

Radio Point Cloud-based Positioning and Mapping

Kyeong-Ju Cha*, Hyunwoo Park*, Soonmin Hwang† and Sunwoo Kim*

*Department of Electronic Engineering, Hanyang University

†Department of Automotive Engineering, Hanyang University

{lovelyckj, stark95, soonminh, remero}@hanyang.ac.kr

Abstract—This paper proposes a point cloud-based radio positioning and mapping algorithm that enables geometry-aware reconstruction from channel impulse responses (CIRs). Traditional radio simultaneous localization and mapping (SLAM) algorithms rely on predefined virtual anchors or scattering models, which restrict their ability to represent arbitrary geometries. The proposed algorithm directly infers reflection points from multi-antenna CIR measurements, aligns them via a UE-to-global transform, and aggregates their global positions over time to construct a flexible radio point cloud map. Simulation results in a simplified urban scenario show 0.4 m UE localization RMSE and 0.25 m mapping Chamfer distance, confirming that the point cloud-based approach enables geometry-aware radio sensing.

Index Terms—Radio point cloud, positioning and mapping, integrated sensing and communication.

I. INTRODUCTION

Integrated sensing and communication (ISAC) systems require fine-grained sensing capabilities beyond simple obstacle detection, in order to support environment-aware services [1], [2]. A central example of such capabilities is simultaneous localization and mapping (SLAM), which jointly estimates the user’s position and a geometric representation of the surrounding environment [3]. Most existing SLAM methods rely on vision or LiDAR sensors to construct dense point cloud maps, but their performance degrades in featureless, reflective, or transparent environments [4]. Motivated by the robustness of radio propagation to illumination and texture variations, radio SLAM has been investigated within the ISAC framework, leveraging multipath channel measurements to extract geometric information from existing wireless infrastructure [5].

Radio SLAM methods have been developed to improve mapping accuracy while reducing computational cost. Random finite set (RFS)-based approaches [6]–[9] jointly estimate propagation landmarks as multi-object states while handling association uncertainty. More recent methods include vector-type belief propagation (BP) SLAM [10], which requires particle filtering at considerable computational cost, and set-type BP SLAM [11], which adopts a set-theoretic message-passing formulation but remains sensitive to noise accumulation. An end-to-end learning-based radio SLAM [12] processes raw channel impulse responses (CIRs) with a Transformer encoder, avoiding explicit channel-parameter estimation. Despite these architectural differences, existing methods still represent the environment as a discrete set of predefined propagation

This work was supported by the National Research Foundation of Korea under Grant RS-2024-0040942.

landmarks, which biases the map toward piecewise-planar structures and makes curved or highly irregular geometries difficult to capture.

This paper proposes a radio point cloud-based positioning and mapping algorithm that does not rely on predefined landmark types. Unlike existing radio SLAM methods, the proposed approach uses a Transformer-based sensing network that takes multi-antenna CIR measurements as input and directly estimates ranges to the base station (BS) and dominant reflection points (RPs) in a UE-centric coordinate frame. From these direction-wise range estimates, UE-centric BS and RP positions are reconstructed and then aligned with the known BS location to recover the UE trajectory and transform all RP estimates into the global coordinate system. By aggregating the globally transformed RP estimates over time, the method constructs a radio point cloud that flexibly captures arbitrary environmental geometries without explicit assumptions on walls, corners, or scattering structures, enabling geometry-aware mapping even in complex urban environments.

II. SYSTEM MODEL

Consider a scenario with a single BS and a UE for positioning and mapping of the surrounding environment. The UE, equipped with a 2×2 antenna array, is mounted on a vehicle. The vehicle moves along a predefined straight-line trajectory, and the position of the vehicle at time step t is represented as $\mathbf{p}_{\text{UE},t} = [x_{\text{UE},t} \ y_{\text{UE},t}]^\top$. In contrast, the BS is fixed over time, and its position is denoted by $\mathbf{p}_{\text{BS}} = [x_{\text{BS}} \ y_{\text{BS}}]^\top$.

