

Human Tracking and Activity Recognition Using Multiple Ceiling-Mounted UWB Radars

Master Thesis

Yanqi Hong



Human Tracking and Activity Recognition Using Multiple Ceiling-Mounted UWB Radars

Master Thesis

by

Yanqi Hong

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on August 26, 2025 at 10:00 AM.

Student number:	5884683
Project duration:	December 15, 2024 – August 26, 2025
Thesis Supervisor:	Dr. Francesco Fioranelli, TU Delft Prof. dr. Guido Dolmans, imec Dr. Utku Kumbul, imec
Thesis committee:	Dr. Francesco Fioranelli, TU Delft, supervisor Dr. Arash Asadi, TU Delft Prof. dr. Guido Dolmans, imec, supervisor

Cover Image: Walker of Light and Shadow in Modern Architecture by ChatGPT (GPT-4o, OpenAI), generated on June 3, 2025. Modified

Acknowledgment

Time flies swiftly, and three years at Delft University of Technology have now come to an end. It feels as if the autumn breeze that welcomed me upon my arrival in Delft were still whispering in my ears, yet today it is already time to bid farewell to this campus. Looking back on these years, I recall not only the challenges and hardships of academic exploration but also the joy and growth that came with them. This journey has allowed me to acquire knowledge and skills, and, more importantly, to understand the true meaning of perseverance and collaboration.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Francesco Fioranelli. Over the past two years, you introduced me to the fascinating world of radar signal processing and guided me forward on the path of academic research. Whether it was the direction of my project, the writing of my thesis, or the daily progress of my work, you have always provided patient guidance and insightful advice. Your rigorous attitude toward research, broad academic vision, and meticulous spirit have benefited me immensely. It has been my greatest privilege and honor to complete this academic journey under your supervision.

I would also like to extend my sincere appreciation to my company supervisor, Prof. dr. Guido Dolmans. You offered me the opportunity to work at imec, a platform where I could pursue my passion for UWB signal processing. With your extensive engineering experience and forward-looking vision, you provided invaluable guidance and inspiration. Thanks to you, I grew not only in theoretical research but also in practical engineering skills. Through our discussions, I came to appreciate the synergy between academic research and industrial application, which further strengthened my determination to continue pursuing this field. At imec, I am equally grateful to my mentor Dr. Utku Kumbul. Thank you for your support and encouragement, as well as for helping me refine my ideas and identify weaknesses throughout my research. Your patient guidance helped me clarify the logic of my project and thesis, while also improving my scientific methodology and academic writing. Your encouragement and recognition have been a constant source of motivation for me.

I would also like to thank all the colleagues and friends who supported me during my graduation project, as well as all the members of my thesis committee. My gratitude goes to Dr. Dingyang Wang and Ir. Binghua Chen, who offered their help and discussed possible technical solutions when my project faced difficulties; and to Siqing Liu, Mingyang Zhang, Zichen Zhao, Yiyang Wu, Yiyun Xu, Jiachang Luo, and Ir. Yue Teng, whose assistance, advice, and care during the data collection process ensured the smooth progress of my experiments. Special thanks also go to all my friends at the imec student corner, at TU Delft, and at TU/e. Your companionship and encouragement made this graduation project not only a scientific journey but also a colorful and joyful life experience.

Last but not least, I owe my deepest gratitude to my parents and grandparents. It is your silent support that has allowed me to pursue the academic path with peace of mind. Your selfless love and the warmth of our family have nurtured me and given me strength and comfort whenever I faced difficulties. I dedicate this work to you as a token of my heartfelt appreciation and respect.

I will forever cherish the invaluable memories of my time in Delft. The challenges, the lessons, and the joy experienced here will continue to inspire me in the years to come, guiding me forward on my journey through life.

*Yanqi Hong
Eindhoven, August 2025*

Abstract

In smart office environments, intelligent control of IoT devices based on human movement and the monitoring of sedentary behavior are crucial. To address object occlusion problem, this thesis proposes an innovative solution: integrating UWB radars into ceiling-mounted office lights, in contrast to conventional wall-mounted setups. The core contribution of this work is the development of an indoor human localization and activity recognition system using ceiling-mounted radars. A novel method is proposed for joint human tracking and activity recognition. The system and the proposed method is validated in a $3.5\text{ m} \times 6\text{ m}$ office environment with three ceiling-mounted Novelda X4 radars. The experimental results demonstrate high performance: trajectory tracking achieves a root mean square error (RMSE) as low as 0.23 m, representing a 52.6% improvement over the baseline method, while activity classification for walking, standing, and sitting reaches an accuracy of up to 98%. These findings demonstrate the feasibility and effectiveness of ceiling-mounted UWB radars for accurate human tracking and activity recognition in office-like environments.

