$$

where $H = J(s)^T J(s)$ and $g = J(s)(\hat{d} - d(s))$.

The above equation allows to find the step size Δs_k for each iteration. The value $s_0 = (r_0, \alpha_0)$ estimated by the IMU is used as the initial value for the iterative calculation.

$$\hat{s} = s_0 + \sum_{k=1}^{n} \Delta s_k, \quad (33)$$

where n is the number of iterations.

According to the above method, it is possible to use the acoustic signal to accurately estimate the pose change between different positions of the robot.

3.2.3. Pose Correction Based on Graph Optimization. According to the above method, we can construct the graph. Every time the robot moves a certain distance or rotates a certain arc, a vertex is added to the graph, and the constraint relationship between the vertices is established according to the pose estimation algorithm. The structure of the graph is shown in Figure 7.

Let $x = (x_1, x_2, \ldots, x_T)$ be a vector of parameters, where x_i describes the pose of node i. The robot moves from the pose node xi to the pose node x_j, $\hat{z}_{ij}$ is the pose transformation estimated by the IMU and z_{ij} is the pose transformation observed by the acoustic sensor. Let $e(x_i, x_j)$ be the error function from x_i to x_j, which is the difference between the robot's predicted observation $\hat{z}_{ij}$ and the actual observation z_{ij}.

$$e_{ij}\left(x_i, x_j\right) = z_{ij}\left(x_i, x_j\right) - \hat{z}_{ij}\left(x_i, x_j\right). \quad (34)$$

The dashed box in Figure 7 shows the constraint relationship between node x_2 to node x_p.

Let C be the set of constraint pairs of nodes in the graph and the set of nodes in the graph of trajectory points x of the robot. The goal of the maximum likelihood method is to find the configuration of nodes x^* that minimizes the negative log likelihood $F(x)$ of all observations.

$$F(x) = \sum_{(i,j)\in c} e_{ij}^T \Omega_{ij} e_{ij}, \quad (35)$$

where Ω_{ij} is the measurement information matrix of the error function e_{ij}.

The objective optimization function is

$$x^* = \arg\min_x F(x). \quad (36)$$

Using the first-order Taylor expansion error function $e_{ij}(x_i, x_j)$.

$$e\left(x_i + \triangle x_i, x_j + \triangle x_j\right) = e_{ij}\left(x_i, x_j\right) + J_{ij}\Delta x. \quad (37)$$

The error function $e_{ij}(x_i, x_j)$ is only related to x_i and x_j. Its Jacobi matrix J_{ij} is

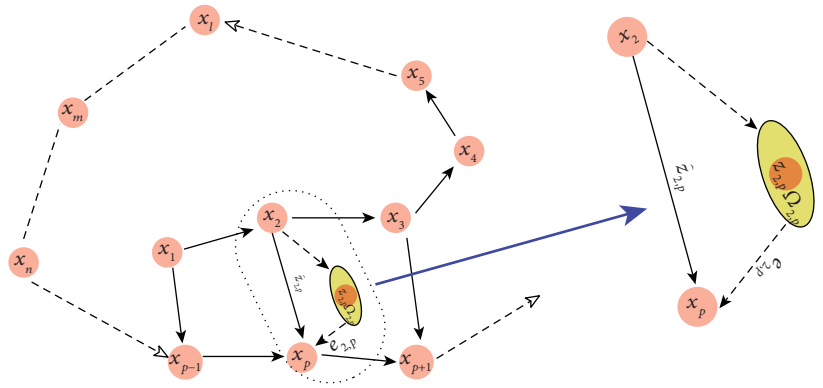


Figure 7: Schematic diagram of the connection between nodes.

$$J_{ij} = \frac{\partial e_{ij}\left(x_i, x_j\right)}{\partial x} = \left(0 \ldots 0, \frac{\partial e_{ij}\left(x_i, x_j\right)}{\partial x_i}, 0 \ldots 0, \frac{\partial e_{ij}\left(x_i, x_j\right)}{\partial x_j}, 0 \ldots 0\right). \tag{38}$$

The nonlinear optimization algorithm is used to iteratively solve for the minimum value of $F(x)$. The step size of each iteration can be solved using equation (32).

$$\Delta x = -(H + \lambda I)^{-1} b, \tag{39}$$

where $H = \sum_{(i,j)\in c} {J_{ij}}^T \Omega_{ij} J_{ij}$ and $b = \sum_{(i,j)\in c} {J_{ij}}^T \Omega_{ij} e_{ij}$.

The optimal robot trajectory point x^* is obtained by iterative calculation.

$$x^* = x_0 + \sum_{k=1}^{n} \Delta x_k, \tag{40}$$

where x_0 is the initial trajectory point of the robot estimated by the IMU and n is the number of iterations.

The robot travels around the room once, constructs the vertices and edges of the graph according to the method in this paper, and solves equation (36) using a nonlinear optimization algorithm. The optimal robot movement trajectory is obtained by optimizing the length of the edges of the graph to minimize $F(x)$.

4. System Implementation and Experimental Verification

This section introduces our self-designed robot prototype and then experimentally verifies the performance of the acoustic SLAM method in this paper.

