

SmokeNav: Millimeter-Wave-Radar/Inertial Measurement Unit Integrated Positioning and Semantic Mapping in Visually Degraded Environments for First Responders

Changhao Chen, Zhiqiang Yao,* Junlin Jiang, Xianfei Pan, Xiaofeng He, Ze Chen, and Bing Wang

First responders often face hazardous and life-threatening situations in environments filled with smoke, posing significant risks to their safety. The existing perception solutions, such as camera or light detection and ranging (LiDAR)-based methods, are inadequate when faced with visually degraded conditions caused by smoke. In this work, SmokeNav, a novel system that combines data from an inertial sensor and millimeter-wave (mmWave) radar, is proposed to enhance situational awareness for first responders in smoky environments. SmokeNav utilizes an inertial positioning module that exploits the human motion constraints with a foot-mounted inertial measurement unit to provide accurate user localization. By integrating this location information with mmWave radar data, it employs a probabilistic occupancy map construction to reconstruct an accurate metric map. To enable semantic understanding of the environment, a DNN-based semantic segmentation model that incorporates radar reflectivity and employs focal loss to improve performance is introduced. Herein, extensive real-world experiments in smoky environments is conducted to demonstrate that SmokeNav precisely localizes the user and generates detailed maps with semantic segmentation. In this work, potentials are held for enhancing the safety and effectiveness of first responders in hazardous conditions.

1. Introduction

First responders frequently encounter hazardous and life-threatening environments, which pose significant risks to their safety. Firefighters, for instance, may face the danger of becoming trapped in hazardous conditions, such as loss of location, trapped in places with smoke. Data collected by the Ministry of Emergency Management of the People's Republic of China revealed that, in the past four years, a total of 177 firefighters lost their lives in China. Given the critical need to rescue and evacuate individuals from burning or collapsed structures, it becomes imperative for emergency responders to enhance their situational awareness.

In search and rescue operations, it is crucial to accurately track the location of first responders and provide a comprehensive map of the unfamiliar environment with semantic understanding.^[1] This capability not only enhances the effectiveness of rescue and exit for first responders themselves, aiding


them in making informed decisions, maintaining situational awareness, and navigating the environment more effectively, but also allows commanders and other teammates outside the immediate vicinity to have a clear visualization of the situation.^[2] Furthermore, key features such as exit routes can be highlighted, assisting in effective navigation and evacuation strategies.

Although self-contained positioning in indoor environments has been largely addressed, the existing perception solutions unfortunately do not meet the requirements of first responders in such settings. Presently, mainstream mapping and perception methods rely on camera or light detection and ranging (LiDAR) technology to perform visual odometry,^[3,4] visual simultaneous localization and mapping,^[5,6] visual-inertial odometry,^[7] or LiDAR localization and mapping.^[8–11] However, camera-based methods are susceptible to degradation or failure in visually degraded environments, such as those filled with heavy smoke in a burning building. Similarly, LiDAR-based solutions encounter obstacles due to the absorption and scattering of laser beams by smoke particles, leading to challenges in accurate object detection and measurement behind the smoke. Furthermore, the scattered light creates interference, resulting in inaccuracies in distance measurements and

C. Chen, X. Pan, X. He, Z. Chen
College of Intelligence Science and Technology
National University of Defense Technology
Changsha 410073, China

Z. Yao, J. Jiang
College of Electronics and Information
Xiangtan University
Xiangtan 411105, China
E-mail: yaozhiqiang@xtu.edu.cn

B. Wang
Department of Aeronautical and Aviation Engineering
The Hong Kong Polytechnic University
11 Yuk Choi Rd, HKSAR, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202400241>.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202400241

making it difficult to distinguish between obstacles and smoke. Consequently, the reliability of mapping and object detection capabilities for first responders in smoky environments is diminished. Moreover, it is crucial for the sensor used in these scenarios to be small sized, lightweight, and energy efficient, as it must align with the limitations of human operators.

With recent advancements in single-chip millimeter-wave (mmWave) radar technology, it has emerged as a promising perception option for first responders due to its low cost (less than 300 dollars), energy efficiency (around 2W), and lightweight design (less than 0.03 kg). Additionally, mmWave radar demonstrates reliable performance in challenging environmental conditions such as smoke, fog, and dust. However, a key challenge lies in the sparsity and noise levels of the point clouds generated by single-chip mmWave radar, which are notably inferior to those produced by LiDAR. Consequently, accurately determining location and creating maps solely using mmWave radar presents difficulties. Existing mapping approaches based on mmWave radar^[12–15] predominantly rely on bulky mechanical radar systems designed for outdoor environments, rendering them unsuitable for our specific requirements. Although Lu et al.^[16] combines wheel odometry with mmWave radar data, this approach is limited to robotic applications and cannot be directly applied to first-responder scenarios.

To tackle these challenges, we propose SmokeNav, a system that integrates data from an inertial sensor and mmWave radar. By leveraging the location information obtained from an inertial positioning module, which utilizes a foot-mounted IMU and exploits human motion constraints, SmokeNav enables the reconstruction of an accurate metric map from mmWave radar data through probabilistic occupancy map construction. To achieve semantic understanding of the environment, we introduce a deep-neural-network-based semantic segmentation model that incorporates radar reflectivity to enhance performance. To address the issue of data imbalance, we employ focal loss. To validate the effectiveness of our proposed SmokeNav system in smoky environments, we conducted extensive real-world experiments. The experimental results demonstrate that SmokeNav achieves precise user localization and generates reliable maps of the environment with semantic segmentation in smoky environments.

In summary, our contributions are as follows. 1) We propose SmokeNav, a novel system that combines inertial sensor, and mmWave radar to accurately localize user and reconstruct a metric map in visual-degraded environments. 2) We present a novel deep neural network framework that achieves object semantic segmentation in smoky environment using very sparse point clouds from mmWave radar, by incorporating radar reflectivity to enhance semantic understanding and utilizing focal loss to address data imbalance problem. 3) Extensive real-world experiments confirm SmokeNav's ability to precisely operate under smoky conditions. Our work holds potentials for enhancing the safety and effectiveness of first responders in hazardous conditions.

2. Related Work

2.1. Milliwave Radar Mapping

Recent advancements in small-size mmWave radar have brought forth exciting possibilities for mapping applications. mmWave

radars, renowned for their wide bandwidth in the GHz range and modulation schemes optimized for range estimation rather than communication (e.g., frequency-modulated continuous wave [FMCW]), have proven successful in object imaging.^[12,15] However, these endeavors have predominantly relied on bulky mechanical radar systems engineered for outdoor settings, where multipath noise has minimal impact. In refs. [14,17], model training was conducted based on simulated mmWave radar input, accurately generating depth images of objects like cars using mmWave radar data. Engelhardt et al.^[13] employed a vehicle-mounted setup encompassing mmWave radar, LiDAR, and depth camera to collect concurrent data; leveraging the LiDAR and depth camera data for training an actual environment model; and applying mmWave radar for outdoor environment mapping. Brodeski et al.^[18] proposed an approach that directly operates on mmWave echo information, establishing an end-to-end network called RadarNet that transforms the echo information into an actual point cloud, thus bypassing traditional radar echo signal processing. Presently, mmWave radar finds primary usage in outdoor autonomous driving for obstacle avoidance. Lu et al.^[16] introduced a generative neural network model based on mmWave data to generate indoor maps. However, this approach still requires wheel odometry information for mapping assistance, making it unsuitable for first responders in our specific context. Thus, there exists a significant gap in investigating the potential of single-chip mmWave radars for imaging indoor environments, marking a critical and unexplored realm of research.

2.2. Pedestrian Inertial Navigation

Strapdown inertial navigation systems (SINS) have been a subject of extensive research for several decades.^[19] Initially, these navigation systems relied on expensive, bulky, and high-precision IMUs, thereby limiting their application to vehicles in motion, such as automobiles, ships, aircraft, submarines, and spacecraft. However, recent advancements in micro-electromechanical system (MEMS) technology have revolutionized IMU production, resulting in significant cost, size, and energy consumption reductions. In the realm of pedestrian tracking, attaching an IMU to the user's foot presents an opportunity to leverage zero-velocity updates (ZUPTs) to compensate for the inherent error drifts of inertial systems.^[20] ZUPTs offer a means to mitigate the cubic error growth associated with standalone SINS by exploiting the stationary intervals during normal gait.^[21] To detect zero-velocity phases and effectively apply ZUPTs within foot-mounted inertial navigation, several approaches have been explored. These methods encompass empirical step-length estimation in navigation systems,^[22,23] motion constraints assuming walking on a flat surface,^[24] systematic sensor error estimation,^[25] gait cycle segmentation,^[26] and discussions on the challenges of step detection at varying gait speeds.^[27] These diverse strategies have contributed to enhancing the accuracy and reliability of zero-velocity detection in foot-mounted inertial navigation systems.

