mmWave Radar-Based Non-Line-of-Sight Pedestrian Localization at T-Junctions Utilizing Road Layout Extraction via Camera

Byeonggyu Park[1], Hee-Yeun Kim[1], Byonghyok Choi[2], Hansang Cho[2], Byungkwan Kim[3],
Soomok Lee[4], Mingu Jeon[1], and Seong-Woo Kim[1]

***Abstract*— Pedestrians Localization in Non-Line-of-Sight (NLoS) regions within urban environments poses a significant challenge for autonomous driving systems. While mmWave radar has demonstrated potential for detecting objects in such scenarios, the 2D radar point cloud (PCD) data is susceptible to distortions caused by multipath reflections, making accurate spatial inference difficult. Additionally, although camera images provide high-resolution visual information, they lack depth perception and cannot directly observe objects in NLoS regions. In this paper, we propose a novel framework that interprets radar PCD through road layout inferred from camera for localization of NLoS pedestrians. The proposed method leverages visual information from the camera to interpret 2D radar PCD, enabling spatial scene reconstruction. The effectiveness of the proposed approach is validated through experiments conducted using a radar-camera system mounted on a real vehicle. The localization performance is evaluated using a dataset collected in outdoor NLoS driving environments, demonstrating the practical applicability of the method.**

***Index Terms*— 2D radar point cloud, sensor-fusion, collision avoidance, multi-target localization, non-line-of-sight**

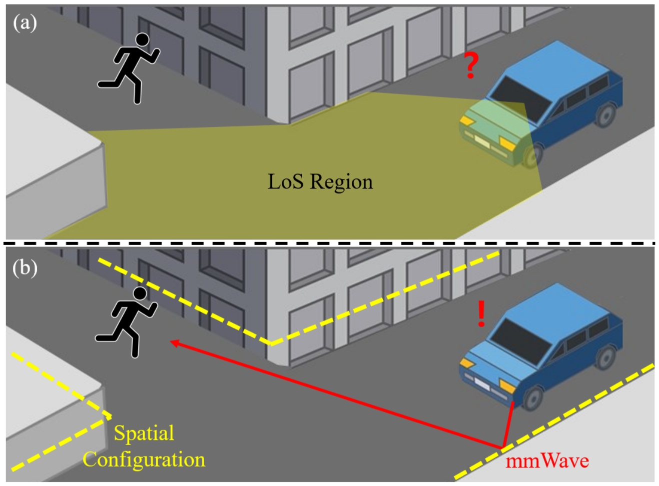

Fig. 1. **Illustration of the results using the proposed method in an intersection composed of backroads.** (a) Using only LoS sensors, it is not possible to detect objects in the NLoS region, (b) The proposed method allows for estimating the position of NLoS objects.

I. Introduction

Recent advancements in autonomous driving technology for robots and vehicles have been rapidly progressing and are being widely applied across various industrial sectors [1]. Currently, autonomous driving systems perceive the surrounding environment in real-time through various sensors, such as LiDAR, camera, and radar. In particular, LiDAR and camera are widely used for object localization and detection, operating based on Line-of-Sight (LoS) principles to recognize the visible regions of the environment.

However, autonomous driving systems that depend solely on LoS-based sensors face inherent limitations in detecting objects within NLoS regions, which are common in complex urban environments. According to the U.S. Department of Transportation, traffic accidents at intersections resulted in 12,036 fatalities in 2022, highlighting the critical challenge of limited situational awareness in such scenarios [2]. LoS sensors are incapable of detecting pedestrians or obstacles obstructed by buildings and fences at intersections, posing significant safety risks. To address this limitation, there is a need for perception methods capable of accurately recognizing pedestrians and obstacles in NLoS environments.

In response to the limitations of existing solutions, various researches have explored NLoS observation through Vehicle-to-Everything (V2X) communication as a potential solution [3]. For instance, cooperative perception is proposed as a method utilizing Vehicle-to-Vehicle (V2V) communication, wherein vehicles within a platoon exchange information about their perceived surroundings to improve safety and driving efficiency [4]–[6]. However, these solutions are often impractical in environments lacking proper infrastructure and incur substantial deployment and maintenance cost.

As a result, a more generalizable approach to NLoS perception is needed, one that involves independent observation and analysis of occluded environments. This approach requires sensors operating in frequency bands capable of reflecting and diffracting signals, such as mmWave radar. mmWave radar leverages the reflectivity to detect objects in NLoS regions and estimate their positions.

Nevertheless, using 2D radar PCD for spatial inference presents several challenges. 2D radar PCD is often sparse, noisy, and prone to positional distortion due to multi-path reflections. Additionally, interpreting 2D radar PCD requires substantial spatial information to analyze reflection paths.

This work was supported by Samsung Electro-Mechanics Co., Ltd., the National Research Foundation of Korea (NRF) through the Ministry of Science and ICT under Grant 2021R1A2C1093957, Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0020536, HRD Program for Industrial Innovation), the Korean Ministry of Land, Infrastructure and Transport (MOLIT) as the Innovative Talent Education Program for Smart City, and by the Institute of Engineering Research at Seoul National University, which provided the research facilities for this work.

[1]Seoul National University, [2]Samsung Electro-Mechanics Co., Ltd., [3]Chungnam National University, [4]Ajou University. Correspondence to: Mingu Jeon and Seong-Woo Kim {bg.park; mingujeon; snwoo}@snu.ac.kr.

Environments with multiple reflectors, such as T-junctions, pose significant challenges for precise spatial inference.

To overcome these limitations, this paper proposes the approach that integrates 2D radar PCD with front camera images to observe and analyze NLoS regions. While front camera images alone cannot provide precise 3D spatial estimates or directly capture NLoS areas, they offer high-resolution visual information that aids in analyzing road structures and layouts. By leveraging the complementary strengths of both sensors, the proposed framework of this paper infers the structure of occluded spaces and estimates pedestrian locations within NLoS regions is proposed, as shown in Fig. 1.