Contents

Acknowledgment	i
Abstract	ii
Abbreviations	x
1 Introduction	1
1.1 Thesis Background	1
1.2 Problem Statement	2
1.2.1 Thesis Scope	2
1.2.2 Problem definition	3
1.2.3 Challenges	5
1.2.4 Research Question	7
1.3 Project Contribution	7
1.4 Thesis Structure	7
2 Technical Background	8
2.1 UWB Technology	8
2.1.1 UWB Introduction	8
2.1.2 UWB Localization	9
2.1.3 Channel Impulse Response	9
2.2 Monostatic UWB Radar	11
2.2.1 Definition	11
2.2.2 Localization using the Monostatic UWB Radar	11
2.3 Related Work in the Literature	12
2.3.1 Tracking	12
2.3.2 Activity Recognition	15
2.4 Summary	16
3 Localization and Tracking	18
3.1 System Model	18
3.1.1 Extended Target Modeling	18
3.1.2 Target Reflection Probability	19
3.2 Overall Pipeline of the Proposed 3D Grid-Based Method	20
3.3 Signal Preprocessing	22
3.4 CIR 3D Grid Mapping	23
3.4.1 CIR Score Calculation	23
3.4.2 Voxel Range Determination	24
3.4.3 CIR Grid Construction	25
3.4.4 CIR Grid Configuration	25
3.5 CIR Grid Fusion and 2D likelihood Projection	27
3.5.1 CIR Grid Fusion	27

3.5.2	2D Likelihood Map Construction	27
3.6	3D Grid Compensation	29
3.6.1	Artifacts Compensation	30
3.6.2	Target Ring Enhancement	34
3.6.3	PDOP Compensation	38
3.7	Particle Filter Tracking	42
3.7.1	Particle Filter Overview	43
3.7.2	State Model	43
3.7.3	Measurement Model	44
3.7.4	Resample and Roughening	44
3.7.5	Initialization	45
3.8	Baseline Method	46
3.8.1	Overall Pipeline of the Baseline Method	46
3.8.2	CFAR Detection	47
3.8.3	Leading Edge Detection	48
3.8.4	Kalman Filter on Target Range	48
3.8.5	Taylor Series based Localization	49
3.8.6	Kalman Filter on 2D Trajectory	51
3.9	Summary	51
4	Activity Recognition	52
4.1	General Introduction of Activity Recognition	52
4.2	Walk-Static Classification	52
4.2.1	Background Signal Energy	53
4.2.2	Movement Signal Energy	53
4.2.3	Static Period Extension	54
4.3	Stand-Sit Classification	54
4.3.1	Human Body Region Construction	55
4.3.2	Human Height Estimation	55
4.3.3	Height Change Decision	56
4.4	Summary	56
5	Experimental Setup and Evaluation	57
5.1	Radar Configuration	57
5.1.1	Radar Selection	57
5.1.2	Radar Deployment	61
5.2	Experimental Setup	62
5.2.1	Experiment Office Room	62
5.2.2	Ground Truth Data Measurement	63
5.2.3	Experiment Content	64
5.3	Evaluation of Localization and Tracking	67
5.3.1	Evaluation metric of Localization and Tracking	67
5.3.2	Experimental Results of Localization and Tracking	68
5.3.3	Analysis of Localization and Tracking Results	74
5.4	Evaluation of Activity Recognition	75
5.4.1	Evaluation Metrics	76
5.4.2	Experimental Results of Activity Recognition	76
5.4.3	Analysis of Activity Recognition Results	82