Signals transmitted from the BS can either reach the UE directly or after being reflected by surrounding objects such as walls, buildings, vehicles, or people. The physical locations on these objects where the signal bounces are referred to as RPs. Assuming that N dominant reflection points exist around the BS–UE link, the propagation between the BS and UE consists of $N + 1$ paths. The first path ($n = 0$) corresponds to the line-of-sight (LoS) path, and the rest of the paths ($n = 1, 2, \dots, N$) are non-line-of-sight (NLoS) multipaths associated with reflections at the RPs. The position of the n -th reflection point is denoted by

$$\mathbf{p}_{\text{RP},n,t} = [x_{\text{RP},n,t} \ y_{\text{RP},n,t}]^\top. \quad (1)$$

The multipath propagation characteristics determine the wireless channel observed at the UE. Assuming a band-limited

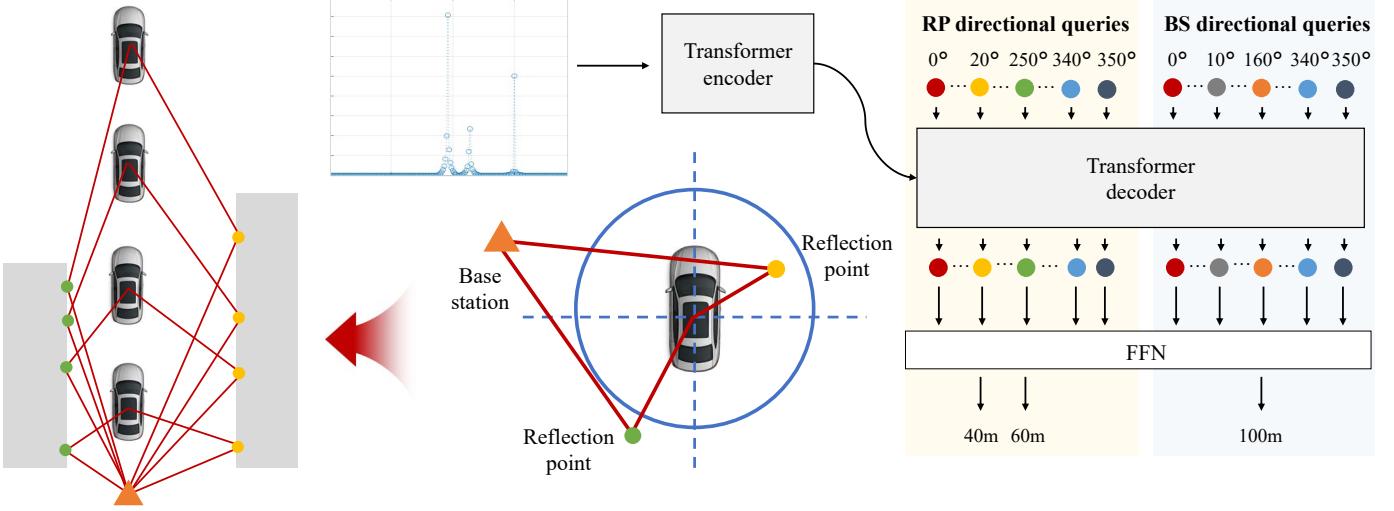


Fig. 1: Radio positioning and mapping algorithm architecture.

system with bandwidth B , the continuous-time CIR at time step t is modeled as [13]

$$h_t(z) = \sum_{n=0}^N \alpha_{n,t} \operatorname{sinc}[B(z - \tau_{n,t})] \quad (2)$$

where $\alpha_{n,t}$ and $\tau_{n,t}$ represent the amplitude and delay of the n -th path at time step t , respectively. By sampling $h_t(z)$ at Z points, the discrete-time CIR is obtained as

$$\mathbf{c}_t = [h_t(0) \dots h_t(Z-1)]^\top. \quad (3)$$

III. RADIO POSITIONING AND MAPPING ALGORITHM

The proposed radio positioning and mapping algorithm operates through a three-stage architecture, as shown in Fig. 1. Each stage performs specialized processing to estimate the UE position and construct a radio point cloud map of the surrounding environment.

A. UE-centric Radio Sensing Network

At each time step t , the network takes the multi-antenna CIR from the 2×2 antenna array as input and estimates the BS and RP positions in the UE coordinate.