The proposed approach extracts road layout information from front camera images, serving as a foundation for interpreting 2D radar PCD to infer spatial details. A ray-tracing technique is then used to correct distorted positional data in the 2D radar PCD. Filtering and clustering methods remove noise and enhance spatial inference accuracy. The effectiveness of the framework is evaluated using real-world datasets from a radar-camera system mounted on a vehicle, demonstrating its ability to localize pedestrians in NLoS environments in practical driving scenarios The main contributions of this research are as follows:

- A novel framework for NLoS pedestrian localization at T-junctions using mmWave radar and road layout inferred from camera images is proposed.
- A method for interpreting 2D radar PCD based on front camera images to achieve accurate spatial inference is proposed.
- The localization performance is validated using real-world outdoor NLoS datasets from a radar-camera system mounted on a vehicle.

II. Related Works

A. Camera-based road layout estimation model

In the context of road environment perception, camera-based Bird's Eye View (BEV) transformation and road layout estimation models have garnered significant attention [7], [8]. Cameras offer a distinct advantage in extracting structures such as lanes, road boundaries, and intersections due to their high-resolution visual information. Recently, deep learning-based techniques for BEV transformation have been widely adopted in the realm of autonomous driving research. Philion *et al.* proposed the Lift-Splat-Shoot model, which processes multi-camera images to reconstruct the 3D environment and generate a BEV-style road layout [9]. This model integrates multi-view information to provide comprehensive spatial representations. Qureshi *et al.* introduced a deep learning-based approach that estimates the BEV layout of urban driving scenarios from a single image [10]. By employing adversarial feature learning and multi-channel semantic occupancy grids, this method can infer occluded regions and reconstruct a plausible scene layout in real-time. Liu *et al.* utilized a transformer-based model to estimate road layouts, including NLoS regions occluded by obstacles, from monocular camera images [11].

TABLE I
Comparison of NLoS Pedestrian Localization Methods.

Methods	Input type	Object		Sensor fusion	Reflector estimation
		NLoS	Dynamic		
Chen *et al.* [12]	Signal	✓			
Pham *et al.* [13]	Signal	✓	✓		
Palffy *et al.* [14]	PCD, Image	✓	✓	✓	
Proposed Method	PCD, Image	✓	✓	✓	✓

B. mmWave radar based NLoS object detection

Recent research on detecting NLoS objects utilizing the reflective properties of mmWave radar has been actively pursued [15], [16]. Chen *et al.* proposed a multipath reflection model using MIMO radar and estimated the Time of Arrival for each path through the Matrix Pencil algorithm to infer the position of NLoS objects [12]. However, this method was only validated in an indoor environment, raising concerns regarding its applicability in outdoor and more complex scenarios. Pham *et al.* introduced a Bayesian-based multipath selection technique to reduce localization ambiguities in NLoS situations [13]. Despite this improvement, their method does not explicitly reconstruct the spatial environment. Palffy *et al.* proposed a radar-camera sensor fusion technique for pedestrian localization in NLoS environments, employing a particle filter to estimate the position of pedestrians occluded by vehicles [14]. However, their approach only detects occluded regions without estimating reflectors, limiting its ability to utilize radar points generated by multiple reflections.

To address these limitations, this paper introduces the method that fuses front-view camera data with mmWave radar, enabling more accurate NLoS pedestrian localization. The key differences between the proposed approach and previous research are summarized in Table I.

III. Problem Definition

Accurately localizing NLoS pedestrians requires analyzing mmWave signal reflection paths, but inferring spatial details from radar PCD alone presents significant challenges. To overcome this limitation, the proposed method incorporates the front camera image to enhance spatial inference, with the radar's static points in PCD interpreted based on the front camera image.

The proposed method involves a series of steps for spatial inference. First, the road layout is extracted from the front camera image, denoted as I_{cam}, using a model F_t that transforms I_{cam} into a BEV layout [11]. Based on this road layout, the radar's static points in PCD, denoted as $S = \{s_1, s_2, \dots, s_n\}$, are interpreted to extract the spatial configuration L. This process is expressed as:

$$L = f(S \mid F_t(I_{\text{cam}})), \tag{1}$$

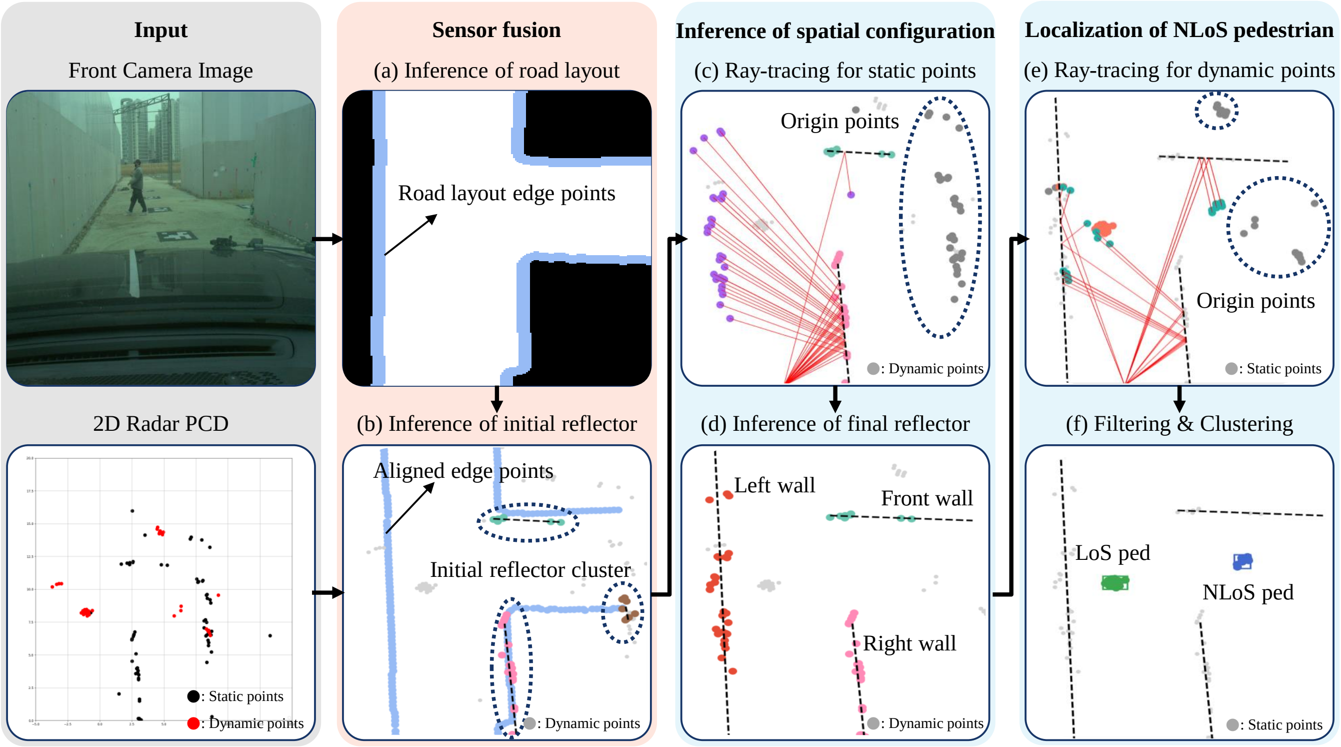
Input
Front Camera Image
2D Radar PCD
●: Static points
●: Dynamic points
Sensor fusion
(a) Inference of road layout
Road layout edge points
(b) Inference of initial reflector
Aligned edge points
Initial reflector cluster
●: Dynamic points
Inference of spatial configuration
(c) Ray-tracing for static points
Origin points
●: Dynamic points
(d) Inference of final reflector
Left wall
Front wall
Right wall
●: Dynamic points
Localization of NLoS pedestrian
(e) Ray-tracing for dynamic points
Origin points
●: Static points
(f) Filtering & Clustering
LoS ped
NLoS ped
●: Static points

Fig. 2. **The overall framework for NLoS pedestrian localization and the results of each algorithm block.**

where f represents the function that interprets the set S based on the road layout extracted from I_{cam}, yielding the spatial configuration L.

Next, a dynamic set of radar points in PCD, denoted as D, is defined from the radar PCD. Utilizing the spatial configuration L, the set D is analyzed to estimate the position of the pedestrian, denoted as X_{pred}. This process is formulated as:

$$X_{\text{pred}} = g(D \mid L), \tag{2}$$

where g is the function that estimates the pedestrian's position by analyzing the set D based on the spatial configuration L. The function g interprets D in the context of L, resulting in the predicted position of the pedestrian, X_{pred}.

Finally, the objective of this method is to minimize the absolute error between the predicted pedestrian position X_{pred} and the ground truth position X_{GT}. The optimization problem is formulated as:

$$X^*_{\text{pred}} = \arg\min_{X_{\text{pred}}} |X_{\text{pred}} - X_{\text{GT}}|. \tag{3}$$

The goal is to find the predicted position X_{pred} that minimizes the absolute error, thereby achieving the most accurate localization of pedestrians in the NLoS region. This process will be further elaborated in Section VI.

IV. NLoS Pedestrian Localization Pipeline through Fusion of Image & PCD

To observe NLoS regions, it is essential to utilize waves with reflective and diffractive properties, and by analyzing these properties, the location of NLoS pedestrians can be inferred. This suggests that the results of such analysis may vary depending on the points of reflection and diffraction, ultimately demonstrating a strong dependence on the spatial configuration. However, accurately analyzing spatial information using a single sensor presents significant limitations.

While mmWave radar PCD provides precise distance measurements, it is constrained by sparse observation data and positional distortion due to multipath reflections. Conversely, camera images offer interpretable information for LoS regions but suffer from relatively inaccurate depth estimation and are unable to detect NLoS objects.

To address these challenges, this section proposes a sensor fusion-based spatial analysis pipeline that integrates the quantitative distance data from mmWave radar with the qualitative visual information from a camera. Based on this fusion approach, we introduce a method for estimating the location of pedestrians in NLoS regions, as shown in Fig. 2,

A. *Inference of spatial configuration*

mmWave radar is capable of distinguishing between static and dynamic objects by analyzing the Doppler effect. In this context, the observation of S refers to the set of static points obtained from the radar PCD. Given the mounting position of the radar in this experiment, these static points can be considered as observations of reflectors. The challenge, however, arises from the fact that static objects can be observed through both direct and reflected paths, as formulated by the