5.5	Summary	83
6	Conclusion and Future Work	84
6.1	Conclusion	84
6.2	Future Work	85
	References	86

List of Figures

1.1	Scope of the thesis project in a sketch with blocks containing the main activities.	3
1.2	Image of the selected office room for data collection and experiment.	3
1.3	Conceptual illustration of multiple ceiling-mounted UWB radars for human tracking and activity recognition in a smart office.	4
1.4	Illustration of limited opening angle issue for the ceiling mounted UWB radars.	5
1.5	Illustration of TOA localization using three radars in a 2D plane.	6
2.1	Comparison of the power spectral density and bandwidth of UWB radio technology versus other technologies, from [7].	9
2.2	Illustration of a CIR sample collected with a UWB radar in two scenarios. Scenario 1 (orange line): no target in front of the radar (background only). Scenario 2 (blue line): one target in front of the radar (CIR with target).	10
2.3	Illustration of range time plot generated by stacking together multiple CIR samples collected over time.	10
2.4	Example of three different radar configurations: monostatic, bistatic, and multistatic. The multistatic radar configuration can be different compared with the example one shown in the plot, but there should be at least three nodes.	11
2.5	Illustration of localization process by using multiple (i.e., 3) monostatic UWB radar nodes.	12
2.6	Typical pipeline for a Detect-Before-Track approach.	13
3.1	Visualization of the human body as an extended target from the side view.	19
3.2	Visualization of the human body as an extended target from the top view.	19
3.3	Overall Pipeline of the Proposed 3D Grid-Based Method.	21
3.4	An example result of raw data vs ACS using the dataset from test ID <i>VH_T_W_P1_H1</i>	22
3.5	Pipeline of mapping CIR onto a 3D grid.	23
3.6	Illustration of the voxel size of the 3D grid and its range with respect to the UWB radar.	24
3.7	An example of 2D slices (along the vertical z -dimension) from the CIR grid is shown for the experimental data sample with test ID <i>VH_T_W_P1_H1</i> , obtained from UWB radar 1 at slow-time index $t = 390$	26
3.8	An example of 2D slices (in the vertical z -dimension) from the target reflection probability grid is shown for the experimental data sample with test ID <i>VH_T_W_P1_H1</i> , obtained at slow-time index $t = 390$	28
3.9	An example of the likelihood map is shown for the experimental data sample with test ID <i>VH_T_W_P1_H1</i> , obtained at slow-time index $t = 390$	28
3.10	An example of the likelihood map is shown for the experimental data sample with test ID <i>VH_T_W_P1_H1</i> , obtained at slow-time index $t = 210$	28
3.11	An example of 2D slices (in the vertical z -dimension) from the target reflection probability grid is shown for the experimental data sample with test ID <i>VH_T_W_P1_H1</i> , obtained at slow-time index $t = 210$	29
3.12	The relationship between the three 3D grid compensation methods applied in this thesis.	29

3.13	Illustration of the artifacts in the CIR grid across different vertical layers.	30
3.14	Pipeline of the Artifact Compensation Process.	31
3.15	Illustration of the radial shrink strategy for artifact compensation.	31
3.16	Illustration of artifact distance calculation described in this section.	33
3.17	Comparison of the results before and after artifact compensation for 3D CIR grid $G_{m,t}$, target reflection probability grid R_t and 2D likelihood map L_t	33
3.18	Illustration of artifact rings across all radars and the resulting false alarms at slow-time $t = 390$ for experimental test sample with ID $VH_T_W_P1_H1$	34
3.19	Pipeline of the Target Ring enhancement process.	35
3.20	Example of peaks along the radial line at $n_z = 6$ in the CIR grid $G_{m,t}$, where radar $m = 1$ and the slow-time index is $t = 390$, from experimental test sample with ID $VH_T_W_P1_H1$	36
3.21	Comparison of the result before and after Target Ring enhancement for 3D CIR grid $G_{m,t}$, target reflection probability grid R_t and 2D likelihood map L_t	38
3.22	Illustration of the geometry of satellite placement affecting PDOP (from [53]): (a) Satellites placed close together result in poor localization accuracy and a higher PDOP value; (b) Satellites placed evenly result in better localization accuracy and a lower PDOP value. The solid circles represent the true distances from the satellites to the user and intersect at the receiver's position. The dotted circles, which are parallel to the solid circles, represent the erroneous measurements obtained by the receiver. The red region indicates the uncertainty area.	39
3.23	Example of normalized PDOP values for horizontal layers of the experiment setup.	40
3.24	Example of the PDOP compensation factor tensor for horizontal layers of the experimental setup.	42
3.25	Comparison of the result before and after PDOP compensation of 2D likelihood map.	42
3.26	Overall pipeline of the baseline method considered for performance comparison.	46
4.1	Pipeline of the proposed walk-static classification procedure.	53
4.2	Illustration of static period extension to account for transition periods.	54
4.3	Pipeline of the proposed stand-sit classification procedure.	54
4.4	Illustration of the extended body region represented by a cuboid.	55
5.1	Image of Novelda X4M03 and Novelda X4M300 [64][65].	58
5.2	Image of Novelda X7F202 [66].	58
5.3	Illustration of Novelda X4 data acquisition program.	59
5.4	Illustration of Novelda X7 data acquisition pipeline.	60
5.5	Image of the Novelda X4 radar installed on the ceiling.	62
5.6	The floor plan of the office room for experiment.	63
5.7	The person records timestamps while walking to provide ground truth for tracking evaluation.	64
5.8	Route of four designed walking trajectory on the floor plan.	66
5.9	Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID $VH_T_W_P1_H1$	69
5.10	Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID $VH_T_W_P1_H1$	69
5.11	Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID $VH_T_W_P2_H1$	70
5.12	Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID $VH_T_W_P2_H1$	70
5.13	Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID $VH_T_W_P3_H1$	71