1) *Input Embedding*: Let $\mathbf{c}_{t,m} \in \mathbb{R}^Z$ denote the discrete-time CIR measured at the m -th antenna element of the UE at time step t . These four CIRs are stacked column-wise to form a multi-antenna CIR matrix

$$\mathbf{C}_t = [\mathbf{c}_{t,1} \mathbf{c}_{t,2} \mathbf{c}_{t,3} \mathbf{c}_{t,4}] \in \mathbb{R}^{Z \times 4} \quad (4)$$

where each column corresponds to one receive antenna. Each row of \mathbf{C}_t is linearly embedded into a dimensional feature vector D by

$$\mathbf{X}_{e,t} = \mathbf{C}_t \mathbf{W}_c \in \mathbb{R}^{Z \times D} \quad (5)$$

where $\mathbf{W}_c \in \mathbb{R}^{4 \times D}$ is the embedding matrix. By incorporating the positional embedding, $\mathbf{X}_t \in \mathbb{R}^{Z \times D}$ is obtained as

$$\mathbf{X}_t = \mathbf{X}_{e,t} + \mathbf{X}_{pe} \quad (6)$$

where $\mathbf{X}_{pe} \in \mathbb{R}^{Z \times D}$ is the positional embedding matrix. The matrix $\mathbf{X}_t \in \mathbb{R}^{Z \times D}$ serves as the input to the transformer encoder.

2) *Transformer Encoder*: The transformer encoder models the dependencies within the embedded multi-antenna CIR. It consists of L stacked layers, each containing a multi-head self-attention (MSA) block followed by a feed-forward network. In the l -th layer, the input is first projected into query(\mathbf{Q}_h), key(\mathbf{K}_h), and value(\mathbf{V}_h) matrices for each of the H attention heads:

$$\mathbf{Q}_h = \mathbf{X}_t^{(l)} \mathbf{W}_h^Q, \quad \mathbf{K}_h = \mathbf{X}_t^{(l)} \mathbf{W}_h^K, \quad \mathbf{V}_h = \mathbf{X}_t^{(l)} \mathbf{W}_h^V$$

where $\mathbf{W}_h^Q, \mathbf{W}_h^K, \mathbf{W}_h^V$ are learnable projection matrices. For each head, the attention is computed as

$$\operatorname{Attention}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \operatorname{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^\top}{\sqrt{A}}\right) \mathbf{V}_h \quad (7)$$

where A denotes the dimensionality of the queries and keys. The outputs from all H heads are concatenated and linearly projected to produce a unified representation, which is then passed through a feed-forward network. Residual connections and layer normalization are applied around both the MSA and feed-forward blocks to stabilize training. The final encoder output is denoted by \mathbf{E}_t , and serves as a compact representation of the CIR.

3) *Directional Query Decoder*: The decoder adopts a set of direction-wise queries to estimate ranges towards the BS and reflection points in the UE-centric frame. The azimuth domain is uniformly quantized into $M = 36$ directions with a 10° spacing $\Theta = \{\theta_1, \theta_2, \dots, \theta_M\} = \{0^\circ, 10^\circ, \dots, 350^\circ\}$. For each direction $\theta_i \in \Theta$, two learnable query vectors are defined: one associated with the BS and one associated with a RP $\mathbf{q}_{BS,i}, \mathbf{q}_{RP,i} \in \mathbb{R}^D$. Collecting all queries yields the decoder input $\mathbf{Q}_{dec} = [\mathbf{q}_{BS,1}^\top \dots \mathbf{q}_{BS,M}^\top \mathbf{q}_{RP,1}^\top \dots \mathbf{q}_{RP,M}^\top]^\top \in \mathbb{R}^{2M \times D}$. Given the encoder output $\mathbf{E}_t \in \mathbb{R}^{Z \times D}$, the trans-

former decoder applies cross-attention over \mathbf{E}_t with \mathbf{Q}_{dec} as queries and produces a set of latent vectors

$$\mathbf{H}_t = \text{Decoder}(\mathbf{Q}_{\text{dec}}, \mathbf{E}_t) \in \mathbb{R}^{2M \times D} \quad (8)$$

where the first M rows correspond to BS-related features and the remaining M rows correspond to RP-related features. The i -th BS and RP feature vectors at time step t are denoted by

$$\mathbf{h}_{\text{BS},t}(\theta_i) = \mathbf{H}_t(i, :), \quad \mathbf{h}_{\text{RP},t}(\theta_i) = \mathbf{H}_t(M+i, :). \quad (9)$$