To the best of our knowledge, there is currently no existing research focused on the application of ZUPTs-based inertial positioning to facilitate mmWave radar mapping. While previous

studies, as mentioned earlier, primarily concentrate on providing location information, our proposed SmokeNav goes beyond that by offering comprehensive environmental information, including maps and semantic segmentation of objects. By combining ZUPT-based inertial positioning with mmWave radar mapping, SmokeNav not only provides accurate location data but also equips responders with valuable contextual information essential for effective decision-making and navigation in challenging environments.

3. Method

As shown in **Figure 1**, our proposed **SmokeNav** system consisting of ZUPT-based pedestrian inertial positioning (Figure 1a), mmWave radar mapping assisted by inertial positioning (Figure 1b), and focal-loss-enhanced semantic segmentation using mmWave radar data (Figure 1c). This section will discuss these modules in details separately.

3.1. ZUPT-Based Pedestrian Inertial Positioning

3.1.1. IMU Integration

In an SINS, the gyroscope in the IMU provides angular velocity output, which is utilized to determine attitude information. In contrast, the accelerometer in the IMU yields specific force output, enabling the calculation of position information. Once the initial position, attitude, and velocity of the foot-mounted IMU are determined, real-time recursive estimation allows for continuous navigation updates.

Initially, the attitude of the foot-mounted IMU is computed. The attitude angle can be represented using Euler angles, quaternions, or direction cosine matrices. However, for computational convenience, the quaternion differential equation is commonly employed to update the attitude:

$$\mathbf{q}_{b_k}^{n_k} = \mathbf{q}_{b_{k-1}}^{n_{k-1}} \mathbf{q}_{b_{k-1}}^{b_{k-1}} \quad (1)$$

where $\mathbf{q}_{b_k}^{b_{k-1}}$ denotes the quaternion update of the rotation in the carrier coordinate system between adjacent moments, while $\mathbf{q}_{b_{k-1}}^{n_{k-1}}$ represents the quaternion update in the navigation coordinate

system. The quaternion update of the rotation, $\mathbf{q}_{b_{k-1}}^{n_{k-1}}$, corresponds to the posture at time $k-1$ in the navigation coordinate system, and $\mathbf{q}_{b_k}^{n_k}$ represents the updated posture at time k in the navigation coordinate system.

The terms $\mathbf{q}_{b_k}^{b_{k-1}}$ and $\mathbf{q}_{n_k}^{n_{k-1}}$ can be inferred via

$$\begin{cases} \mathbf{q}_{b_k}^{b_{k-1}} = \begin{bmatrix} \cos \frac{\sigma}{2} \\ \frac{\sin(\sigma/2)}{\sigma} \vec{\sigma} \end{bmatrix} \\ \mathbf{q}_{n_k}^{n_{k-1}} = \begin{bmatrix} \cos \frac{\xi}{2} \\ \frac{\sin(\xi/2)}{\xi} \vec{\xi} \end{bmatrix} \end{cases} \quad (2)$$

In these equations, ξ represents the rotation vector in the navigation coordinate system at adjacent moments, while σ corresponds to the rotation vector in the carrier coordinates. These vectors can be computed as follows:

$$\begin{cases} \xi = \int_{t_{k-1}}^{t_k} \omega_{in}^n d\tau \\ \sigma = \Delta\theta_1 + \Delta\theta_2 + \frac{2}{3}(\Delta\theta_1 \times \Delta\theta_2) \end{cases} \quad (3)$$

Here, $\omega_{in}^n = \omega_{ie}^n + \omega_{en}^n$ represents the rotational angular velocity of the navigation coordinate system relative to the inertial space coordinate system. Furthermore, $\Delta\theta_1$ and $\Delta\theta_2$ denote the angular increments of the gyroscope output between time $(t_{k-1}, t_{k-1} + \frac{t_k - t_{k-1}}{2})$ and $(t_{k-1} + \frac{t_k - t_{k-1}}{2}, t_k)$, respectively.

Subsequently, the current speed is determined using the following equation:

$$\mathbf{v}_k^n = \mathbf{v}_{k-1}^n + \Delta\mathbf{v}_{\text{sflk}} + \mathbf{v}_{g/\text{cork}} \quad (4)$$

Here, \mathbf{v}_k^n represents the velocity at the current timestep, \mathbf{v}_{k-1}^n denotes the velocity at the previous timestep, $\Delta\mathbf{v}_{\text{sflk}} = \int_{t_{k-1}}^{t_k} \mathbf{f}^n dt$ corresponds to the specific force integral increment, and $\mathbf{v}_{g/\text{cork}} = \int_{t_{k-1}}^{t_k} [\mathbf{g}^n - (2\omega_{ie}^n + \omega_{en}^n) \times \mathbf{v}_e^n] dt$ represents the integral increment of gravitational acceleration and Coriolis acceleration.

By combining the previous speed and the current speed, the pedestrian position \mathbf{L} can be obtained using trapezoidal integration:

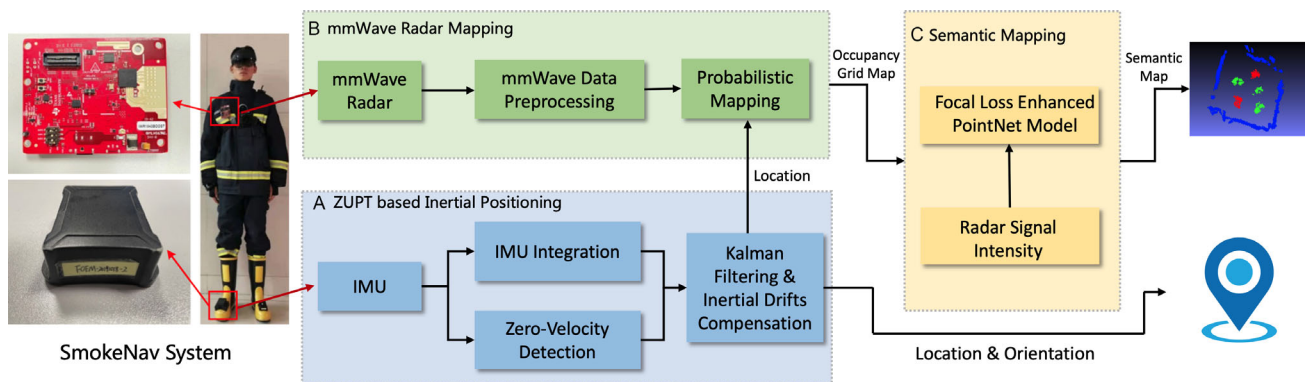


Figure 1. An overview of our SmokeNav framework, consisting of A) ZUPT-based pedestrian inertial positioning, B) mmWave mapping assisted by inertial positioning, and C) focal loss-based semantic segmentation for mmWave radar.

$$\mathbf{L}_k = \mathbf{L}_{k-1} + \frac{\mathbf{v}_k + \mathbf{v}_{k-1}}{2} \Delta T \quad (5)$$

In this equation, ΔT represents the time interval between consecutive solution steps.

3.1.2. Error Model and Kalman Filtering

Based on the principles discussed in the previous part regarding strapdown inertial navigation, the error model for the inertial navigation solution can be derived as follows:

$$\begin{cases} \delta \dot{\mathbf{p}} = \delta \mathbf{v} \\ \delta \dot{\mathbf{v}} = [\mathbf{f}^n \times] \Psi + \mathbf{C}_b^n \delta \mathbf{f}^b - (2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \delta \mathbf{v} \\ \quad - (2\delta \boldsymbol{\omega}_{ie}^n + \delta \boldsymbol{\omega}_{en}^n) \times \mathbf{v} - \delta \mathbf{g}^n \\ \delta \dot{\Psi} = -\boldsymbol{\omega}_{in}^n \times \Psi + \delta \boldsymbol{\omega}_{in}^n - \mathbf{C}_b^n \delta \boldsymbol{\omega}_{ib}^b \end{cases} \quad (6)$$

In this system of equations, $\delta \dot{\mathbf{p}}$, $\delta \mathbf{v}$, and $\delta \dot{\Psi}$ represent the error terms for position, velocity, and attitude, respectively. $\delta \mathbf{f}^b$ denotes the accelerometer error term, $\boldsymbol{\omega}_{ib}^b$ represents the gyroscope error term, and $\delta \mathbf{g}^n$ signifies the gravitational disturbance term.

Consequently, a 15D state Kalman filter can be designed to constrain the error drifts of inertial system:

$$\begin{cases} \delta \dot{\mathbf{x}} = \mathbf{F} \delta \mathbf{x} + \mathbf{G} \mathbf{u} \\ \mathbf{z} = \mathbf{H} \mathbf{x} + \mathbf{v} \end{cases} \quad (7)$$

where $\delta \dot{\mathbf{x}}$ represents the time derivative of the error state vector, \mathbf{F} represents the system matrix, \mathbf{G} represents the control matrix, \mathbf{u} represents the control input vector, \mathbf{z} represents the measurement vector, \mathbf{H} represents the observation matrix, \mathbf{x} represents the error state vector, and \mathbf{v} represents the measurement noise.