5.14	Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID <i>VH_T_W_P3_H1</i>	71
5.15	Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID <i>VH_T_W_P4_H1</i>	72
5.16	Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID <i>VH_T_W_P4_H1</i>	72
5.17	Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID <i>VH_T_W_P1_H2</i>	73
5.18	Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID <i>VH_T_W_P1_H2</i>	73
5.19	Tracking trajectory of Walk-Stand test at 'F1'.	77
5.20	Distance error and CDF of the proposed 3D grid-based method for Walk-Stand test at 'F1'.	77
5.21	Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Stand test at 'F1'.	78
5.22	Tracking trajectory of Walk-Stand test at 'D3'.	78
5.23	Distance error and CDF of the proposed 3D grid-based method for Walk-Stand test at 'D3'.	79
5.24	Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Stand test at 'D3'.	79
5.25	Distance error and CDF of the proposed 3D grid-based method for Walk-Sit test at 'F1'.	79
5.26	Tracking trajectory of Walk-Sit test at 'F1'.	80
5.27	Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Sit test at 'F1'.	80
5.28	Tracking trajectory of Walk-Sit test at 'D3'.	81
5.29	Distance error and CDF of the proposed 3D grid-based method for Walk-Sit test at 'D3'.	81
5.30	Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Sit test at 'D3'.	82

List of Tables

2.1	Summary of Wall-Mounted UWB Radar-Based Human Tracking Approaches.	14
2.2	Summary of Ceiling-Mounted Non-UWB Radar Based Human Tracking Approaches.	15
2.3	Summary of Radar-Based Activity Recognition Approaches.	16
3.1	Input arguments in the findpeaks function used within the processing pipeline for Target Ring enhancement.	36
5.1	Specifications of Novelda X4 and X7 Radars [67] [68].	59
5.2	Configuration Parameters of Novelda X4 UWB Radar.	61
5.3	Designed walking trajectories for localization and tracking experiments.	65
5.4	Explanation of the Test Data ID Format.	66
5.5	Tracking performance comparison of proposed 3D grid-based method and baseline.	74
5.6	Binary confusion matrix (rows: actual, columns: predicted).	76
5.7	Summary of activity recognition results.	82

Abbreviations

Abbreviation	Definition
AOA	Angle of Arrival
AWGN	Additive White Gaussian Noise
BiLSTM	Bidirectional Long Short-Term Memory
CA-CFAR	Cell-Averaging Constant False Alarm Rate
CDF	Cumulative Distribution Function
CFAR	Constant False Alarm Rate
CIR	Channel Impulse Response
CNN	Convolutional Neural Network
CUT	Cell Under Test
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DBT	Detect-Before-Track
ESS	Effective Sample Size
GDOP	Geometric Dilution of Precision
GPS	Global Positioning System
IoT	Internet of Things
KF	Kalman Filter
ML	Machine Learning
MPC	Multipath Component
MTI	Moving Target Indicator
MVDR	Minimum Variance Distortionless Response
NLOS	Non-Line-of-Sight
PDF	Probability Density Function
PDOP	Position Dilution of Precision
PF	Particle Filter
RCS	Radar Cross Section
RMSE	Root Mean Squared Error
RPE	Relative Position Error
RTI	Radio Tomographic Imaging
Rx	Receiver
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
TBD	Track-Before-Detect
TDOA	Time Difference of Arrival
TOA	Time of Arrival
Tx	Transmitter
UWB	Ultra-Wideband
VATS	Variance-based Temporal-Spatial