Algorithm 1: Inference of spatial configuration

```
Input: $I_{cam}$: Front camera image,
       $S = \{s_1, s_2, \dots, s_n\}$: Set of static points
       from 2D Radar PCD
Output: $L$: Final set of linear regressions
        representing reflectors
// Inference of road layout
$I_{layout} \leftarrow F_t(I_{cam})$;
// Extract edge points in $I_{layout}$
$I_{edge} \leftarrow$ ExtractEdges($I_{layout}$);
$x_{\text{wall}} \leftarrow X_{\text{scale}} \cdot (u - o_x) + x_{\text{offset}}$   where $(u, v) \in I_{\text{edge}}$;
$y_{\text{wall}} \leftarrow Y_{\text{scale}} \cdot (o_y - v) + y_{\text{offset}}$;
$W \leftarrow \{w_j = (x_{\text{wall}}, y_{\text{wall}}) \mid j = 1, 2, \dots, m\}$;
// Alignment of $W$ & radar PCD
$W_{\text{near}} \leftarrow \{w_j \mid \|s_i - w_j\| < \epsilon\}$;
$\theta, T \leftarrow \arg\min_{\theta, t_x, t_y} \sum_{w_j \in W_{\text{near}}} \min_{s_i \in S} \|s_i - w_j\|$;
$W_{\text{aligned}} \leftarrow R(\theta) W_{\text{near}} + T$;
// Initial interpretation $S$ using $W_{aligned}$
$C \leftarrow \{C_1, C_2, ..., C_k\}$; // DBSCAN on $W_{aligned}$
foreach $C_{ref} \in C$ do
    $S_j \leftarrow \{s_i | \exists w \in W_j,\ \|s_i - w\| < \delta\}$;
    $l_j \leftarrow LinearRegression(S_j)$;
end
// Ray tracing for reflected static points
foreach $r \in S_{reflect}$ do
    $\alpha, \beta \leftarrow$ FindIntersection$(r, L)$;
    $r'_x \leftarrow 2\alpha(r_y - \beta) + r_x \cdot \frac{\alpha^2+1}{\alpha^2+1-r_x}$;
    $r'_y \leftarrow 2\left(\alpha\left(\frac{r'_x + r_x}{2}\right) + \beta\right) - r_y$;
    $S_{\text{relocated}} \leftarrow S_{\text{relocated}} \cup \{r'\}$;
end
// Inference of final spatial configuration
$S_{final} \leftarrow S_{\text{direct}} \cup S_{\text{relocated}}$;
foreach $C_{ref} \in C$ do
    $S'_j \leftarrow \{s'_i \mid s'_i \in S_{\text{final}}, \exists w \in W_j,\ \|s'_i - w\| < \delta\}$;
    $l_j \leftarrow LinearRegression(S'_j)$;
end
return $L \leftarrow \{l_j | j \in \{1, 2, ..., k\}\}$;
```

following expression:

$$S = S_{\text{direct}} \cup S_{\text{reflect}}, \tag{4}$$

where S_{direct} represents the static points observed through the direct path, and S_{reflect} represents the static points observed through the reflected path.

In this context, the coordinates of points observed through the reflected path do not correspond to the actual position of the object but instead contain distorted positional information. Thus, a method for interpreting these points is required. To address this, the present method proposes an algorithm that interprets S through I_{layout} to infer the spatial configuration as summarized in Algorithm 1.

1) Inference of road layout: Recent research on road layout inference models utilizing camera images have been actively pursued. In this research, a model that can infer the road layout, including previously unseen road areas, from monocular camera images is required. One such model, which satisfies this requirement, takes I_{cam} as input and infers the road layout image I_{layout}, as formulated follows:

$$I_{\text{layout}} = F_t(I_{\text{cam}}). \tag{5}$$

2) Extract edge points in I_{layout}: The road layout image I_{layout} is divided into two regions: the drivable space and the undrivable space, which are represented in a binary occupancy map format. Specifically, the pixel values of the drivable region are set to 255, while those of the undrivable region are set to 0. To extract the boundaries of the road from I_{layout}, we detect edges by examining the value changes between adjacent pixels, where the Euclidean distance between pixels is 1. This process is applied to pixels (u, v) in I_{layout} where the pixel value is 255. By doing so, we extract a set of points, I_{edge}, that represent the boundaries of the road.

The coordinates of the points in the extracted I_{edge} are initially expressed in pixel units, necessitating a conversion process into radar coordinate units (meters). The transformation of m transformed image points, $W = \{w_j = (x_{\text{wall}}, y_{\text{wall}}) | j \in \{1, 2, \dots, m\}\}$, is given by the following equations:

$$x_{\text{wall}} = (u - o_x) \cdot X_{\text{scale}} + x_{\text{offset}}, \tag{6}$$

$$y_{\text{wall}} = (o_y - v) \cdot Y_{\text{scale}} + y_{\text{offset}}. \tag{7}$$

where X_{scale} and Y_{scale} are the scale factor values, which are hyperparameters. x_{offset} and y_{offset} are the correction values that account for the difference in mounting positions between the radar and the camera. o_x represents the horizontal central coordinate of the front camera in I_{layout}, and o_y represents the vertical bottom coordinate.

3) Alignment of W & radar PCD: The set S contains observations of multiple reflectors, and without distinguishing each reflector as an individual instance, it is impossible to infer accurate spatial information. However, due to the sparse nature of the observation data, it is difficult to determine whether the points originate from a single structure or multiple reflectors. Thus, relying solely on S may lead to incorrect spatial inferences. To classify the reflectors as distinct instances in S, it is necessary to interpret S based on the qualitative form of W. However, since W provides inaccurate quantitative location information, we must reinterpret S through W first.

To achieve this, we first identify W_{near}, the set of points in W that are within a certain distance ϵ from each point $s \in S$, as formulated below:

$$W_{\text{near}} = \{w_j \mid \|s_i - w_j\| < \epsilon\}. \tag{8}$$

Next, we convert W_{near} into an aligned set of road boundary points in the radar coordinate system, denoted as W_{aligned}, as follows:

$$W_{\text{aligned}} = R(\theta) W_{\text{near}} + T, \tag{9}$$

where $R(\theta)$ is the rotation matrix and $T = (t_x, t_y)$ is the translation vector. To optimize θ and T, we employ a method that minimizes the sum of Euclidean distances between points in W_{near} and S. The optimization process is formulated as:

$$\theta, T \leftarrow \arg\min_{\theta, t_x, t_y} \sum_{w_j \in W_{\text{near}}} \min_{s_i \in S} \|s_i - w_j\|. \tag{10}$$

The optimization process adjusts the values of θ, t_x, and t_y using the Nelder-Mead method, which allows us to obtain the optimal transformation parameters [17]. Once the optimization is complete, the obtained rotation matrix $R(\theta)$ and translation vector T are applied to the entire set of W_{near} to transform it into the radar-aligned set W_{aligned}.