4.1. System Implementation. Our robot is based on the Turtlebot3 Waffle Pi robot, a small, low-cost, fully programmable, ROS-based mobile robot, as shown in Figure 8(b). Turtlebot3 consists mainly of Raspberry Pi3 and OpenCR control board (with IMU sensor inside). In this system, OpenCR is responsible for collecting the built-in IMU sensor data and sound data as well as driving the robot to move, Raspberry Pi 3 is responsible for processing and calculating the data, and Raspberry Pi 3 is connected to OpenCR via USB 2.0. The system architecture connection diagram of the robot is shown in Figure 8(a). Since the sound source is close to the microphone array, which may result in larger direct waves and affect the reception of other reflected waves, a sound insulation panel was designed between the microphone array and the sound source to isolate the direct sound. In addition, the sound insulation panel can also isolate the reflected waves from the upper and lower walls. The physical diagram of the robot is shown in Figure 8(c).

The robot uses an active acoustic scheme for self-localization and room contour estimation, and to prevent the sound signals emitted during the robot's work from affecting people's life and work, this paper uses sound pulse signals in the pseudoultrasonic band (16k–24k), which has a wavelength between 1.4 cm and 2.1 cm and can be guaranteed to be received by a small microphone array, and the sound signals in this band are insensitive to the human ear but can be picked up by the robot's acoustic sensors. Sound source emits a sound signal of 16 kHz–20 kHz chirp pulse signal, which can avoid the leakage of sensitive information such as indoor human voice and ensure the security of privacy, but most speakers do not support the sound signal of the band, the need for speaker selection. Sound generation equipment was used, that is, Huawei Sound is used as the sound source. Huawei Sound is a 360-degree omnidirectional speaker with a frequency response range of 55 Hz–40 kHz and supports 3.5 mm wired audio input.

The process of acquiring the sound pulse signal from the robot itself is the process of converting the sound signal from a mechanical wave to a digital signal. The sound signal is first converted into a voltage signal through a microphone, and since this voltage signal is usually small, it needs to be amplified by an amplifier; then, the amplified signal is passed through a 15k–24k bandpass filter to filter out the noise signal outside the pseudoultrasonic band. Finally, the filtered signal is converted to a digital signal by an ADC.

Based on the acquisition process of the sound signal, the microphone array signal acquisition board for the robot was designed. The microphone in the acquisition board is a 130F22 omnidirectional microphone from PCB, which has a frequency response range of 10 Hz–20 kHz, and the

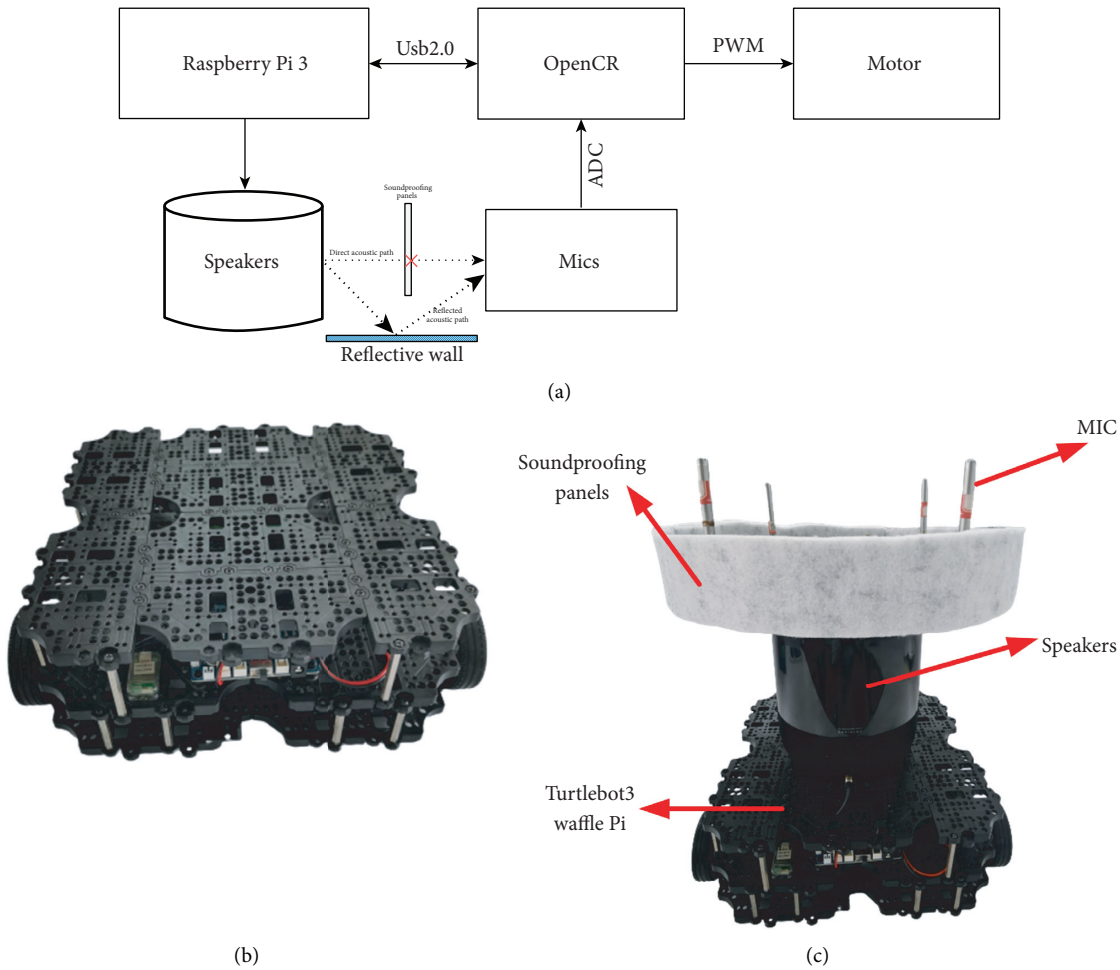


Figure 8: (a) Connection diagram of each module of the robot: (b) physical diagram of Turtlebot3 Waffle Pi; (c) physical diagram of the robot.

microphone is an SMB interface that can be plugged into the acquisition board standing up. The acquisition board is a four-channel microphone array, and the microphones are distributed at equal intervals on a circle with a radius of 0.14 m. The physical diagram of the acquisition board is shown in Figure 9.