Specifically, the state error vector, $\delta \mathbf{x}$, is a 15D vector given by

$$\delta \mathbf{x} = [\delta \Psi \quad \delta \mathbf{v} \quad \delta \mathbf{r} \quad \delta \boldsymbol{\eta}_b \quad \Delta_b]^T \quad (8)$$

It comprises attitude error, speed error, position error, gyro bias, and accelerometer bias. In contrast, the error term for gyro zero bias and accelerometer zero bias is represented by the vector:

$$\mathbf{u} = [\boldsymbol{\eta}_b \quad \Delta_b]^T \quad (9)$$

To effectively estimate the system error and mitigate the divergence of navigation error, it is crucial to carefully select an appropriate observation value, denoted as \mathbf{z} .

3.1.3. Zero-Velocity Detection and Drifts Compensation

The schematic process of the human body movement cycle consists of four stages: zero-speed rest, lift-off from the ground, swing, and touch the ground.^[28] During the foot's resting phase, ideally, the IMU's speed output should be zero. However, due to inherent device errors and cumulative errors, the speed obtained from the inertial navigation solution deviates from zero. To address this, the speed when the foot is at zero speed is artificially set to zero and utilized as a virtual observation to timely correct the system error. This corrective measure aids in suppressing the

accumulation of inertial navigation errors and enhancing positioning accuracy.

The key to implementing this approach lies in acquiring zero-speed gait information. In our work, the statistical value denoted as T serves as the basis for assessment:

$$T = \frac{1}{W} \sum_{k=n}^{n+W+1} \frac{1}{\sigma_a^2} \left\| \mathbf{y}_k^a - \mathbf{g} \frac{\bar{\mathbf{y}}_n^a}{\|\bar{\mathbf{y}}_n^a\|} \right\| + \frac{1}{\sigma_\omega^2} \|\bar{\mathbf{y}}_k^\omega\| \quad (10)$$

Here, n represents the initial moment of the sliding window, W denotes the sliding window size, and σ_a^2 and σ_ω^2 correspond to the measurement noise variance of the accelerometer and gyroscope, respectively. Additionally, \mathbf{g} symbolizes the acceleration due to gravity, and $\bar{\mathbf{y}}_n^a$ denotes the measured acceleration value output by the IMU at the sliding window's initial moment. Furthermore, \mathbf{y}_k^a and \mathbf{y}_k^ω represent the acceleration and angular velocity measurements output by the IMU at time k , respectively.

By employing a predetermined threshold τ , the current moment can be classified as either a zero-speed (stationary) state or a stepping (swinging) moment using the following formula:

$$\delta = \begin{cases} \text{stationary,} & T \leq \tau, \\ \text{swinging,} & T > \tau \end{cases} \quad (11)$$

Here, the threshold τ is established in advance based on body characteristics and motion behavior.

Once the zero-speed moment is detected, the observation quantity of the system becomes

$$\mathbf{z} = \delta \mathbf{v}_k = \mathbf{v}_k - [0 \quad 0 \quad 0] \quad (12)$$

Furthermore, the observation matrix \mathbf{H} of the system is defined as

$$\mathbf{H} = [\mathbf{0}^{3 \times 3} \quad \mathbf{I}^{3 \times 3} \quad \mathbf{0}^{3 \times 3} \quad \mathbf{0}^{3 \times 3} \quad \mathbf{0}^{3 \times 3}] \quad (13)$$

By utilizing the prediction equation and update equation of the Kalman filter, the navigation error during the zero-speed stage of pedestrians can be corrected, leading to an enhancement in pedestrian positioning accuracy.

3.2. mmWave-Mapping-Assisted by Inertial Positioning

3.2.1. mmWave Data Preprocessing

In our solution, the employed 77GHz small single-channel mmWave radar operates as a continuous wave radar utilizing an FMCW transmission signal. The primary measurements performed by the mmWave radar encompass distance (range), velocity (velocity), and angle of arrival (angle) estimation for objects located in front of the radar. Through the mmWave radar signal processing flow, the distance, velocity, and angle are estimated, and the signal-to-noise ratio is calculated using the energy peak value to generate point cloud data.

Upon receiving the point cloud acquired by the data receiving module of the mmWave radar board, the data preprocessing module initiates an input-level filtering process. This step is necessary due to potential deviations in the point cloud's positional accuracy caused by factors such as the radar board's placement or

echo processing. To rectify the point cloud's generation, the point cloud correction module is employed to align the points with their correct positions. Subsequently, the filter module eliminates noise points generated by external disturbances affecting the mmWave, retaining only the relevant environmental information crucial for accurate mapping purposes.

3.2.2. Occupancy Map Construction

Given accurate sensor measurements and location information \mathbf{L} , local map can be directly mapped to the world coordinate system, allowing for the rapid creation of a global map. However, both sensor data \mathbf{z} from adopted mmWave radar sensor and the acquired pedestrian location \mathbf{L} from Equation (5) contain inevitable errors and noises, necessitating the use of probabilistic occupancy grid mapping as a solution.^[29,30]

In the occupancy grid mapping, each point is assigned one of two states: occupied (represented by 1) or free (represented by 0), corresponding to the presence or absence of obstacles, respectively. However, in the occupancy grid map, instead of binary values, probabilities are employed to represent the states of the points. Specifically, $p(s = 1)$ denotes the probability of a point being in the free state, while $p(s = 0)$ represents the probability of it being in the occupied state. These probabilities sum up to 1, and their ratio is introduced as the measure of the point's state:

$$\text{map}(s) = \frac{p(s = 1)}{p(s = 0)} \quad (14)$$

For each point, a measurement observation $\mathbf{z} \sim (0, 1)$ is incorporated, and subsequently, the status of the measurement observation is updated. Prior to obtaining the measured value, the state of the point is denoted as $\text{map}(s)$, which is updated using the following formula:

$$\text{map}(s|\mathbf{z}) = \frac{p(s = 1|\mathbf{z})}{p(s = 0|\mathbf{z})} \quad (15)$$

Applying Bayes' formula yields

$$\begin{cases} p(s = 1|\mathbf{z}) = \frac{p(\mathbf{z}|s = 1)p(s = 1)}{p(\mathbf{z})} \\ p(s = 0|\mathbf{z}) = \frac{p(\mathbf{z}|s = 0)p(s = 0)}{p(\mathbf{z})} \end{cases} \quad (16)$$

By substituting Equation (16) into Equation (15), the following expression is obtained:

$$\text{map}(s|\mathbf{z}) = \frac{p(\mathbf{z}|s = 1)}{p(\mathbf{z}|s = 0)} \text{map}(s) \quad (17)$$

By taking the logarithm of both sides of the equation, the following expression is obtained:

$$\log \text{map}(s|\mathbf{z}) = \log \frac{p(\mathbf{z}|s = 1)}{p(\mathbf{z}|s = 0)} + \log \text{map}(s) \quad (18)$$

Since only the term $\log \frac{p(\mathbf{z}|s = 1)}{p(\mathbf{z}|s = 0)}$ incorporates the measurement observation, it is defined as the measurement model (measurement model):

$$\text{meas} = \log \frac{p(\mathbf{z}|s = 1)}{p(\mathbf{z}|s = 0)} \quad (19)$$

For the measurement model meas , it exhibits two states, namely "free" and "occupied":

$$\begin{cases} \text{free} = \log \frac{p(\mathbf{z} = 0|s = 1)}{p(\mathbf{z} = 0|s = 0)} \\ \text{occu} = \log \frac{p(\mathbf{z} = 1|s = 1)}{p(\mathbf{z} = 1|s = 0)} \end{cases} \quad (20)$$

By introducing the state variable S to represent the position s , the update rule can be further simplified as follows:

$$S^+ = S^- + \log \frac{p(\mathbf{z}|s = 1)}{p(\mathbf{z}|s = 0)} \quad (21)$$

Here, S^+ represents the state of grid s after incorporating the measurement value, and S^- represents the state of grid s before the measurement value. The initial state of a grid is denoted as S_{init} . Since the default probabilities for grid being idle or occupied are both 0.5, we have

$$S_{\text{init}} = \log \frac{p(s = 1)}{p(s = 0)} = 0 \quad (22)$$

After constructing the model and updating each grid state through a series of transformations, only simple addition is required:

$$S^+ = \begin{cases} S^- + \text{free}, & \text{or} \\ S^- + \text{occu} \end{cases} \quad (23)$$

3.2.3. Occupancy Map Update

To update the global map, radial insertion is employed based on the point cloud data from our adopted mmWave radar, as depicted in **Figure 2**. The black region in the figure represents the grid containing an obstacle. Initially, a frame of mmWave radar data is acquired, along with corresponding location information \mathbf{L} from Equation (5) using a foot-mounted IMU. By calculating the grid number associated with the IMU position within the frame, each scan point in the mmWave radar data is traversed. The grid serial number of every obstacle hit within the pixel coordinate system is determined, connecting the

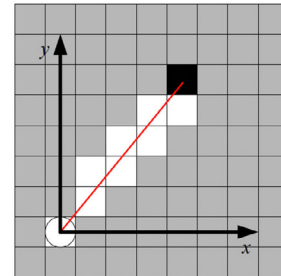


Figure 2. The process of probabilistic occupancy grid map update.

current pedestrian location's grid serial number to the grid serial number of the mmWave radar scanning point with a line. This line signifies an empty raster marker and records the position within the collection.