4) Initial interpretation of S using $W_{aligned}$: To accurately infer the environment based on the observations from S, it is essential to identify the points corresponding to the reflectors. To achieve this, we first apply the DBSCAN algorithm [18] to W_{aligned} to partition the data into distinct clusters, each representing a potential reflector, denoted as $\{W_j \mid j \in \{1, 2, \ldots, k\}\}$.

Subsequently, to identify the points in S that correspond to specific reflectors, we extract the subset S_j from S where the distance to any point $w \in W_j$ is within a defined threshold δ, as formulated below:

$$S_j = \{s_i \mid \exists w \in W_j,\ \|s_i - w\| < \delta\}. \tag{11}$$

Next, a linear regression model is applied to the points in S_j to derive the line segment that best represents the geometry of each reflector. The regression process utilizes the Least Squares Method (LSM) to estimate the slope and y-intercept of the line representing each cluster S_j, denoted by α and β, respectively. The resulting linear equation l_j for each reflector is expressed as:

$$l_j : y = \alpha x + \beta. \tag{12}$$

Through this process, we obtain a set of linear models, $L = \{l_1, l_2, \ldots, l_k\}$, each representing an inferred reflector in the environment.

5) Ray tracing for reflect static points: The estimated set L contains spatial information inferred as S_{reflect}, and the process of transforming S_{reflect} to infer the environment is required. In this method, based on the previously inferred set L, we differentiate S_{reflect} from S_{direct} and apply ray-tracing techniques to adjust the position of S_{reflect}.

The ray-tracing model traces the path from the origin O of the Ego-vehicle to a point $s \in S_{\text{reflect}}$, identifying the intersection points where this path intersects the line segments L, which represent the boundaries of reflectors. Among these intersection points, the one closest to the origin is selected as q. The point $r = (r_x, r_y) \in S_{\text{reflect}}$ is then symmetrically reflected across the line $l_j : y = \alpha x + \beta$, which passes through q. The recalibrated points, $S_{\text{relocated}} = \{r' = (r'_x, r'_y)\}$, are computed as follows:

$$r'_x = \frac{2\alpha(r_y - \beta) + r_x}{\alpha^2 + 1} - r_x, \tag{13}$$

$$r'_y = 2\left(\alpha \frac{(r'_x + r_x)}{2} + \beta\right) - r_y. \tag{14}$$

6) Inference final spatial configuration: The set of points S_{final}, reconstructed to align with the actual road environment, is defined as follows:

$$S_{\text{final}} = S_{\text{direct}} \cup S_{\text{relocated}}. \tag{15}$$

Similar to the procedure in Section IV-A.4, linear regression is once again performed on S_{final} to ultimately derive the final spatial information L.

B. Localization of NLoS pedestrian

A pedestrian is observed as a dynamic object represented by a set of points D in the 2D radar PCD. To estimate the actual position of a NLoS pedestrian, ray tracing, as described in Section IV-A, is employed to infer the real-world position of the pedestrian. However, due to the potential presence of clutter and noise in the 2D radar PCD, which may arise from high reflectivity, localization is conducted through a series of filtering and clustering processes. The algorithm is outlined as follows.

1) Ray tracing of dynamic points: The set D represents the observed points of moving objects, which include the set of points acquired through direct paths, denoted as D_{direct}, and the set of points observed after being reflected by reflectors (walls), denoted as D_{reflect}. The set of points in D_{reflect} that are corrected via ray tracing is referred to as the set of relocated points, $D_{\text{relocated}}$. Thus, the complete set of dynamic points, including the relocated points, can be expressed as follows:

$$D' = D_{\text{direct}} \cup D_{\text{relocated}}. \tag{16}$$

2) Filtering & clustering: As the observation distance increases in the 2D radar PCD, the detection error distance becomes larger, and noise arises due to hardware limitations and environmental factors, which subsequently reduce the accuracy of NLoS pedestrian detection. To address these issues, filtering and clustering processes are applied.

The set of relocated points, $D_{\text{relocated}}$, can be represented as the union of points in the LoS region D'_{los} and the set of points in the NLoS region D'_{nlos}, as formulated below:

$$D_{\text{relocated}} = D'_{\text{los}} \cup D'_{\text{nlos}}. \tag{17}$$

Pedestrians in the LoS region are observed through both D_{direct} and D'_{los}. However, since D'_{los} is located further away than D_{direct}, its reliability is lower. Thus, the position of LoS objects can be more accurately estimated using only D_{direct}. Therefore, for LoS objects, only the more reliable D_{direct} is used, and based on Equation 16, the points ultimately used for pedestrian localization, D_{final}, are defined as follows:

$$D_{\text{final}} = D_{\text{direct}} \cup D'_{\text{nlos}}. \tag{18}$$

Subsequently, DBSCAN is applied to D_{final} to estimate the pedestrian's position X_{pred} while simultaneously removing residual noise, as formulated below:

$$X_{\text{pred}} = \text{DBSCAN}(D_{\text{final}}). \tag{19}$$

V. Dataset

A. Test bed & data acquisition vehicle

In real-world road environments, data collection is constrained by safety, and NLoS areas cannot be directly observed. To address this challenge, a testbed with dimensions of 53.5 m × 33.5 m, encompassing various road conditions, was constructed. To estimate pedestrian positions in the NLoS regions, a camera equipped with a fisheye lens was installed at a height of $7m$ at the center of the intersection to provide a BEV of the entire experimental area. For data collection, various sensors were mounted on an SUV, with sensor positions chosen based on those found in actual vehicles. For NLoS object detection, two 77 GHz mmWave radar sensors were placed on the front-right side of the vehicle, and a front camera was mounted at the rearview mirror position to differentiate between NLoS and LoS conditions of objects. Additionally, a LiDAR was integrated to calibrate the sensors and evaluate the inferred spatial configuration. Data from all sensors were collected at 100 ms intervals. Further details can be found in Jeon *et al.* [19].