4.2. Experimental Verification. The experimental part is divided into echo labelling experiments, pose correction and room shape estimation experiments, and real room experiments.

4.2.1. Echo Labelling Algorithm Performance Simulation. For the echo labelling experiments, echo labelling simulations were conducted in three different shapes of rooms: square, rectangular, and pipeline. The shape of the room is schematically shown in Figure 10, and w1, w2, w3, and w4 were used to denote the four walls of the room, and their corresponding reflection coefficients are $[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$. The radius of the microphone array of the robot is 0.2 m, and a location in the room is randomly selected, the RIR of each microphone under that location is simulated using the image source method, and the echoes are labelled using the method in Section 3. To verify the robustness of the algorithm to noise, the Gaussian white noise was added to the propagation distance of the signal as follows:

$$\widehat{d} = d + \varepsilon, \tag{41}$$

where d is the true propagation distance and ε is the additive noise, and its standard deviation σ varies from 0.01 to 0.05 in steps of 0.005. Similar to [15], the F1-score was used to evaluate the goodness of the echo markers.

Literature [15] solved ELP by means of elliptical iterations, and for comparison, experiments in a square room with reflection coefficients for each wall of [0.8, 0.8, 0.8, 0.8] were performed. Randomly 300 points were chosen to conduct echo labelling experiments with the method of this paper and compare with the method of literature [15]. The experimental results are shown in Figure 11(b), where method 1 is the above method and method 2 is the method of this paper.

Then, the F1-score of different numbers was simulated of microphone array robotic echo markers in each of the three rooms in Figure 10, where the reflection coefficient of the room is [0.5, 0.6, 0.7, 0.8]. The experimental results are

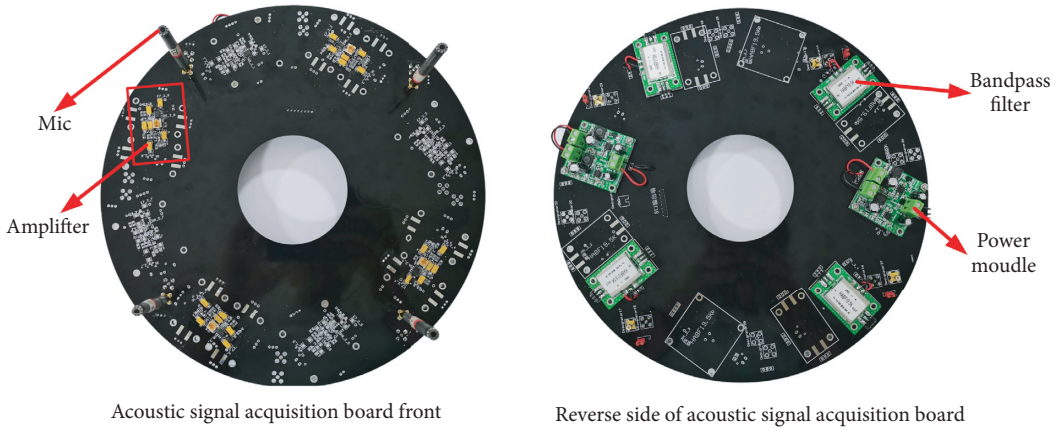


Figure 9: Physical diagram of the acoustic signal acquisition board.

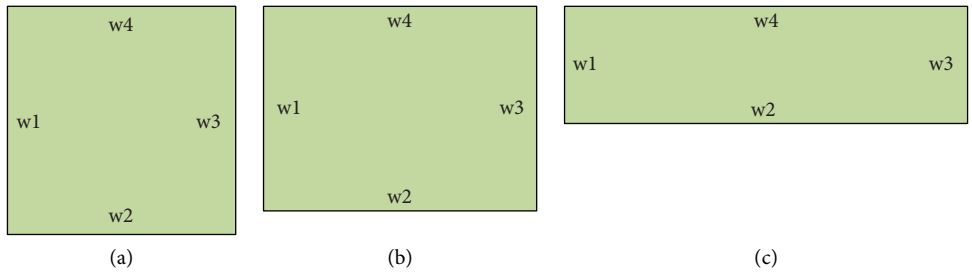


Figure 10: (a) Square room (8 ∗ 8). (b) Rectangle room (8 ∗ 6). (c) Pipeline room (10 ∗ 4).