1

Introduction

In this chapter, the background and motivation of the thesis are first introduced. Then, the research problem is scoped, defined, and summarized. A summary of the main contributions is presented, followed by an overview of the thesis structure.

1.1. Thesis Background

With the rapid development of smart buildings and Internet of Things (IoT) technologies, the concept of the smart office has emerged as a key direction in the transformation of modern workplace environments. A smart office not only aims to optimize spatial resource allocation and energy consumption, but also focuses on enhancing user comfort and interactive experience. Among the enabling technologies that support intelligent office management, high precision indoor human tracking and activity recognition play a critical role by providing foundational data for intelligent automation.

In smart office environments, data such as occupant distribution, activity status, and space utilization can enable a wide range of intelligent functions, including automated IoT device control, security monitoring, and detection of prolonged sedentary behavior [1][2][3]. For instance, precise localization and tracking of employees' trajectories facilitates dynamic control of lighting and air-conditioning systems, thereby enhancing energy efficiency. It also enables automatic monitoring of desk occupancy, contributing to improved spatial utilization. Furthermore, activity recognition can help determine an employee's working status. For example, detecting extended periods of sitting and prompting timely reminders to stand and take a break. In emergency situations such as fires, accurate localization and tracking of individuals can ensure personal safety and a better emergency response.

Traditional indoor tracking and activity sensing technologies such as Wi-Fi, Bluetooth, Radio Frequency Identification (RFID), infrared, and vision-based systems often suffer from limitations including low positioning accuracy, privacy concerns, vulnerability to multipath effects, and relatively high energy consumption [4][5] [6]. These challenges are evident in office environments, where dynamic occlusions, frequent wall reflections, strict privacy requirements (e.g., business confidentiality), and increasing demand for energy-efficient solutions further worsen the limitations of conventional methods. Therefore, there is a growing need for robust, accurate, and privacy-preserving sensing technologies that can operate reliably in complex and dynamic indoor office settings.

In recent years, Ultra-Wideband (UWB) radar has attracted growing interest in a variety of human-centric sensing

applications, such as presence detection, activity recognition, device-free localization, and vital sign monitoring, due to its fine range resolution, low transmission power, and strong resilience to narrow-band interference [7]. Compared to Wi-Fi or other infrared-based technologies, the wide bandwidth of UWB enables high range resolution, allowing the system to detect slight human movements such as posture shifts or breathing patterns with significantly greater sensitivity [7]. Unlike vision-based sensing, UWB also preserves user privacy by avoiding the collection of identifiable visual or audio data, making it particularly suitable for office environments where confidentiality and unobtrusiveness are critical concerns. These advantages collectively position UWB radar as a promising candidate for smart office scenarios that demand accurate, privacy preserving, and energy efficient human tracking and activity sensing.

In this project, an innovative deployment strategy by mounting UWB radars on the ceiling is proposed. Several UWB radars are installed on the ceiling, facing downward, for human tracking and activity recognition. Compared to the conventional wall-mounted setup which is typically positioned around 1 m above the ground, ceiling-mounted deployment offers several advantages in office environments [7] [8] [9] [10] [11]. For example, it can reduce occlusion effects caused by furniture and dense crowds, which is a common issue near walls where shelves, monitors, or desks are often placed [12]. Moreover, some office walls are made of glass or other materials that are not suitable for device installation, further limiting wall-mounted deployment options.

In addition, multiple ceiling-mounted UWB radars may provide a broader and less obstructed field of view, enabling more consistent detection performance across the network of radars. Looking ahead, UWB radars can potentially be integrated into office ceiling light, which would not only streamline installation, but also improve system unobtrusiveness and maintain aesthetic consistency within modern office interiors. While this special deployment also includes challenge which is discussed in section 1.2.3.