This process is repeated for each identified obstacle point, resulting in a set of blank grids based on each frame's point cloud. Grids containing obstacles are marked as occupied within this set. Simultaneously, within the set of empty grids, any grids previously marked as occupied due to the addition of an obstacle point are removed. After inputting each IMU location and its corresponding sensor point cloud, the global grid is iteratively updated using this procedure, ultimately completing the update of the occupancy grid map.

3.3. Focal-Loss-Enhanced Semantic Segmentation Using mmWave Radar Data

Inspired by PointNet++,^[31] we design an encoder-decoder-based semantic segmentation framework that inputs mmWave radar map from Equation (14)–(23) of last subsection and radar intensity, and outputs semantic map, as depicted in **Figure 3**. PointNet++ is a well-suited framework for processing point cloud data, owing to its hierarchical structure that effectively captures features at various scales. However, unlike LiDAR data (with ten thousands of points), point clouds obtained from mmWave radar are considerably less abundant (with only hundreds of points), making it challenging to perform semantic segmentation according to object shape alone. To address this problem, along with point clouds, we incorporate the radar's signal intensity, which provides complementary information about material properties and object surface, into our semantic segmentation neural network model. This integration enables a more robust discrimination between different object classes based on their radar reflectivity characteristics. Additionally, to mitigate the data imbalance issue, we introduce focal loss^[32] as an auxiliary during the training process.

3.3.1. Semantic Segmentation Framework

As shown in Figure 3, the feature encoder module of our mmWave semantic segmentation framework performs multi-level downsampling using set abstraction structures to extract point-wise features at different scales,^[33] from a combination of mmWave point clouds and radar intensities as input. Set abstraction consists of three modules: sampling, grouping, and feature extraction. First, farthest point sampling is used to downsample the point set, reducing its scale while ensuring uniformity. Grouping creates local neighborhoods by finding neighbors for each key point. Feature extraction module processes these neighborhoods, producing an output of feature. This downsampling characterization process is repeated at each level of set abstraction, generating smaller point sets.

The decoder module for the segmentation task employs an upsampling process. This is achieved through reverse interpolation and skip connections, enabling the acquisition of both local and global point-wise features to enhance the discriminative nature of the final representation. This framework introduces a reverse interpolation method to implement the structure of the upsampling decoder and obtain discriminative point-wise features through reverse interpolation and skip connections. By iteratively performing reverse interpolation and incorporating skip connections, the decoder progressively samples local and global point-wise features. The resulting global point-wise feature, which captures discriminative characteristics, is then utilized for the segmentation task.

3.3.2. Optimization Loss

The original PointNet++ is trained with the cross-entropy loss function (CE Loss), defined as

$$\text{CE Loss} = - \sum_i^C t_i \log(s_i) \quad (24)$$

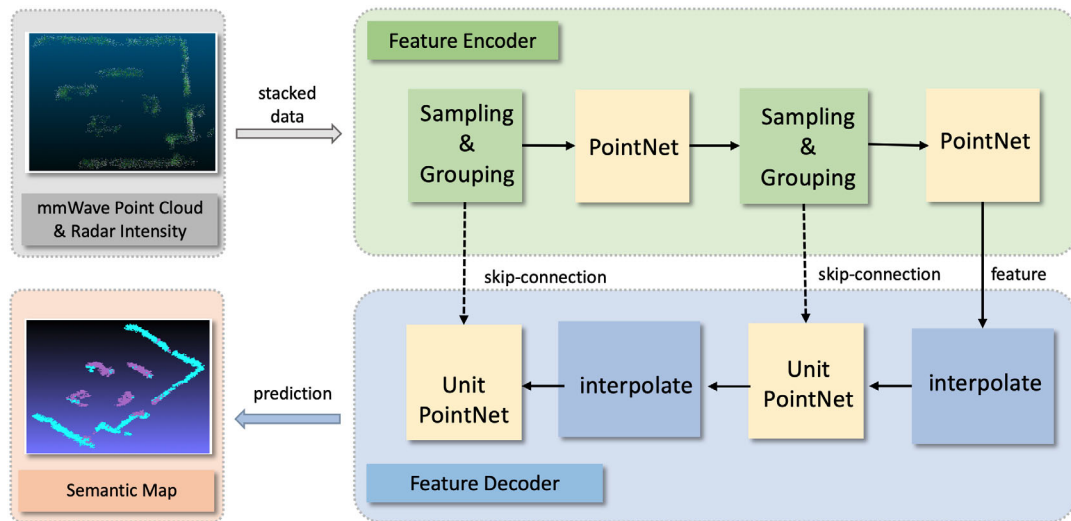


Figure 3. The deep-neural-network-based semantic segmentation model for processing mmWave radar data in our proposed SmokeNav framework.

Table 1. The comparison of mmWave radar and LiDAR in terms of their cost, weight, and power.

	Cost [\$]	Weight [Kg]	Power [W]
LiDAR (Livox Avia)	1000	0.5	9
mmWave radar (TI IWR-1843)	200	0.03	2

where C represents the number of categories, t_i denotes the true value, and s_i refers to the predicted score.

Our map semantic segmentation algorithm, using mmWave radar point cloud and reflectivity information, achieves favorable results when categorizing the point cloud map into two classes (e.g., walls, obstacle boxes), as these two classes are different in their shapes. However, further categorization into three and more classes (e.g., walls, paper boxes, and metal boxes) poses challenges due to imbalanced samples, leading to varying weights within the training model, in our experiments. To address this data imbalance problem in the segmentation task, focal loss^[32] is employed, assigning higher weights to

challenging examples and mitigating the impact of imbalanced data. Specifically, in our model, the focal loss function is introduced, improving upon the weighted CE Loss and ensuring a balanced allocation of penalties across different categories. Specifically, focal loss is formulated as

$$\text{Focal Loss} = - \sum_i^C (1 - s_i)^\gamma t_i \log(s_i) \quad (25)$$

where $(1 - s_i)^\gamma$ represents the weight factor, where $\gamma \geq 0$. This factor is used to reduce the penalty for correctly classified categories, effectively adjusting the balance of the categories without increasing computational complexity. By employing the focal loss function, the issue of data imbalance in the segmentation of mmWave radar point clouds can be addressed.

4. Experimental Section

In this section, we conducted extensive experiments to demonstrate the effectiveness and robustness of our proposed