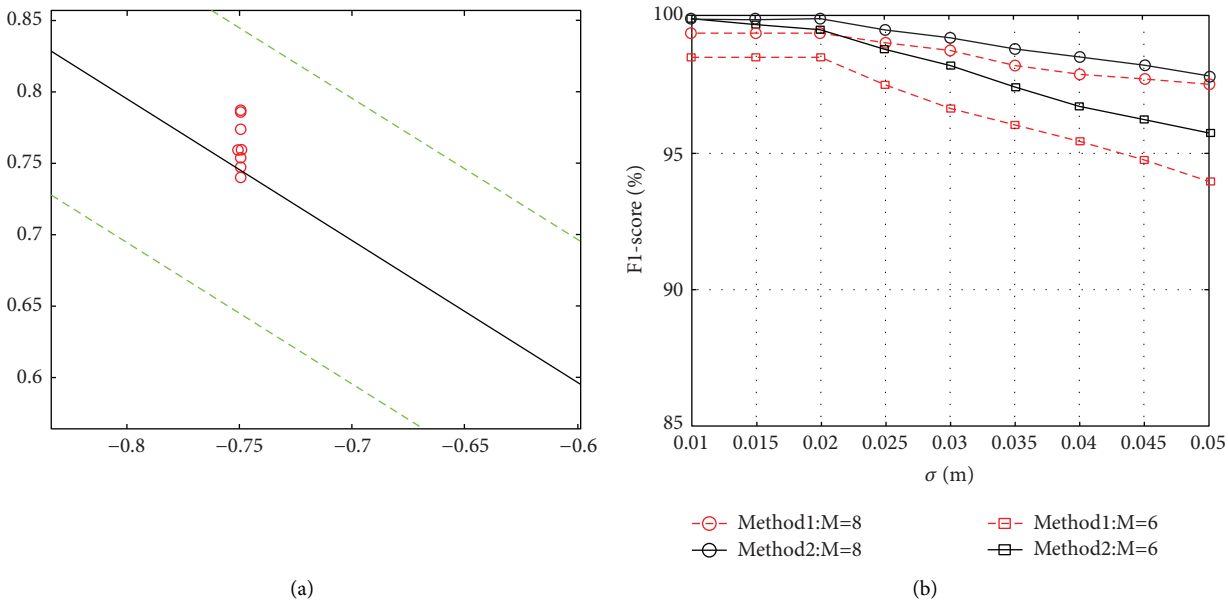


Figure 11: (a) Schematic diagram of the RIR cepstrum echo markers. The red circle is the RIR cepstrum, and the green dashed line is the upper and lower boundaries of the RIR cepstrum. (b) Comparison of the F1-scores of method 1 and method 2.

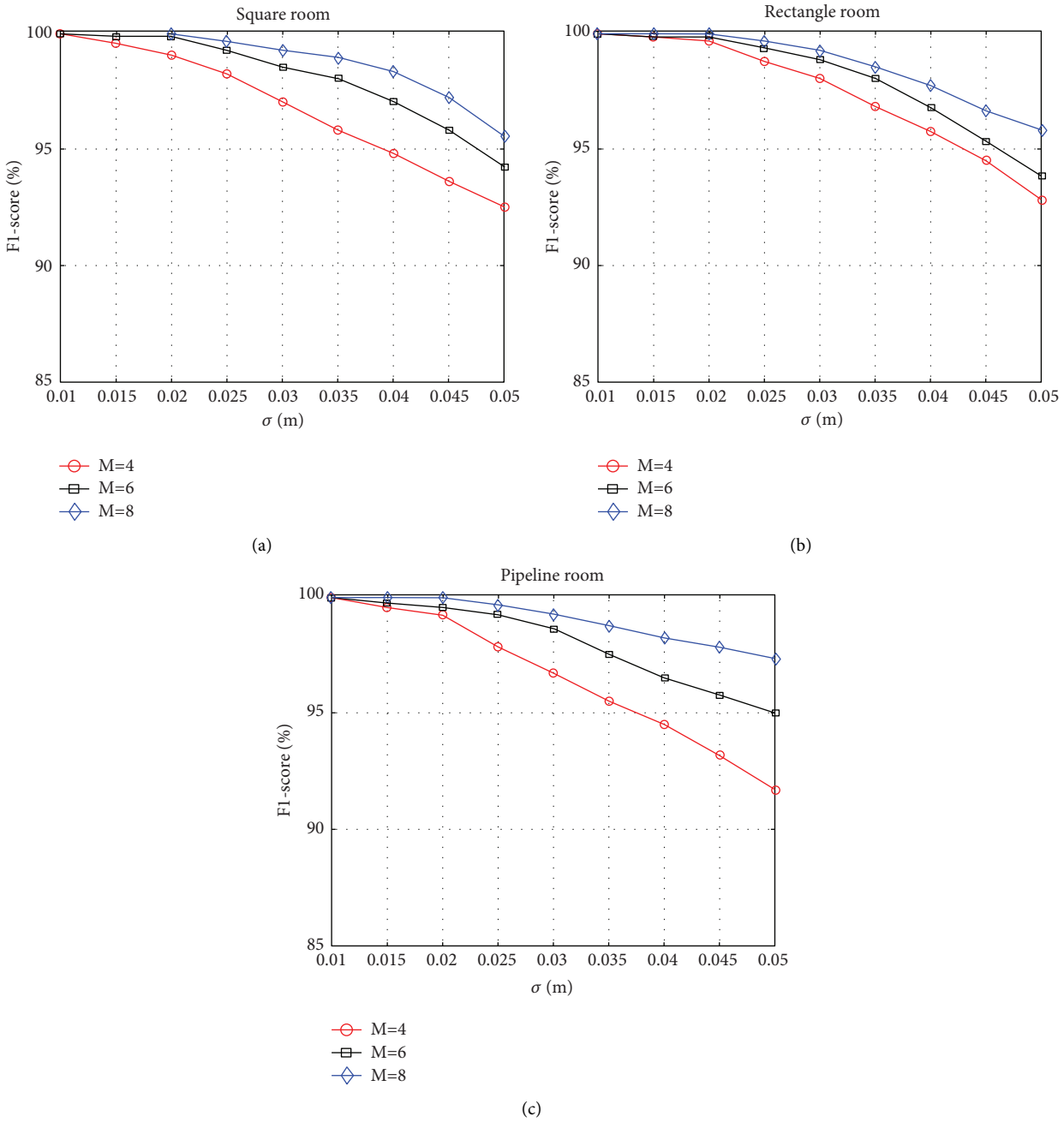


Figure 12: F1-score results for different σ values: (a) square room, (b) rectangle room, and (c) pipeline room.

shown in Figure 12. The experiments show that the F1-score can be maintained above 95% in all three rooms under low noise conditions ($\sigma \leq 0.03$ m).