4) *Range Regression and Masking*: Each direction-wise feature vector is mapped to a scalar range estimate through a regression head. For direction $\theta_i \in \Theta$, the BS and RP ranges from the UE are estimated as

$$\hat{r}_{\text{BS},t}(\theta_i) = \mathbf{w}_{\text{BS}}^\top \mathbf{h}_{\text{BS},t}(\theta_i), \quad \hat{r}_{\text{RP},t}(\theta_i) = \mathbf{w}_{\text{RP}}^\top \mathbf{h}_{\text{RP},t}(\theta_i) \quad (10)$$

where $\mathbf{w}_{\text{BS}}, \mathbf{w}_{\text{RP}} \in \mathbb{R}^D$ are learnable regression vectors. Not all directions contain a valid BS or RP path. For directions where no BS or RP is present, binary masks $m_{\text{BS},t}(\theta_i), m_{\text{RP},t}(\theta_i) \in \{0, 1\}$ are used to exclude the corresponding terms from the regression loss, so that only directions with a valid path contribute to the training objective.

5) *UE-Centric Position Reconstruction*: Once the direction-wise ranges are obtained, BS and RP positions are reconstructed only for directions with valid paths. Let $\Theta_t^{\text{BS}} = \{\theta_i \in \Theta \mid m_{\text{BS},t}(\theta_i) = 1\}$ and $\Theta_t^{\text{RP}} = \{\theta_i \in \Theta \mid m_{\text{RP},t}(\theta_i) = 1\}$ denote the sets of directions with valid BS and RP paths at time t , respectively. In the considered scenario, Θ_t^{BS} contains a single direction θ_i^* corresponding to the BS. The UE-centric BS position is then given by

$$\hat{\mathbf{p}}_{\text{BS},t}^{\text{UE}} = \begin{bmatrix} \hat{r}_{\text{BS},t}(\theta_i^*) \cos \theta_i^* \\ \hat{r}_{\text{BS},t}(\theta_i^*) \sin \theta_i^* \end{bmatrix}. \quad (11)$$

Similarly, UE-centric RP positions are reconstructed for all $\theta_i \in \Theta_t^{\text{RP}}$ as

$$\hat{\mathbf{p}}_{\text{RP},n,t}^{\text{UE}} = \hat{\mathbf{p}}_{\text{RP},t}^{\text{UE}}(\theta_i) = \begin{bmatrix} \hat{r}_{\text{RP},t}(\theta_i) \cos \theta_i \\ \hat{r}_{\text{RP},t}(\theta_i) \sin \theta_i \end{bmatrix}. \quad (12)$$

B. Global Coordinate Transformation

This subsection aligns the estimated UE-centric BS position with the known global BS location to obtain the UE position and to transform all estimates into the global coordinate system. Since both \mathbf{p}_{BS} and $\hat{\mathbf{p}}_{\text{BS},t}^{\text{UE}}$ represent the same physical BS expressed in different coordinate frames, their alignment determines the UE position. Given the UE heading $\mathbf{R}_t \in \mathbb{R}^{2 \times 2}$, the UE position in the global frame is obtained by

$$\hat{\mathbf{p}}_{\text{UE},t} = \mathbf{p}_{\text{BS}} - \mathbf{R}_t \hat{\mathbf{p}}_{\text{BS},t}^{\text{UE}}. \quad (13)$$

The same rigid transform is applied to the UE-centric RP estimates. The corresponding global RP position is given by

$$\hat{\mathbf{p}}_{\text{RP},n,t} = \mathbf{R}_t \hat{\mathbf{p}}_{\text{RP},n,t}^{\text{UE}} + \hat{\mathbf{p}}_{\text{UE},t}. \quad (14)$$

By applying this transform jointly to the UE-centric BS and RP estimates, the absolute BS and RP positions are obtained in the global coordinate system.