B. Data for road layout model

A total of 2,931 images were utilized as training data, and 2,200 images were used for validation to train and evaluate the road layout model. The training dataset consisted of annotated images captured while the ego-vehicle was operating within the testbed. Initially, the obtained BEV images were annotated by marking the ego-vehicle's position with a bounding box. For road layout annotation, segmentation was performed, where drivable pixels were represented by a value of 255, and undrivable pixels were represented by a value of 0, resulting in the creation of an occupancy map. Subsequently, based on the annotated position of the ego-vehicle, a Region of Interest was defined. A 1400 × 1400 pixel area was then cropped from the occupancy map, centered around the ego-vehicle's bounding box, to create the corresponding ground truth image.

VI. Experimental Results

To evaluate the performance of the proposed method, validation was carried out in four distinct scenarios, assuming that the ego-vehicle is stationary at a T-junction. The scenarios are categorized into two main cases based on the position of the ego-vehicle: **B1** refers to the case where the ego-vehicle is located on the Left Main Road, with no wall opposite the T-junction, while **B2** corresponds to the case where the ego-vehicle is situated on the Branch Road, with a wall opposite the T-junction. Additionally, four specific scenarios were defined based on the movement of pedestrians approaching the intersection. **S1** represents the scenario where two NLoS pedestrians approach from the right. **S2** involves one NLoS pedestrian approaching from the right, while one LoS pedestrian moves away from the ego-vehicle. **S3** describes the case where two NLoS pedestrians approach from the left, and one NLoS pedestrian approaches from the right. Lastly, **S4** refers to the scenario where two NLoS pedestrians approach from the left, while one LoS pedestrian moves away from the ego-vehicle. Each of these scenarios consists of 60 to 90 frames, allowing for a comprehensive evaluation of the proposed method in various environments and confirming its ability to accurately localize NLoS pedestrians in T-junction scenarios.

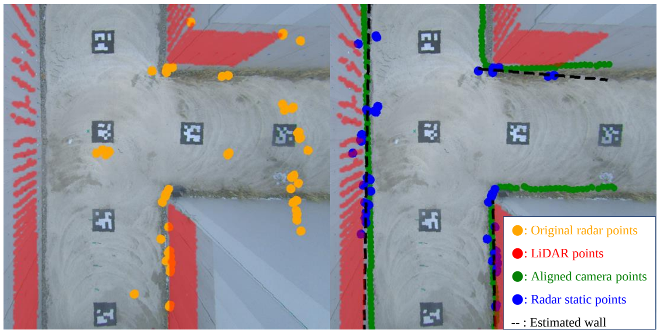

Fig. 3. **Comparison of walls estimated Using LiDAR PCD, aligned camera points, and radar PCD.** (Left) The original radar static points are sparse and contain noise. (Right) The walls estimated using the proposed method are closely aligned with the LiDAR points, showcasing the effectiveness of the approach.

TABLE II

Evaluation of the Spatial Configuration Inference Model.

Scenarios		Sensors	Difference [°]		Max. Diff. with LiDAR
Sites	Cases		F-R	F-L	
B1	**S1**	Radar only	83.77	79.59	7.74
		Proposed	84.94	83.64	3.69
		LiDAR (GT)	87.38	87.33	-
	S2	Radar only	77.41	79.10	9.99
		Proposed	85.72	85.00	2.34
		LiDAR (GT)	87.40	87.34	-
B2	**S3**	Radar only	94.2	76.49	10.89
		Proposed	84.94	83.64	3.74
		LiDAR (GT)	87.44	87.38	-
	S4	Radar only	89.17	79.96	7.44
		Proposed	85.72	85.00	2.4
		LiDAR (GT)	87.44	87.40	-

A. Analysis of 2D Radar PCD using camera data

By analyzing the spatial information of 2D radar PCD alongside data from a front camera, the accuracy of spatial inference can be significantly enhanced. As shown in Fig. 3, radar points corresponding to the front wall are sparsely distributed. When inferring the spatial structure using only radar points, it becomes difficult to determine whether each point originates from a single reflector or multiple reflectors. To address this, the road layout from the front camera can be used to classify the radar points. This classification enables more effective interpretation of the radar points, improving the overall interpretability of the radar PCD.

B. Evaluation of the spatial configuration inference model

In this paper, the pedestrian localization algorithm was evaluated to verify the performance of the proposed spatial configuration inference model. Various sensor configurations, including LiDAR (used as the ground truth), Radar-only, and camera-radar fusion (proposed), were tested. Furthermore, to minimize calibration errors of each sensor, the angular differences between the front wall and the right wall, as well as