Because of the advantages of ceiling-mounted UWB radars, this thesis focuses on human tracking and activity recognition in indoor office environments using this deployment strategy, considering as an implementation example a network of three radars.

1.2. Problem Statement

In this section, the scope of the thesis is clarified at first. Then, the definition of the core problem to be addressed is presented. The specific technical challenges associated with the proposed system setup with radars on the ceiling are then analyzed. Finally, the main research question is formulated, to guide the remaining part of the thesis.

1.2.1. Thesis Scope

This thesis is conducted at the Interuniversity Microelectronics Centre (imec) in Eindhoven, a globally renowned research institute specializing in nanoelectronics and digital technologies [13]. The focus of this work is on the complete signal processing chain for human tracking and activity recognition, including data collection, algorithm development, and performance evaluation. The overall scope of the thesis is illustrated in Figure 1.1.

The experimental data used in this thesis were collected in a real-life office environment using UWB radar sensors and a data collection plan. The experiment data are specifically collected in the ‘Workplace Vitality Hub’, a data-driven smart office facility [14]. A specific office room within the Hub, measuring 6 m in length, 3.5 m in width, and 2.7 m in height, was selected for data collection. Ceiling-mounted UWB radars were installed to monitor and record human activity. The room is a typical office setting in terms of size and interior furnishings, providing a realistic environment for evaluating the proposed methods. An image of the selected office room is shown in Figure 1.2, where the position of red circles denoting the position of ceiling-mounted UWB radars.

The signal processing framework developed in this thesis includes both tracking and activity recognition algorithms

which is capable to track human's walking trajectory and classify human's activity at the same time. After the processing stage, the results are quantitatively evaluated to assess the system's accuracy and robustness.

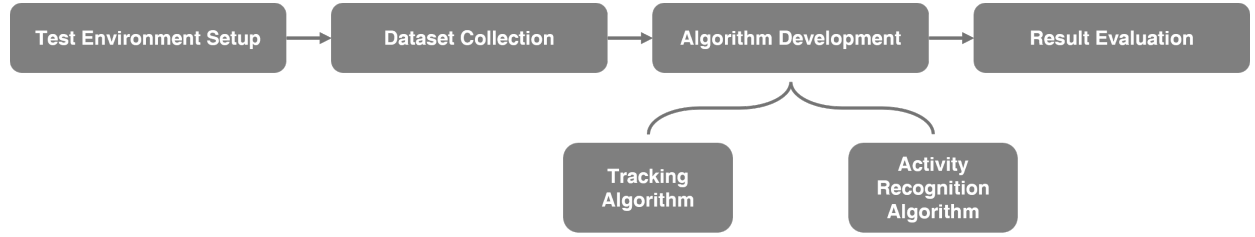


Figure 1.1: Scope of the thesis project in a sketch with blocks containing the main activities.



Figure 1.2: Image of the selected office room for data collection and experiment.

1.2.2. Problem definition

Based on the defined scope of this thesis, the research is structured around three main topics:

1. Data collection in a real office environment
2. Human localization and tracking
3. Human activity recognition

The first topic (*Data collection plan*) focuses on designing an effective data collection procedure. This not only includes selecting suitable UWB radar sensors and implementing a ceiling-mounted installation strategy that fits typical office environments, but also defining what kind of data should be collected. The dataset to include different activities (walking, standing, sitting), multiple subjects of varying height, and diverse trajectories to reflect realistic office scenarios should be included to the data collection plan.

The second topic addresses the challenge of *human localization and tracking*. Although many studies have explored the use of UWB radar systems for human tracking, most of them rely on wall-mounted configurations. These radars are usually installed about one meter above the ground and oriented horizontally toward the human body [7] [8] [9] [10] [11]. In comparison, very few studies have investigated ceiling-mounted placements. Among those, one installs

the radar at a ceiling height of 2.5 meters but facing to the side wall [15]. While another paper places the radar facing the floor from the ceiling, but in that work the target is lying down, which is not relevant to office scenarios [16]. Therefore, there is a lack of systematic research on tracking walking individuals using multiple downward-facing UWB radars mounted on the ceiling. This thesis aims to address this gap by proposing and evaluating a ceiling-based deployment strategy specifically designed for smart office environments.