TABLE I: Simulation Parameters

Parameter	Value
Input dimension	400×4
Hidden dimension	256
Attention heads	8
Encoder layers	4
Decoder layers	4
Feedforward dimension	2048
Output activation	Sigmoid
Optimizer	AdamW
Learning rate	5×10^{-4}
Batch size	128
Epochs	200
Loss function	MSE + empty penalty

C. Radio Point Cloud Mapping

After transforming all estimated BS and RP positions into the global coordinate system, the environment can be reconstructed by aggregating the global RP estimates across time. Let $\hat{\mathcal{P}}_t = \{\hat{\mathbf{p}}_{\text{RP},n,t} \mid n \in \mathcal{N}_t\}$ denote the set of RP positions detected at time t , where \mathcal{N}_t is the index set of valid RP paths. As the UE moves along its trajectory, each timestep contributes a new set of RP points whose union forms the radio point cloud map

$$\hat{\mathcal{M}} = \bigcup_{t=1}^T \hat{\mathcal{P}}_t. \quad (15)$$

Because each RP corresponds to a surface interaction point between the transmitted signal and the environment, the accumulated point cloud implicitly captures the geometry of surrounding structures such as walls and scatterers.

IV. CASE STUDIES

A simplified 2D urban-like scenario is considered. The environment consists of one BS and two buildings that serve as dominant reflectors. The BS is located at $\mathbf{p}_{\text{BS}} = [0 \text{ m} \ 100 \text{ m}]^\top$. Building 1 is modeled as a vertical line segment around $x = -45 \text{ m}$ and Building 2 around $x = 40 \text{ m}$. RPs are placed along these buildings at positions corresponding to 10° angular intervals from the UE perspective. A single UE moves along a straight trajectory on the y -axis, beginning at $[0 \text{ m} \ 50 \text{ m}]^\top$ and ending at $[0 \text{ m} \ 10 \text{ m}]^\top$, and a total of 40 UE positions are sampled uniformly along this path.

The carrier frequency, sample rate, and bandwidth are set to 28 GHz, 245.76 MHz, and 100 MHz, respectively. The simulation parameters are shown in Table I. The network is trained end-to-end using a combined loss function that consists of regression losses and empty penalties:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BS}} + \mathcal{L}_{\text{RP}} + \lambda(\mathcal{L}_{\text{empty}}^{\text{BS}} + \mathcal{L}_{\text{empty}}^{\text{RP}}) \quad (16)$$

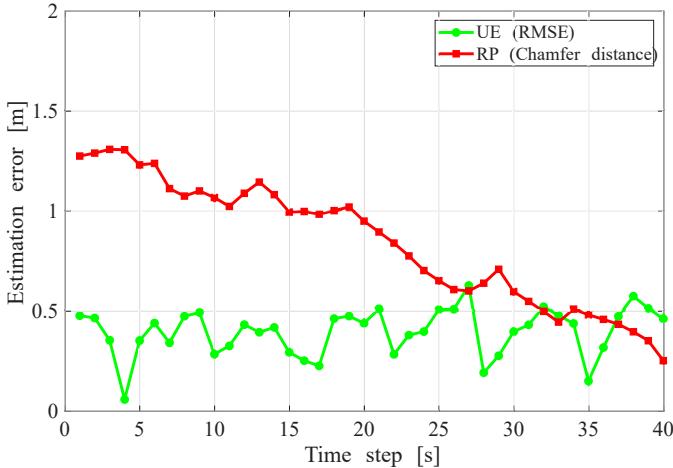


Fig. 2: UE position estimation RMSE and RP mapping Chamfer distance.

where the regression losses for BS and RP range estimates are defined as

$$\begin{aligned}\mathcal{L}_{\text{BS}} &= \frac{1}{|\Theta^{\text{BS}}|} \sum_{t=1}^T \sum_{\theta_i \in \Theta_t^{\text{BS}}} |\hat{r}_{\text{BS},t}(\theta_i) - r_{\text{BS},t}(\theta_i)|^2, \\ \mathcal{L}_{\text{RP}} &= \frac{1}{|\Theta^{\text{RP}}|} \sum_{t=1}^T \sum_{\theta_i \in \Theta_t^{\text{RP}}} |\hat{r}_{\text{RP},t}(\theta_i) - r_{\text{RP},t}(\theta_i)|^2\end{aligned}\quad (17)$$