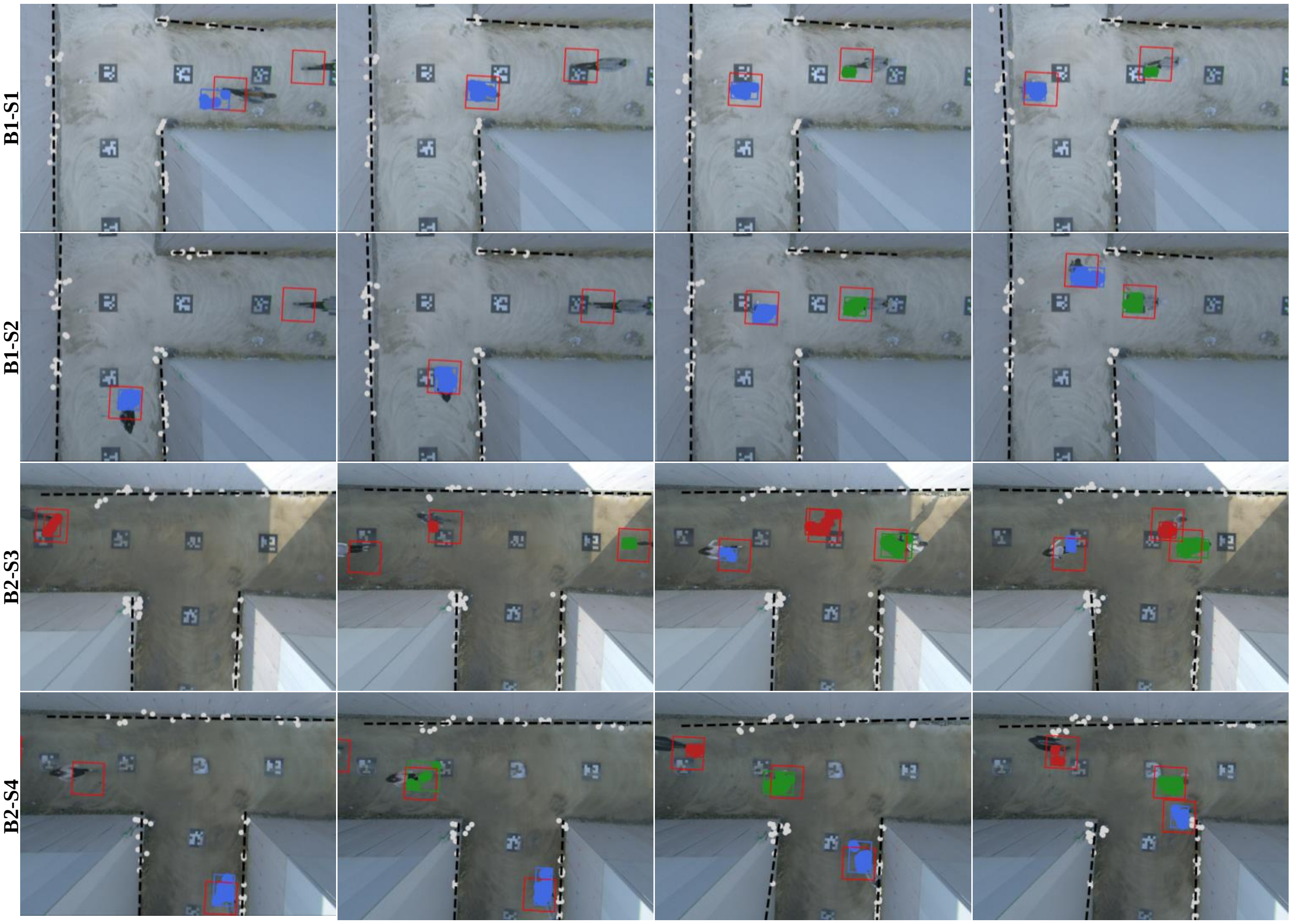

Fig. 4. **Qualitative results in each scenarios.** The figure is organized into four scenarios, displayed from top to bottom. Each scenario is evaluated over time, with frames progressing left to right. Additionally, the distance between each AR mark is 4m and speed of pedestrians is approximately 1.5 m/s.

between the front wall and the left wall, were computed for evaluation. The Front Wall-Right Wall Angular Difference (F-R) and Front Wall-Left Wall Angular Difference (F-L) are defined as follows:

$$\text{F-R} = W_a^{\text{Front}} - W_a^{\text{Right}}, \tag{20}$$

$$\text{F-L} = W_a^{\text{Front}} - W_a^{\text{Left}}, \tag{21}$$

where W_a represents the angle formed with the x-axis.

As presented in Table II, the proposed method exhibited an average error of 2–4 degrees relative to the LiDAR-based measurements, which served as the ground truth, demonstrating high spatial inference accuracy. The evaluation results underscore the effectiveness of the proposed method in reducing the angular difference between F-R and F-L in comparison to the Radar-only method. For instance, in the B1-S1 case, the F-R difference for the Radar-only method was 83.77°, and the proposed method improved this to 84.94°, representing a reduction of 1.40%. In the B1-S2 case, the F-R difference for the Radar-only method was 77.41°, and the proposed method improved it to 85.72°, reflecting a 10.70% enhancement. Similarly, in the B2-S3 and B2-S4 scenarios, the proposed method achieved improvements in F-R, with reductions of 10.00% and 3.89%, respectively.

TABLE III

EVALUATION OF PEDESTRIAN LOCALIZATION.

Scenarios		Methods	Localization Error [AE]		
Sites	Cases		NLoS	LoS	AVG
B1	**S1**	Radar only	2.13	1.85	1.81
		Proposed	0.86	0.26	**0.40**
	S2	Radar only	1.33	1.62	1.61
		Proposed	0.33	0.37	**0.36**
B2	**S3**	Radar only	0.98	1.73	1.55
		Proposed	0.29	0.43	**0.37**
	S4	Radar only	0.64	1.69	1.61
		Proposed	0.43	0.44	**0.44**

In all four scenarios, the proposed method showed a 6.5% improvement in F-R and a 4.6% improvement in F-L angular differences compared to the Radar-only method. The average angular differences for F-R and F-L were reduced from 5.52° and 6.86° to 2.08° and 3.04°, respectively, highlighting the proposed method's ability to enhance spatial configuration inference. The Radar-only method, due to the radar's sparse left wall points and the radar's position on the front-right side of the vehicle, showed greater error in F-L localization. However, the proposed method improved F-L localization accuracy, outperforming Radar-only's F-R estimation. This demonstrates the effectiveness of radar-camera fusion, particularly in NLoS environments, where camera-derived road layout information compensates for the radar's limited view, improving robustness for autonomous driving.

C. Evaluation of pedestrian localization model

The pedestrian localization performance of the proposed method was quantitatively evaluated based on the Absolute Error (AE), which is defined in Equation 3.