4.2.2. Simulation of Pose Correction and Room Shape Estimation. Simulation experiments were conducted on robot self-localization and room position estimation in a room with wooden door and glass windows of size 7 m ∗ 7 m. Among them, the sound reflection coefficient of wooden door is 0.7, the sound reflection coefficient of glass window is 0.8, and the sound reflection coefficient of wall is 0.9. The robot's microphone array is a four-channel microphone array with a radius of 0.2 m. The robot travels around the room along the wall, and every 0.4 m, the robot actively emits sound and simulates the RIR of the current position ($\sigma = 0.05$ m).

The blue diamond line in Figure 13(b) shows the trajectory of the robot without pose correction, and the green line is the closed-loop result detected by the method in this paper. Figure 13(c) shows the trajectory after the pose correction based on the graph optimization, and the optimized path trajectory is basically consistent with the real trajectory. Figure 13(d) shows the location of the reflector estimated based on the optimized path, where the green "×" is the estimated location of the wall, the red "×" is the

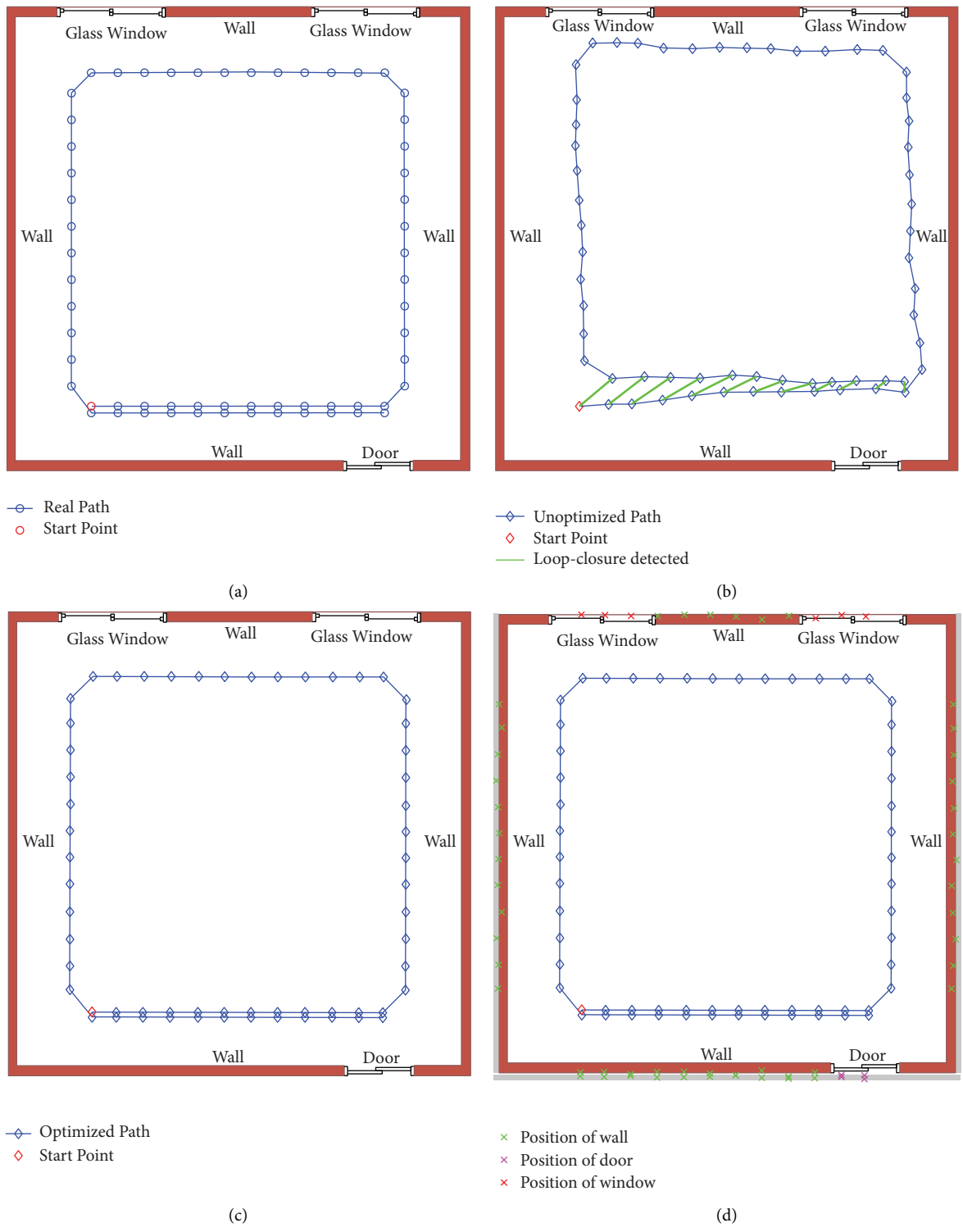


FIGURE 13: (a) Real room contour and real trajectory of the robot, (b) unoptimized robot trajectory, (c) optimized robot trajectory, and (d) estimated wall position based on the optimized trajectory.

location of the glass, and the pink "×" is the location of the door.