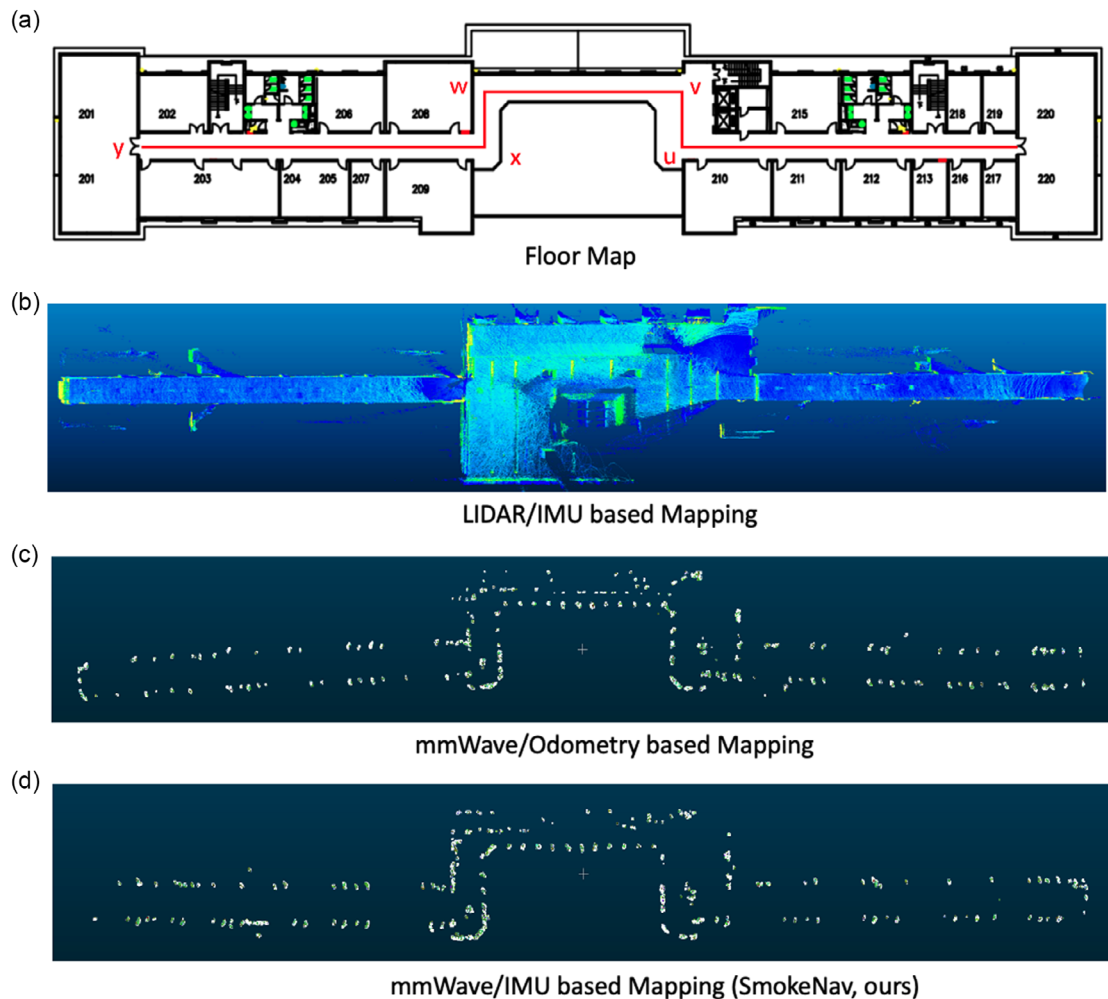


Figure 4. Comparing our proposed d) mmWave/IMU-based mapping with a) floor map, b) LiDAR/IMU-based mapping, and c) mmWave/odometry-based mapping in a corridor.

SmokeNav system in complex real-world situations. First, an evaluation of the system's performance in localizing and mapping was conducted in a long corridor. Then, we investigated the system's generalization ability in smoke-filled environments, assessing its mapping and semantic segmentation capabilities across diverse rooms. Finally, we extended our evaluation to encompass more intricate scenarios, specifically those featuring obstacles composed of various materials.

4.1. Experimental Platform

Figure 1 illustrates our proposed SmokeNav system, comprising a small-sized mmWave radar and a MEMS IMU (MIMU). We used TI IWR 1843 as mmWave radar in our SmokeNav system. The participant was equipped with a MIMU sensor affixed to their feet and a compact single-channel mmWave radar mounted on their body. Our methodology was compared against a representative LiDAR-inertial odometry algorithm, i.e., Fast-LIO₂,^[10] as well as a typical mmWave-odometry-based solution.^[16] These algorithms served as baselines in our localization and mapping experiments.

We compared the cost, weight, and power consumption of the mmWave radar and LiDAR devices used in our experiments. As shown in Table 1, the single-chip mmWave radar was relatively low cost, lightweight, and energy efficient compared to solid-state LiDAR. Therefore, mmWave radar was more suitable for

integration into wearable localization and mapping systems for first responders.

In the context of semantic segmentation, commonly utilized evaluation metrics included Intersection over Union (IoU) and Mean Intersection over Union (Mean IoU). These metrics provided an assessment of the segmentation results' quality. The segmentation algorithm assigned a predicted semantic label to each point within the point cloud, which was then compared to the ground-truth label. Regarding category labels, three metrics were computed: true positive, false positive, and false negative. The intersection ratio represents the ratio of the intersection between the prediction result of a specific category and the actual value to the union. The IoU was calculated via

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (26)$$

4.2. Test 1: Localization and Mapping in Corridor

To assess the localization and mapping performance of our proposed SmokeNav system, we initially conducted experiments within a 100 m long corridor. Two boxes, representing obstacles, were strategically positioned at the opposite corners of this corridor. We selected five positioning marks to evaluate the accuracy of our system's positioning, which are illustrated in the plan map depicted in Figure 4a.

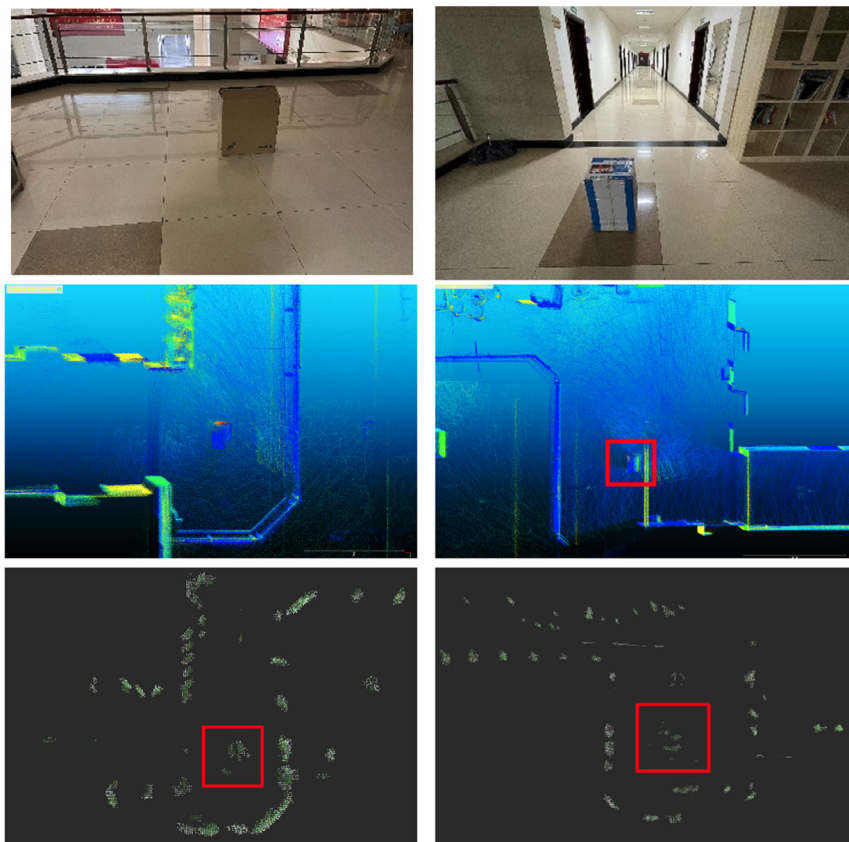


Figure 5. The obstacles are placed in experimental scenes. (Up) Real scene, (middle) LiDAR-generated map, and (down) mmWave-radar-generated map from our proposed SmokeNav.

Figure 4 show cases the mapping results generated by three different methods: our SmokeNav system employing mmWave radar and IMU (referred to as “mmWave + IMU”), FastLIO2^[10] utilizing LiDAR and IMU, and a conventional mmWave radar mapping solution that utilized mmWave radar and wheel odometry (referred to as “mmWave + odometry”). LiDAR sensors were capable of producing a substantial amount of precise point cloud data, enabling more dense and accurate mapping compared to mmWave-radar-based mapping. In contrast, the mmWave radar sensor we employed was relatively small, lightweight, and cost-effective when compared to LiDAR sensors. Our SmokeNav system generated maps with contours similar to those generated by LiDAR, whereas the map produced by the “mmWave radar + odometry” solution exhibited noticeable shape changes, attributed to localization drifts in odometry.

Furthermore, to evaluate the obstacle detection capability of our SmokeNav system, we introduced two boxes as obstacles within the experimental scene, as depicted in Figure 5 (up). Notably, both our SmokeNav system and the LiDAR/IMU-based solution accurately depicted the obstacles in the resulting map, as demonstrated in Figure 5 (down) and (up).

Subsequently, we conducted an evaluation of the positioning performance exhibited by our proposed SmokeNav system in comparison to other baseline methods. The trajectories generated by our SmokeNav system utilizing mmWave radar and IMU (mmWave + IMU), FastLIO2^[10] employing LiDAR and IMU, and the “mmWave-radar + odometry”-based solutions are depicted in Figure 6. Both our SmokeNav system and FastLIO2^[10] demonstrated trajectories that closely resembled the ground-truth trajectory, while mmWave + odometry solution saw a large localization drifts. To gather quantitative results, we carefully selected five positioning marks. The positioning results of our SmokeNav system and the baseline methods are summarized in Table 2. Our proposed method achieved a positioning accuracy of 2.32 m, which closely aligned with the accuracy of 2.53 m attained by the “LiDAR + IMU”-based method. Importantly, when compared to the positioning accuracy of 10.03 m obtained by fusing mmWave radar with wheel odometry, our approach combined data from mmWave radar and IMU, resulting in an approximate 2.33-fold increase in positioning accuracy.