In this method, considering the presence of multiple pedestrians, the predicted pedestrian position X_{pred} was matched with the closest actual pedestrian position X_{GT} to calculate the AE. If only one pedestrian was predicted in a given frame, the error between the predicted position and the closest ground truth position was calculated. In cases where multiple pedestrians were predicted, the average AE for each frame was used as the performance metric. The AE was computed separately for three categories: NLoS pedestrian estimation, LoS pedestrian estimation, and the average of both. The experimental results are summarized in Table III. In the B1-S1 scenario, the AE was recorded as 0.40 m, and in the B1-S2 scenario, it was 0.36 m. In the B2-S3 scenario, the AE was 0.37 m, and in the B2-S4 scenario, it was 0.44 m. The inference process takes 0.07 seconds per frame, resulting in a frame rate of 13.7 FPS on a GTX 1080 GPU. The proposed method demonstrated consistent localization performance across different scenarios, with the AE consistently within the 0.44 m range for all scenarios. In the **B1** branch experiment, NLoS pedestrian localization is achieved through the radar→front wall→pedestrian→front wall→radar reflection path. However, in the Radar Only method, due to the sparsity of the radar data, the inference of the front wall is inaccurate, which results in a degradation of NLoS pedestrian localization performance. On the other hand, the proposed method enhances the performance by effectively addressing this issue, demonstrating improved localization accuracy, as shown in Fig. 4.

VII. Conclusions

This paper introduced a novel sensor fusion framework that combined 2D radar PCD with front camera images to achieve accurate pedestrian localization in NLoS environments. The proposed method leveraged front camera images to enhance the interpretation of 2D radar PCD, facilitating spatial inference and the localization of NLoS pedestrians at T-junctions. The proposed method showed a 6.5% improvement in F-R angular difference and a 4.6% improvement in F-L angular difference, compared to the radar only method. Furthermore, when the localization performance was evaluated using AE, the proposed method demonstrated an accuracy of at least 1.18 m across all scenarios, outperforming the radar only method, with an average accuracy of 1.25 m. Additionally, all estimated results of the proposed method were within 0.44 m.These results demonstrate the effectiveness of the proposed method in improving pedestrian localization in NLoS environments.

References

[1] S.-W. Kim, G.-P. Gwon, W.-S. Hur, D. Hyeon, D.-Y. Kim, S.-H. Kim, D.-K. Kye, S.-H. Lee, S. Lee, M.-O. Shin *et al.*, "Autonomous campus mobility services using driverless taxi," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3513–3526, 2017.

[2] National Center for Statistics and Analysis, "Children: 2022 data ((Traffic Safety Facts. Report No. DOT HS 813 575)," National Highway Traffic Safety Administration (NHTSA), June 2024. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813575

[3] B. Rebsamen, T. Bandyopadhyay, T. Wongpiromsarn, S. Kim, Z. Chong, B. Qin, M. Ang, E. Frazzoli, and D. Rus, "Utilizing the infrastructure to assist autonomous vehicles in a mobility on demand context," in *Tencon 2012 IEEE Region 10 Conference*. IEEE, 2012, pp. 1–5.

[4] S.-W. Kim, Z. J. Chong, B. Qin, X. Shen, Z. Cheng, W. Liu, and M. H. Ang, "Cooperative perception for autonomous vehicle control on the road: Motivation and experimental results," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 5059–5066.

[5] W. Liu, S.-W. Kim, Z. J. Chong, X. Shen, and M. H. Ang, "Motion planning using cooperative perception on urban road," in *2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. IEEE, 2013, pp. 130–137.

[6] S.-W. Kim, B. Qin, Z. J. Chong, X. Shen, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "Multivehicle cooperative driving using cooperative perception: Design and experimental validation," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 663–680, 2014.

[7] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Q. Yu, and J. Dai, "Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," 2022. [Online]. Available: https://arxiv.org/abs/2203.17270

[8] L. Peng, Z. Chen, Z. Fu, P. Liang, and E. Cheng, "Bevsegformer: Bird's eye view semantic segmentation from arbitrary camera rigs," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 5935–5943.

[9] J. Philion and S. Fidler, "Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*. Springer, 2020, pp. 194–210.

[10] K. Mani, S. Daga, S. Garg, S. S. Narasimhan, M. Krishna, and K. M. Jatavallabhula, "Monolayout: Amodal scene layout from a single image," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1689–1697.

[11] W. Liu, Q. Li, W. Yang, J. Cai, Y. Yu, Y. Ma, S. He, and J. Pan, "Monocular bev perception of road scenes via front-to-top view projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[12] J. Chen, S. Guo, H. Luo, N. Li, and G. Cui, "Non-line-of-sight multi-target localization algorithm for driver-assistance radar system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 5332–5337, 2022.

[13] B.-H. Pham, O. Rabaste, J. Bosse, I. Hinostroza, and T. Chonavel, "Multipath model order selection for non-line of sight radar localization in urban environment," in *2023 IEEE Radar Conference (RadarConf23)*. IEEE, 2023, pp. 1–6.

[14] A. Palffy, J. F. Kooij, and D. M. Gavrila, "Detecting darting out pedestrians with occlusion aware sensor fusion of radar and stereo camera," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1459–1472, 2022.

[15] Z. Zhu, S. Guo, J. Chen, S. Xue, Z. Xu, P. Wu, G. Cui, and L. Kong, "Non-line-of-sight targets localization algorithm via joint estimation of dod and doa," *IEEE Trans. Instrum. Meas.*, 2023.

[16] S. Fan, Y. Wang, G. Cui, S. Li, S. Guo, M. Wang, and L. Kong, "Moving target localization behind l-shaped corner with a uwb radar," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–5.

[17] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 01 1965. [Online]. Available: https://doi.org/10.1093/comjnl/7.4.308

[18] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[19] M. Jeon, J.-K. Cho, H.-Y. Kim, B. Park, S.-W. Seo, and S.-W. Kim, "Non-line-of-sight vehicle localization based on sound," *IEEE Trans. Intell. Transp. Syst.*, 2024.