The position error $\text{Err} = \sqrt{X_{\text{err}}^2 + Y_{\text{err}}^2}$ was used to measure the robot's self-positioning error and mapping error, where X_{err} is the X-axis coordinate error and Y_{err} is the Y-axis coordinate error. Figure 14(a) shows the average self-localization error and mapping error statistics of the robot traveling some of the position points according to the route in Figure 13(a). The self-positioning error of the robot is less than 3.18 cm with 60% probability, and the average mapping error is less than 4.86 cm with 58% probability.

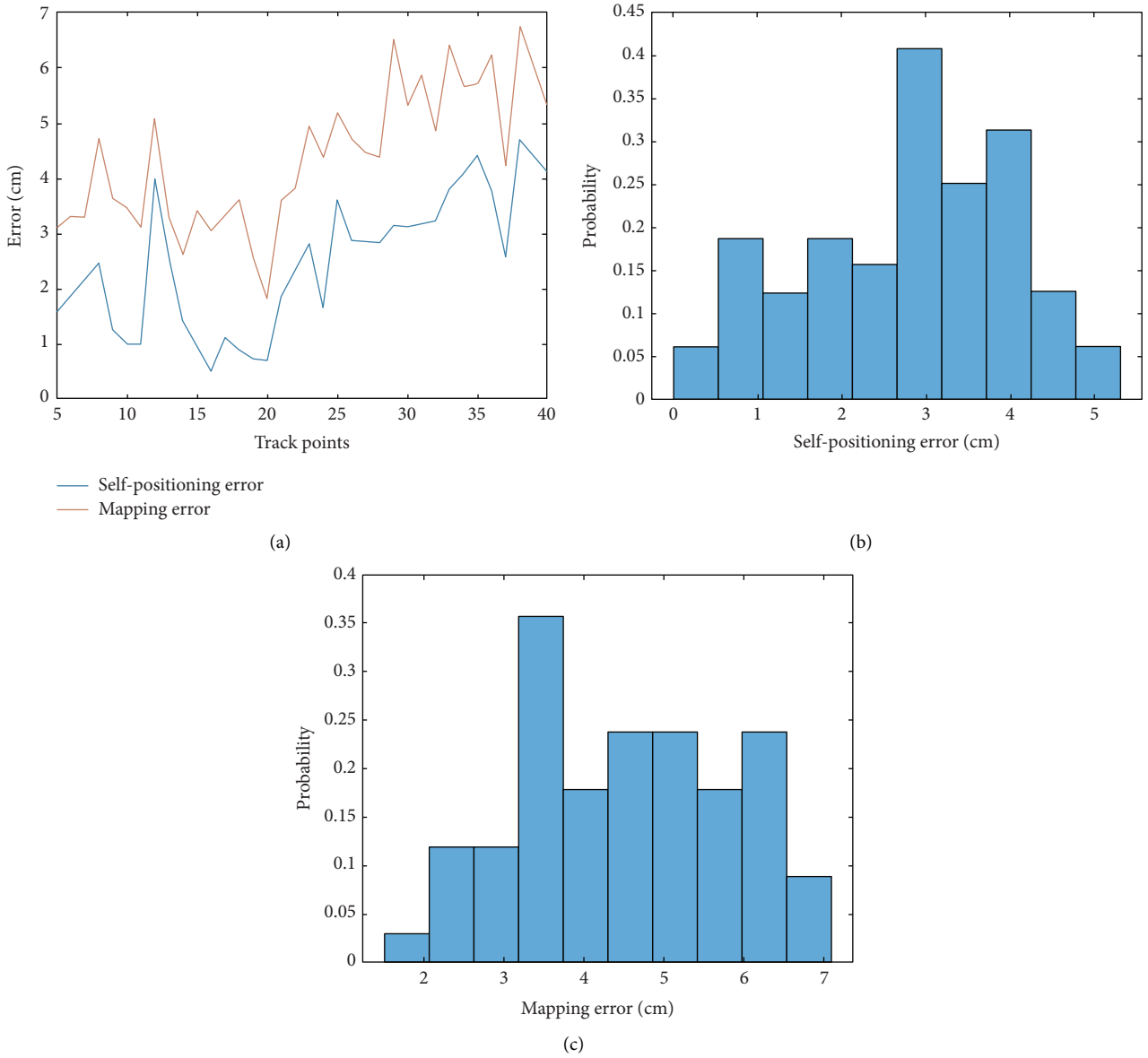


FIGURE 14: (a) Statistics of robot self-positioning error and reflected wall position estimation error. (b) Histogram of probability distribution of self-positioning error. (c) Histogram of probability distribution of mapping error.

The comparison experiments between the method in this paper and the method based on KEF filtering was done, in which the robot drove around the room randomly once in the above room and simulated 100 Monte Carlo experiments, respectively. The average self-localization error and mapping error of the experiments are shown in Table 1, where method 1 is the method without path optimization, method 2 is the method of path optimization by KEF filtering, and method 3 is the method of this paper.