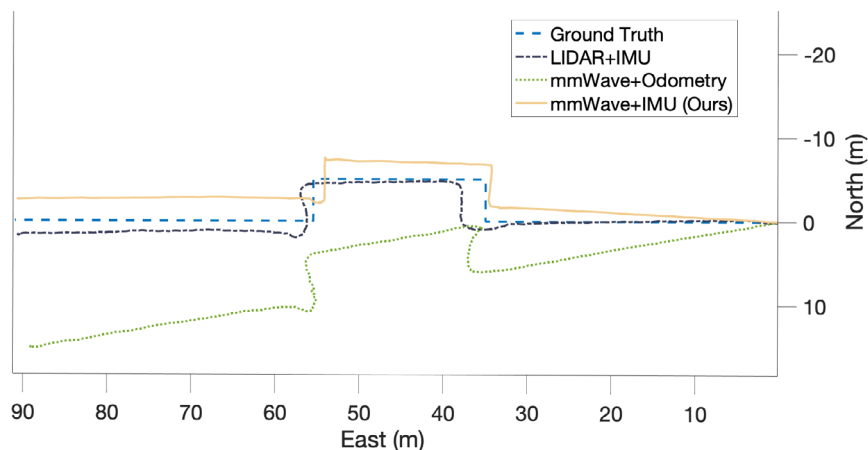


Figure 6. The generated trajectories from our proposed mmWave + IMU-based (i.e., our SmokeNav), LiDAR + IMU-based (i.e., FastLIO2^[10]), and mmWave + odometry-based solutions.

Table 2. The localization results (m) of our proposed SmokeNav (i.e., mmWave + IMU) comparing with LiDAR + IMU and mmWave + odometry in the corridor experiment.

Point	LiDAR + IMU	mmWave + odometry	mmWave + IMU (ours)
1	2.66	6.14	1.91
2	3.27	5.72	1.84
3	1.74	9.36	2.90
4	2.87	10.72	2.24
5	1.73	15.16	2.56
Average	2.53	10.03	2.32

4.3. Test 2: Mapping in Smoky Environment

We proceeded to evaluate the performance of our proposed SmokeNav system in smoky environments.

We first evaluated the mapping performance in an outdoor area with seven randomly placed obstacles. As shown in Figure 7, our SmokeNav system consistently produced reliable mapping results, while the LiDAR system struggled to map the obstacles due to the smoke, which obstructed the LiDAR signal and caused the obstacles to appear cloudlike.

We then conducted a mapping experiment indoors. The data collection took place in two distinct rooms, Room A and Room B. To evaluate the system’s generalization capability in unfamiliar settings, we trained a deep-neural-network-based semantic segmentation model on data gathered from Room A and tested it in Room B.

Figure 8a provides a visualization of Room B under smoke-free conditions, while Figure 8b showcases Room B in an environment filled with smoke. Initially, we employed a “LiDAR + IMU”-based solution to map the surroundings. However, it was evident that LiDAR-based mapping failed to distinguish obstacles from smoke. This limitation arose from the absorption and scattering of LiDAR’s laser beams by smoke particles, impeding accurate object detection and measurement behind the smoke. Moreover, the scattered light created interference, resulting

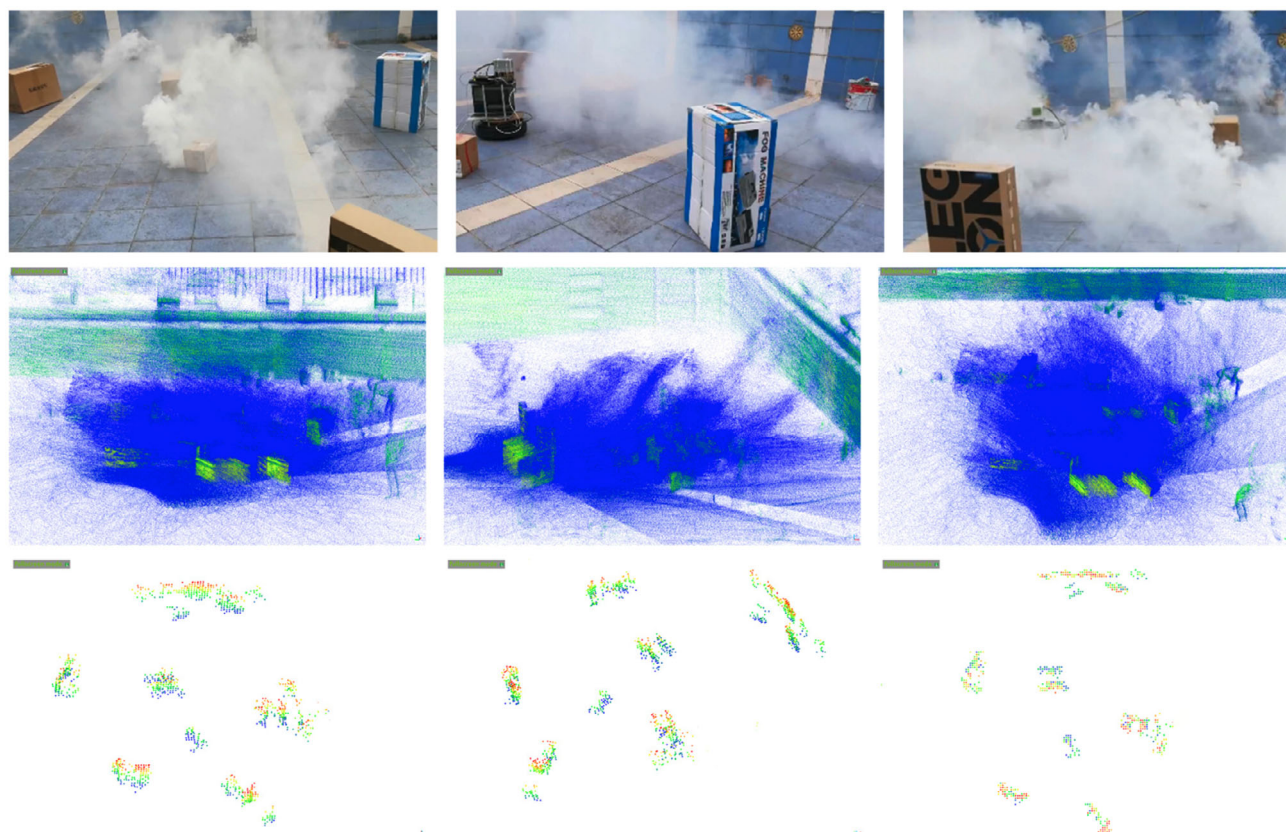


Figure 7. Obstacles are randomly placed in an outdoor area with smoke. (Top) Real scene, (middle) LiDAR-generated map, and (bottom) mmWave-radar-generated map from our proposed SmokeNav system.

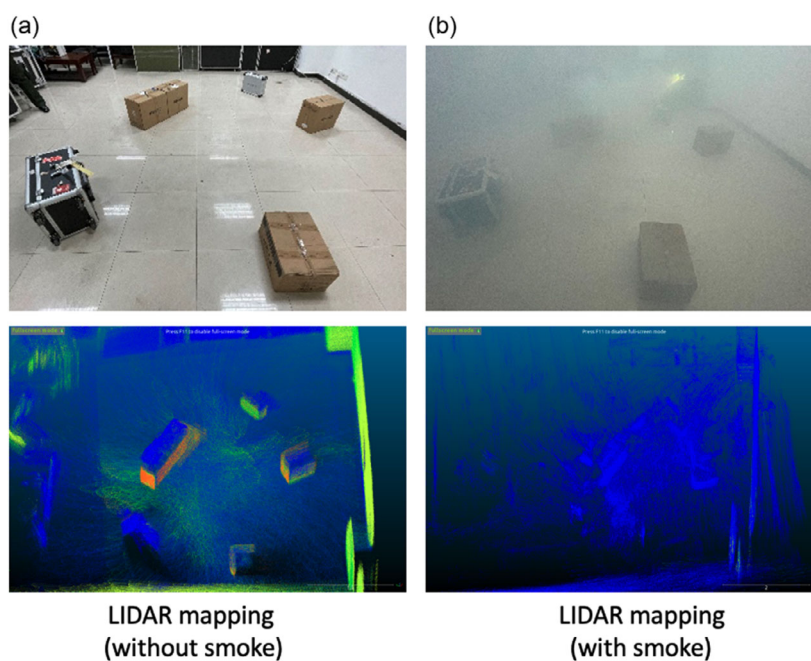


Figure 8. LiDAR/IMU-based mapping in the environments a) without or b) with smoke.

in distance measurement inaccuracies and making it arduous to differentiate between obstacles and smoke. Consequently, this diminished the reliability of mapping and object detection capabilities for first responders in smoky environments.