where $r_{\text{BS},t}(\theta_i)$ and $r_{\text{RP},t}(\theta_i)$ denote the ground truth ranges, and Θ_t^{BS} and Θ_t^{RP} represent the sets of valid directions for BS and RP at time t , respectively. The empty penalties suppress predictions in invalid directions:

$$\begin{aligned}\mathcal{L}_{\text{empty}}^{\text{BS}} &= \frac{1}{|\bar{\Theta}_t^{\text{BS}}|} \sum_{t=1}^T \sum_{\theta_i \in \bar{\Theta}_t^{\text{BS}}} \hat{r}_{\text{BS},t}(\theta_i)^2, \\ \mathcal{L}_{\text{empty}}^{\text{RP}} &= \frac{1}{|\bar{\Theta}_t^{\text{RP}}|} \sum_{t=1}^T \sum_{\theta_i \in \bar{\Theta}_t^{\text{RP}}} \hat{r}_{\text{RP},t}(\theta_i)^2\end{aligned}\quad (18)$$

where $\bar{\Theta}_t^{\text{BS}}$ and $\bar{\Theta}_t^{\text{RP}}$ denote the sets of empty directions at time t . The empty penalty weight is set to $\lambda = 0.5$.

Positioning and mapping performance are evaluated using root mean square error (RMSE) and Chamfer distance [14], respectively. The Chamfer distance quantifies the geometric dissimilarity between two point sets by measuring the average nearest-neighbor distance in both directions. For the ground truth RP map $\mathcal{M} = \{\mathbf{p}_{\text{RP},n,t} \mid n \in N_t, t = 1, \dots, T\}$ and the estimated RP map $\hat{\mathcal{M}}$, the Chamfer distance is defined as

$$\begin{aligned}d_{\text{CD}}(\mathcal{M}, \hat{\mathcal{M}}) &= \frac{1}{|\mathcal{M}|} \sum_{\mathbf{p} \in \mathcal{M}} \min_{\hat{\mathbf{p}} \in \hat{\mathcal{M}}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2 \\ &+ \frac{1}{|\hat{\mathcal{M}}|} \sum_{\hat{\mathbf{p}} \in \hat{\mathcal{M}}} \min_{\mathbf{p} \in \mathcal{M}} \|\hat{\mathbf{p}} - \mathbf{p}\|_2\end{aligned}\quad (19)$$

where \mathbf{p} and $\hat{\mathbf{p}}$ denote RP positions in the ground truth and estimated maps, respectively. The first term measures the

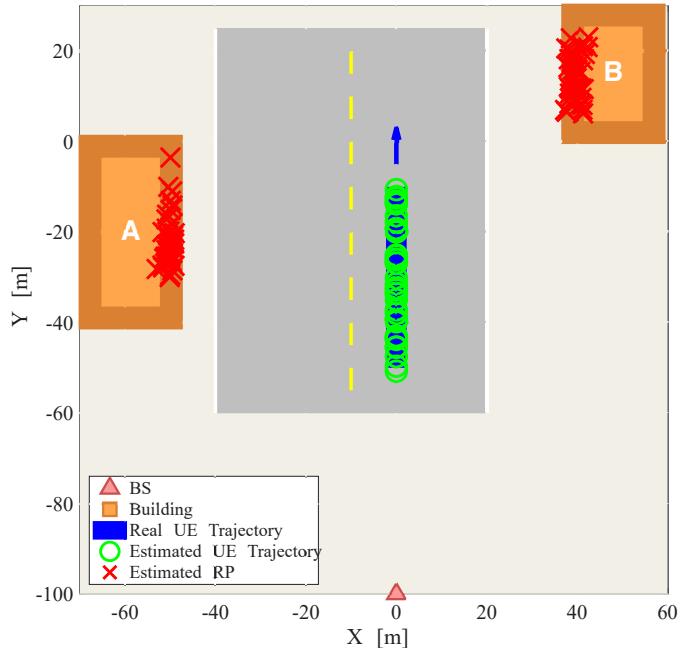


Fig. 3: 2D visualization of UE positioning and radio point cloud mapping.

average distance from ground truth points to their nearest estimated points, and the second term measures the reverse direction. This bidirectional formulation ensures robustness to both missing detections and false alarms in the estimated point cloud.