4.2.3. Real Room Experiments. To verify the stability of our own designed robot and the practical performance of the method in this paper, the experiments were conducted in a real room of $4.3\,\text{m} \times 5.5\,\text{m} \times 3\,\text{m}$ (room dimensions were obtained using total station measurements). The total station was placed at the doorway and was used to measure the actual position of the robot as well as the actual position of the walls. The positions of the total station and the robot in the room are shown in Figure 15(a). Due to the height limitation of the robot, the robot can only measure the reflected wall under the red line in the right figure in Figure 15(a). The robot travels around the room close to the wall, and every time it moves, the robot actively vocalizes once (moving distance is less than 0.5 m) and records the RIR of the current position. Every time the robot moves during the experiment, the real position of the robot is measured with the total station and recorded. The RIRs obtained by the four microphones are shown in Figure 15(b) (the ambient temperature of the experiment is 30 degrees, and the corresponding sound speed is 349.75 m/s). Red marker points are first-order echoes from the wall, and red marker points of the same shape are first-order echoes from the same wall.

Table 1: Position error simulation results.

	Self-positioning error (cm)	Mapping error (cm)
Method 1	24.5	25.8
Method 2	3.01	4.53
Method 3	2.78	4.38

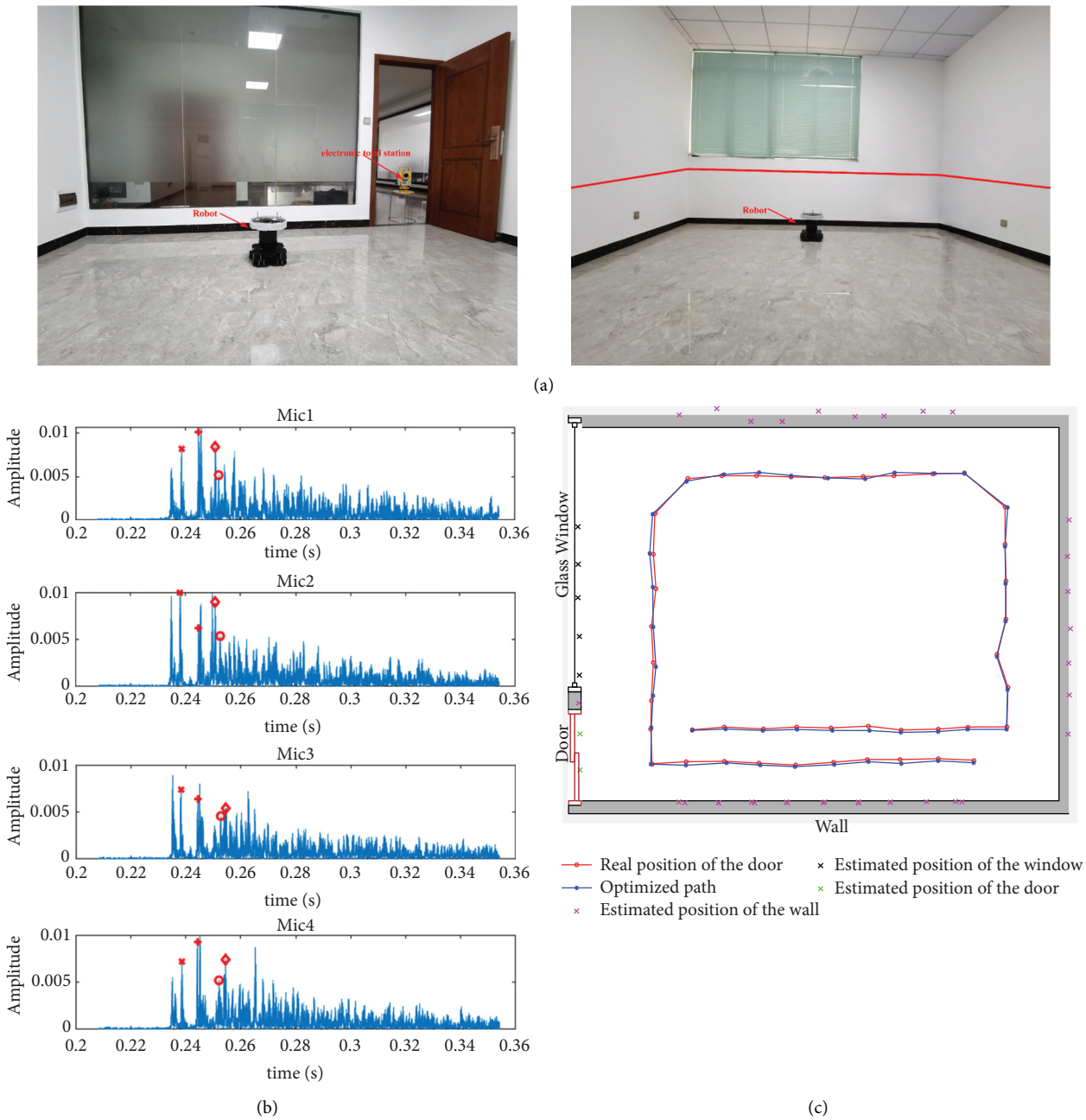


Figure 15: Experimental setup and results: (a) experimental room scene, (b) extracted first-order echo marker results in RIR, (c) experimental results graph.

Since the sound insulation panels were added between the speaker and the microphone receiver board and the sound propagation direction of the speaker is 360 degrees in the horizontal direction, there is no first-order reflection echo from the upper and lower walls. The final experimental results are shown in 2D, as shown in Figure 15(c). The

overall average self-positioning error of the robot is 2.84 cm, and the average mapping error is 4.86 cm.

5. Conclusion

This study introduces a graph optimization-based acoustic SLAM edge computing system and a method that provide new ideas for the solution of the acoustic SLAM problem. Based on the solution in this study, the robot can use acoustic signals to achieve self-localization and centimeter-level room map construction services containing door and window information. The current method in this paper has better performance in an empty room. In the future, acoustic SLAM research will be conducted in more complex indoor spaces.