This test setup required the segmentation of objects into two distinct categories: walls and obstacles. **Figure 9** exhibits the semantic segmentation outcomes obtained from mmWave-radar-based mapping in smoky Room B. Our proposed SmokeNav system utilizing mmWave radar and IMU was compared against the “mmWave radar + odometry”-based mapping approach. To demonstrate the effectiveness of introducing reflectivity, we compared the recognition accuracy of these methods with and without the utilization of reflectivity. It was apparent that the introduction of reflectivity significantly enhanced the accuracy of semantic segmentation recognition under typical smoky conditions, regardless of whether mmWave radar/odometry-based or mmWave radar/IMU-based method was employed. The quantitative results of semantic segmentation in this test configuration are presented in **Table 3**. Our SmokeNav system notably outperformed the other baseline methods. It was worth noting that SmokeNav was trained in Room A but successfully tested in Room B, thus highlighting its effectiveness in real-world unfamiliar environments.

4.4. Test 3: Semantic Mapping to Detect Obstacles with Different Materials

In addition to the generalization test discussed in the previous subsection, Test 3 aimed to assess the performance of

Table 3. The results of semantic segmentation in smoky environment (IoU).

	Wall	Box	Average
mmWave + odometry (without intensity)	0.4412	0.2741	0.3577
mmWave + odometry (with intensity)	0.8543	0.5176	0.6860
mmWave + IMU (without intensity)	0.3622	0.1802	0.2712
mmWave + IMU (with intensity) (ours)	0.9028	0.6685	0.7857

semantic mapping in both Room A and Room B under more complex setups. Specifically, the objects were required to be categorized into walls, carton boxes, and metal boxes. This evaluation intended to demonstrate the effectiveness of incorporating focal loss into deep-neural-network-based semantic segmentation.

Figure 10 showcases the results of semantic segmentation obtained in Room A and Room B, respectively. The IoU metric was calculated to quantify the accuracy of semantic segmentation in both environments, as presented in **Table 4**. Remarkable enhancements in recognition accuracy were observed following the integration of mmWave radar reflectivity information into the segmentation process. Our SmokeNav system (utilizing mmWave radar and IMU with reflectivity) significantly outperformed both the “mmWave radar + odometry”-based mapping and the “mmWave radar + IMU”-based mapping without reflectivity.

Specifically, in Room A, the introduction of point cloud reflectance information resulted in substantial improvements.

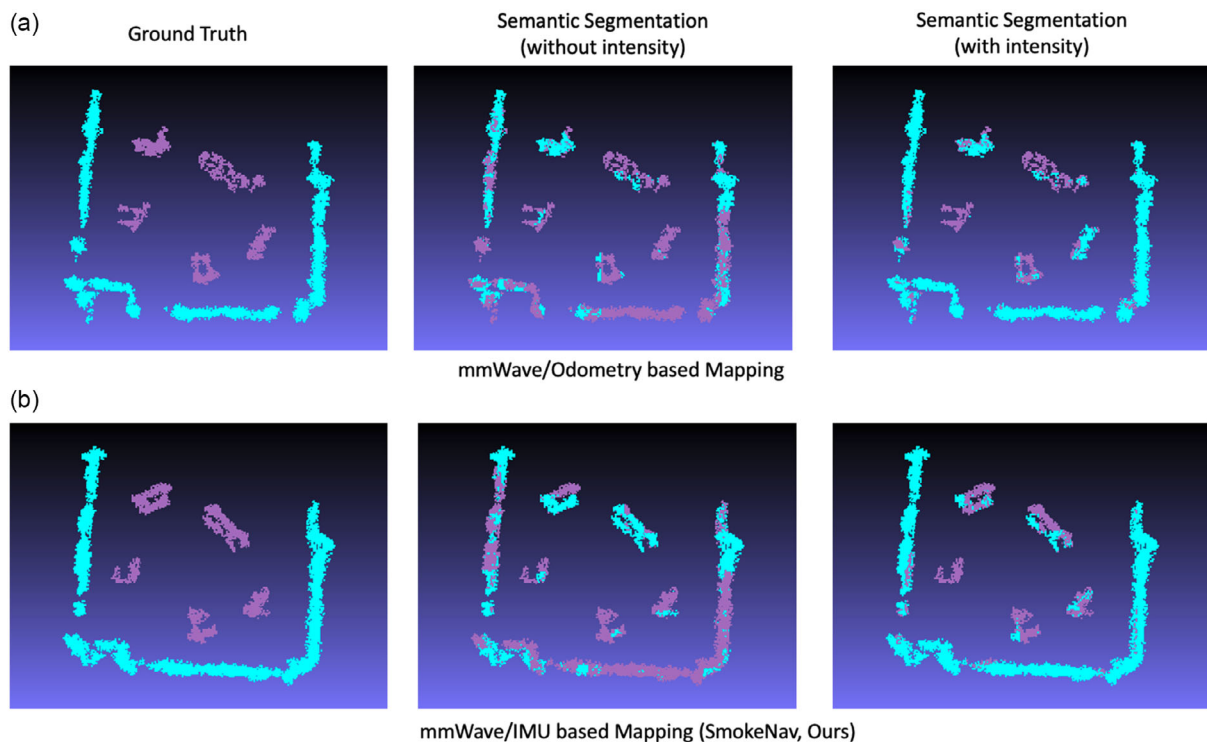


Figure 9. The results of mmWave-radar-based mapping and semantic segmentation in smoky environments of Test 2: a) mmWave/odometry-based mapping and b) mmWave/IMU-based mapping (SmokeNav, ours).

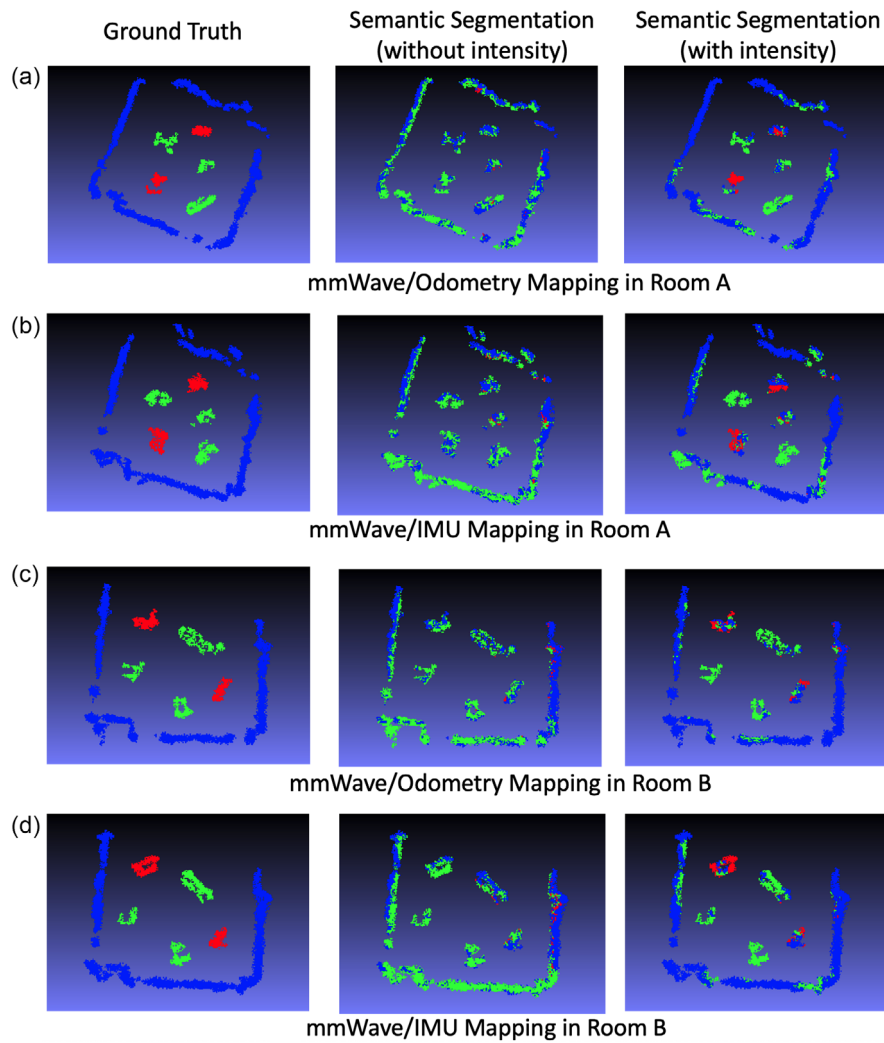


Figure 10. The comparison of different semantic mapping methods under smoky conditions in Test 3: a) mmWave/odometry-based mapping in Room A; b) mmWave/IMU-based mapping (SmokeNav, ours) in Room A; c) mmWave/odometry-based mapping in Room B; and d) mmWave/IMU-based mapping (SmokeNav, ours) in Room B.

Table 4. The results of semantic segmentation in smoky environments (IoU).