The positioning and mapping results in Fig. 2 show how the errors vary over 40 time steps. The UE localization RMSE stays within 0.1 m–0.6 m, indicating centimeter-level tracking accuracy. The RP mapping error, measured by the Chamfer distance, clearly converges from about 1.3 m to 0.25 m, validating the effectiveness of point cloud aggregation whereby multiple observations from different viewpoints progressively refine the radio map. This combination of stable localization and steadily improving mapping accuracy demonstrates that the proposed method can simultaneously achieve reliable positioning and progressive environment reconstruction.

Fig. 3 visualizes the complete positioning and mapping result, including ground truth buildings, the estimated UE trajectory, and the estimated RPs. The estimated trajectory closely aligns with the ground truth, with an average error of 0.4 m, while the estimated RPs form two distinct vertical clusters that accurately represent the geometry of Buildings A and B. Although individual RP estimates exhibit some scatter due to measurement noise and the 10° angular quantization, the aggregated point cloud successfully captures the building structure, achieving a final Chamfer distance of 0.25 m. These results highlight the flexibility of the proposed point cloud mapping, which captures building geometry directly from CIRs rather than through sparse landmark sets or simplified geometric models.

V. CONCLUSIONS

This paper presents a point cloud-based radio positioning and mapping algorithm that directly estimates RP clouds from multi-antenna CIR measurements. By combining a Transformer-based sensing network with coordinate alignment using known BS positions, the method jointly performs UE localization and global point cloud map construction. Simulations in a simplified urban scenario demonstrate a UE localization RMSE of about 0.04 m and mapping performance converging to 0.025 m Chamfer distance, validating that environmental geometry can be reconstructed directly from CIRs without predefined landmark types or explicit wall and corner models. Future work will extend the framework to multi-BS scenarios and more complex urban environments.

REFERENCES

- [1] A. K. et al., “Toward Integrated Sensing and Communications for 6G: Key Enabling Technologies, Standardization, and Challenges,” *IEEE Commun. Stand. Mag.*, vol. 8, no. 2, pp. 52–59, 2024.
- [2] Y. Zeng and X. Xu, “Toward Environment-Aware 6G Communications via Channel Knowledge Map,” *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 84–91, 2021.
- [3] H. Q. et al., “Cooperative Mapping, Localization, and Beam Management via Multi-Modal SLAM in ISAC Systems,” *IEEE Trans. Commun.*, 2025.
- [4] S. Pyykölä, N. Joswig, and L. Ruotsalainen, “Non-Lambertian surfaces and their challenges for visual SLAM,” *IEEE Open Comput. Soc.*, vol. 5, pp. 430–445, 2024.
- [5] B. A. et al., “Radio SLAM: A review on radio-based simultaneous localization and mapping,” *IEEE Access*, vol. 11, pp. 9260–9278, 2023.
- [6] H. W. et al., “5G mmWave cooperative positioning and mapping using multi-model PHD filter and map fusion,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3782–3795, 2020.
- [7] H. D. et al., “The labeled multi-Bernoulli SLAM filter,” *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1561–1565, 2015.
- [8] H. K. et al., “PMBM-based SLAM filters in 5G mmWave vehicular networks,” *IEEE Trans. Intell. Veh.*, vol. 71, no. 8, pp. 8646–8661, 2022.
- [9] Y. G. et al., “A computationally efficient EK-PMBM filter for bistatic mmWave radio SLAM,” *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2179–2192, 2022.
- [10] E. L. et al., “A belief propagation algorithm for multipath-based SLAM,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5613–5629, 2019.
- [11] H. K. et al., “Set-type belief propagation with applications to poisson multi-bernoulli slam,” *IEEE Trans. Signal Process.*, vol. 72, pp. 1989–2005, 2024.
- [12] K.-J. Cha, H. Park, W. Choe, S. Hwang, and S. Kim, “End-to-end Radio SLAM for Autonomous Systems,” in *2025 IEEE Workshop on Signal Processing Systems (SiPS)*, 2025, pp. 1–5.
- [13] P.-H. T. et al., “Ray-tracing-assisted fingerprinting based on channel impulse response measurement for indoor positioning,” *IEEE Trans. Instrum. Meas.*, vol. 66, no. 5, pp. 1032–1045, 2017.
- [14] H. G. B. et al., “Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching,” Tech. Rep., 1977.