Data Availability

The simulation and experimental data used to support the findings of this study have not been made available because this paper is funded by the Guangxi Science and Technology Plan Project (No. AD18281044). The grant is still in the research phase, and all research data are currently restricted to disclose within the project team.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research work was funded by the Guangxi Science and Technology Plan Project (Nos. AD18281044 and AD18281020), the Guangxi Keypoint Research and Invention Program (No. AB18221011), the Dean Project of Key Laboratory of Cognitive Radio and Information Processing, Ministry of Education (Nos. CRKL190104 and CRKL200107), and the Innovation Project of Guangxi Graduate Education (No. 2020YCXS024).

References

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[2] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, "An overview to visual o and visual SLAM: applications to mobile robotics," *Intelligent Industrial Systems*, vol. 1, no. 4, pp. 289–311, 2015.

[3] R. Frikha, R. Ejbali, and M. Zaied, "Camera pose estimation for augmented reality in a small indoor dynamic scene," *Journal of Electronic Imaging*, vol. 26, no. 5, Article ID 053029, 2017.

[4] C. Häne, L. Heng, G. H. Lee et al., "3D visual perception for self-driving cars using a multi-camera system: calibration, mapping, localization, and obstacle detection," *Image and Vision Computing*, vol. 68, pp. 14–27, 2017.

[5] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," in *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Stockholm, Sweden, May 2016.

[6] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, 2007.

[7] J. O'Reilly, S. Cirstea, M. Cirstea, and J. Zhang, "A novel development of acoustic SLAM," in *Proceedings of the 2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, pp. 525–531, IEEE, Istanbul, Turkey, August 2019.

[8] M. Crocco, A. Trucco, and A. Del Bue, "Room reflectors estimation from sound by greedy iterative approach," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6877–6881, IEEE, Calgary, Canada, April 2018.

[9] I. Dokmanic, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12186–12191, 2013.

[10] I. Dokmanić, L. Daudet, and M. Vetterli, "From acoustic room reconstruction to slam," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6345–6349, IEEE, Shanghai, China, March 2016.

[11] I. Jager, R. Heusdens, and N. D. Gaubitch, "Room geometry estimation from acoustic echoes using graph-based echo labeling," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, Shanghai, China, March 2016.

[12] M. Coutino, M. B. Møller, J. K. Nielsen, and R. Heusdens, "Greedy alternative for room geometry estimation from acoustic echoes: a subspace-based method," in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 366–370, IEEE, New Orleans, LA, USA, March 2017.

[13] F. Antonacci, J. Filos, M. R. P. Thomas et al., "Inference of room geometry from acoustic impulse responses," *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 10, pp. 2683–2695, 2012.

[14] X. Alameda-Pineda and R. Horaud, "A geometric approach to sound source localization from time-delay estimates," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 6, pp. 1082–1095, 2014.

[15] S. Park and J.-W. Choi, "Iterative echo labeling algorithm with convex hull expansion for room geometry estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1463–1478, 2021.

[16] C. Evers, A. H. Moore, and P. A. Naylor, "Acoustic simultaneous localization and mapping (a-SLAM) of a moving microphone array and its surrounding speakers," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6–10, IEEE, Shanghai, China, March 2016.

[17] C. Evers and P. A. Naylor, "Acoustic SLAM," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 9, pp. 1484–1498, 2018.

[18] M. Krekovic, I. Dokmanic, and M. Vetterli, "EchoSLAM: simultaneous localization and mapping with acoustic echoes," in *Proceedings of the IEEE International Conference on Acoustics*, March 2016.

[19] A. H. Moore, M. Brookes, and P. A. Naylor, "Room geometry estimation from a single channel acoustic impulse response," in *Proceedings of the Signal Processing Conference*, pp. 1–5, IEEE, Uttar Pradesh, India, April 2014.

[20] V. Maya, N. Yair, and S. Gannot, "The hybrid Cramér-Rao lower bound for simultaneous self-localization and room geometry estimation," *EURASIP Journal on Applied Signal Processing*, vol. 2021, no. 1, pp. 1–22, 2021.

[21] L. Nguyen, J. V. Miro, and X. Qiu, "Can a robot hear the shape and dimensions of a room," 2019, https://arxiv.org/abs/1907.01169.

[22] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small - room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[23] J. Borish, "Extension of the image model to arbitrary polyhedra," *Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, 1984.

[24] H. Guo, M. Li, X. Zhang, Q. Liu, and X. Gao, "Research on indoor wireless positioning precision optimization based on UWB," *Journal of Web Engineering (JWE)*, pp. 94–116, 2020.

[25] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard, "A Tutorial on Graph-Based Slam," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2011.

[26] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: a general framework for graph optimization," in *Proceedings of the IEEE International Conference on Robotics & Automation*, IEEE, Shanghai, China, May 2011.

[27] C. Cadena, L. Carlone, H. Carrillo et al., "Past, present, and future of simultaneous localization and mapping: toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[28] K. Imoto and N. Ono, "Spatial-feature-based acoustic scene analysis using distributed microphone array," in *Proceedings of the European Signal Processing Conference (EUSIPCO) 2015*, September 2015.

[29] X. Song, M. Wang, H. Qiu, and L Luo, "Indoor pedestrian self-positioning based on image acoustic source impulse using a sensor-rich smartphone," *Sensors*, vol. 18, no. 12, 2018.