	Room A				Room B				Average
	Wall	Carton box	Metal box	Average	Wall	Carton box	Metal box	Average	
mmWave + odometry (without intensity)	0.3633	0.1134	0.0154	0.1640	0.4981	0.1839	0.0166	0.2329	0.1984
mmWave + odometry (with intensity)	0.6911	0.2676	0.4979	0.4855	0.8667	0.5496	0.4062	0.6075	0.5465
mmWave + IMU (without intensity)	0.3976	0.1187	0.0151	0.1771	0.3955	0.1544	0.0125	0.1875	0.1823
mmWave + IMU (with intensity)	0.8002	0.3942	0.5522	0.5822	0.8136	0.4523	0.6140	0.6267	0.6044

The wall recognition accuracy rate exhibited a notable increase of 101.26%, while the accuracy rate for recognizing paper obstacle boxes improved by 232.10%. Moreover, the visibility of metal obstacle boxes was significantly enhanced, making them easily distinguishable. As a result, the overall recognition accuracy rate experienced a substantial boost of 228.74%. Similarly, in

Room B, the inclusion of point cloud reflectance information led to a wall recognition accuracy rate increase of 105.71%, a paper obstacle box recognition accuracy rate increase of 192.94%, and transformed the metal obstacle boxes from being unrecognizable to visible. Consequently, there was a remarkable overall recognition accuracy rate improvement of 234.24%.

5. Conclusion and Discussion

This work aims to address the problems of localization, mapping, and semantic understanding in challenging smoky environments, particularly in scenarios involving first responders. Our proposed SmokeNav system, which combines mmWave radar and IMU data, exhibits superior performance across various evaluation scenarios. The integration of reflectivity information effectively enhances the recognition accuracy of walls and various object categories, surpassing baseline methods such as “mmWave radar + odometry”- and “mmWave radar + IMU”-based method. The promising results obtained in different environments validate the effectiveness and adaptability of SmokeNav, making it a valuable solution for accurate localization, mapping, and obstacle detection in challenging real-world settings. SmokeNav enables improved situational awareness, facilitating more effective navigation and decision-making for emergency responders in challenging and hazardous conditions.

Although mmWave radar systems are relatively low cost, energy efficient, and capable of penetrating smoke and fog, they have certain limitations due to the intrinsic properties of their signals and devices. For instance, mmWave radar typically offers lower spatial resolution than LiDAR, leading to less detailed maps and reduced accuracy in detecting smaller objects or fine features in the environment. Additionally, mmWave radar is susceptible to multipath effects, where signals reflect off multiple surfaces, creating clutter and ghost objects that complicate accurate localization and mapping. While mmWave radar can penetrate through smoke, fog, and dust, its effective range is generally shorter than that of LiDAR, particularly in clear conditions, which can limit its utility in large open spaces. The angular resolution of mmWave radar is also often inferior to that of LiDAR, making it less effective in distinguishing closely spaced objects or in providing precise angular information. Therefore, future research is needed to develop novel data processing techniques and more advanced measurement platforms to further enhance the performance of mmWave radar systems.

Future research can focus on expanding the capabilities of SmokeNav by incorporating additional sensing modalities and advanced machine-learning techniques. Exploring the integration of other data sources, such as thermal imaging or gas sensors, could further enhance the system’s performance and enable comprehensive perception in highly dynamic and unpredictable environments. Additionally, refining the system’s real-time processing capabilities and exploring optimization techniques can contribute to its practical deployment in real-world scenarios, supporting first responders in their critical missions.

Acknowledgements

This work was supported by National Natural Science Foundation of China (NFSF) under the grant nos. 62103427, 62073331, and 62103430. Changhao Chen is supported by the Young Elite Scientist Sponsorship Program of China Association for Science and Technology (CAST).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

localizations, mappings, millimeter-wave radars, pedestrian navigations, semantic segmentations

Received: March 26, 2024

Revised: August 19, 2024

Published online: October 31, 2024

- [1] C. Fischer, H. Gellersen, *IEEE Pervasive Comput.* **2010**, 9, 38.
- [2] A. Dhekne, A. Chakraborty, K. Sundaresan, S. Rangarajan, in *NSDI*, Boston, USA **2019**, Vol. 19, pp. 751–764.
- [3] C. Forster, M. Pizzoli, D. Scaramuzza, in *2014 IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, Piscataway, NJ **2014**, pp. 15–22.
- [4] D. Scaramuzza, F. Fraundorfer, *IEEE Robot. Autom. Mag.* **2011**, 18, 80.
- [5] C. Campos, R. Elvira, J. J. Gómez Rodríguez, J. M. M. Montiel, J. D. Tardós, *IEEE Trans. Robot.* **2021**, 37, 1874.
- [6] Y. Li, C. P. Chen, N. Maitlo, L. Mi, W. Zhang, J. Chen, *Adv. Intell. Syst.* **2020**, 2, 1900107.
- [7] T. Qin, P. Li, S. Shen, *IEEE Trans. Robot.* **2018**, 34, 1004.
- [8] D. He, W. Xu, N. Chen, F. Kong, C. Yuan, F. Zhang, *Adv. Intell. Syst.* **2023**, 5, 2200459.
- [9] Y. Tian, F. Liu, H. Liu, Y. Liu, H. Suwoyo, T. Jin, L. Li, J. Wang, *Adv. Intell. Syst.* **2023**, 5, 2300138.
- [10] W. Xu, Y. Cai, D. He, J. Lin, F. Zhang, *IEEE Trans. Robot.* **2022**, 38, 2053.
- [11] J. Zhang, S. Singh, in *Robotics: Science and Systems*, Berkeley, CA **2014**, Vol. 2, pp. 1–9.
- [12] G. Brooker, D. Johnson, J. Underwood, J. Martinez, L. Xuan, *J. Field Robot.* **2015**, 32, 3.
- [13] N. Engelhardt, R. Pérez, Q. Rao, in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, Piscataway, NJ **2019**, pp. 2866–2871.
- [14] J. Guan, S. Madani, S. Jog, H. Hassanieh, High resolution millimeter wave imaging for self-driving cars. *arXiv preprint arXiv:1912.09579*, **2019**.
- [15] R. Rouveure, P. Faure, M.-O. Monod, *Robot. Autonomous Syst.* **2016**, 81, 1.
- [16] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, A. Markham, in *Proc. of the 18th Int. Conf. on Mobile Systems, Applications, and Services*, Toronto, Canada **2020**, pp. 14–27.
- [17] J. Guan, S. Madani, S. Jog, S. Gupta, H. Hassanieh, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, Washington **2020**, pp. 11464–11473.
- [18] D. Brodeski, I. Bilik, R. Giryes, in *2019 IEEE Radar Conf. (RadarConf)*, IEEE, Piscataway, NJ **2019**, pp. 1–6.
- [19] P. G. Savage, *J. Guid. Control Dyn.* **1998**, 21, 19.
- [20] E. Foxlin, *IEEE Comput. Graph. Appl.* **2005**, 25, 38.
- [21] R. Harle, *IEEE Commun. Surv. Tutor.* **2013**, 15, 1281.
- [22] C. Randell, C. Djalllis, H. Muller, in *Seventh IEEE Int. Symp. on Wearable Computers*, 2003, IEEE Computer Society, Washington, DC **2003**, pp. 166–166.
- [23] R. Stirling, J. Collin, K. Fyfe, G. Lachapelle, in *Proc. European Navigation Conf. GNSS*, Graz, Austria **2003**, Vol. 110.
- [24] P. H. Veltink, P. Slycke, J. Hemsems, R. Buschman, G. Bultstra, H. Hermens, *Med. Eng. Phys.* **2003**, 25, 21.
- [25] S. Scapellato, F. Cavallo, C. Martelloni, A. M. Sabatini, *Sens. Actuators, A* **2005**, 123, 418.
- [26] I. P. I. Pappas, M. R. Popovic, T. Keller, V. Dietz, M. Morari, *IEEE Trans. Neural Syst. Rehabil. Eng.* **2001**, 9, 113.

- [27] A. M. Sabatini, C. Martelloni, S. Scapellato, F. Cavallo, *IEEE Trans. Biomed. Eng.* **2005**, 52, 486.
- [28] J. Wahlström, I. Skog, *IEEE Sens. J.* **2020**, 21, 1139.
- [29] D. Meyer-Delius, M. Beinhofer, W. Burgard, in *Proc. of the AAAI Conf. on Artificial Intelligence*, Toronto, Canada **2012**, Vol. 26, pp. 2024–2030.
- [30] S. Thrun, *Auton. Robot.* **2003**, 15, 111.
- [31] C. R. Qi, L. Yi, H. Su, L. J. Guibas, in *Advances in Neural Information Processing Systems*, Vol. 30 **2017**.
- [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, in *Proc. of the IEEE Int. Conf. on Computer Vision*, IEEE, Piscataway, NJ **2017**, pp. 2980–2988.
- [33] C. R. Qi, H. Su, K. Mo, L. J. Guibas, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2017**, pp. 652–660.