Based on the motivation described in section 1.1, the detection of prolonged sedentary behavior is also a key objective of this thesis. This requires the system to *recognize different human activities* within an office environment. To simplify the problem, three representative activity types are selected for classification during the tracking process: walking, standing, and sitting. Similar to the tracking problem, most existing studies on UWB-based activity recognition employ wall-mounted radar configurations [7] [17]. Only a limited number of works have explored ceiling-mounted UWB radar for activity recognition, and among them, only one adopts a machine learning (ML)-based approach (CNN) using such a configuration [12]. Therefore, research on activity recognition using ceiling-mounted UWB radar remains scarce. Furthermore, activity transitions can occur during movement, such as transitioning from walking to sitting or standing. This thesis also aims to explore how to continuously recognize human activities along a movement trajectory, which is another limited explored topic in the field.

Figure 1.3¹ demonstrates the problem of this thesis with a conceptual illustration. It is worth noting that, due to time constraints and the limited size of the dataset collected, this thesis focuses on single-person tracking and activity recognition. The dataset was collected manually by the author in a real office environment. In addition, since a large amount of data required for training ML models like a Neral Network is not available, the activity recognition component will rely on non-ML-based algorithms.

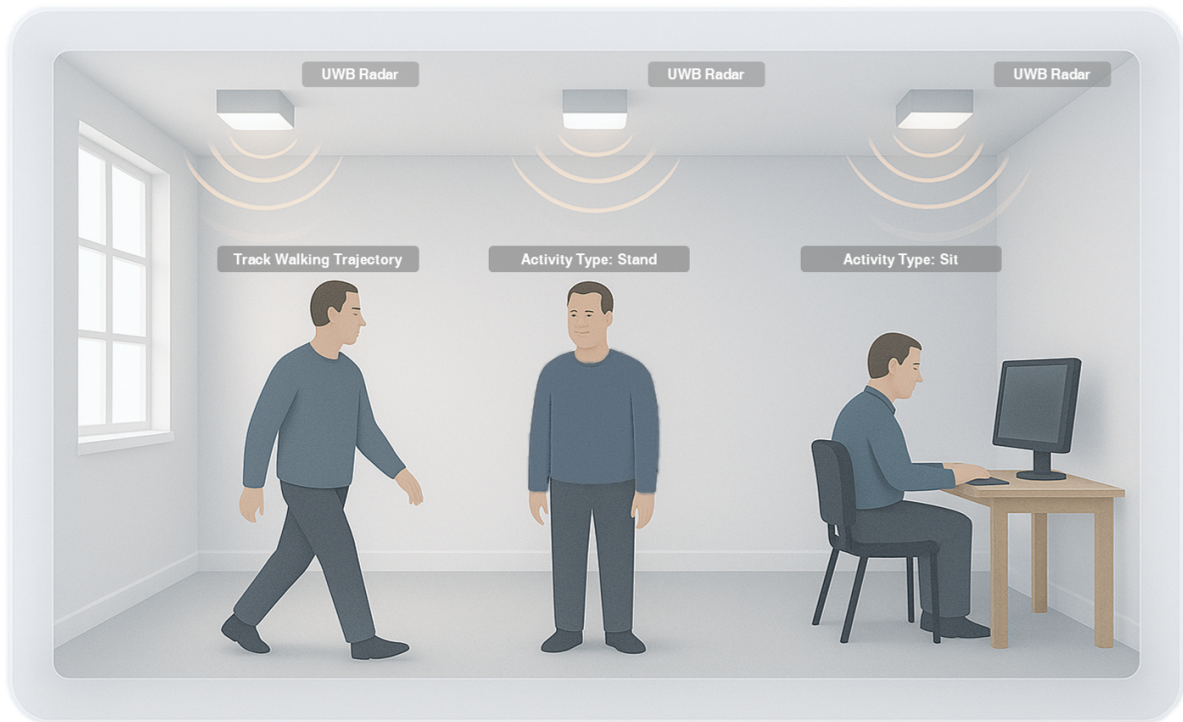


Figure 1.3: Conceptual illustration of multiple ceiling-mounted UWB radars for human tracking and activity recognition in a smart office.

¹Modification based on GPT-4o figure generation

1.2.3. Challenges

As described in the previous sections, one of the main innovations of this project is the use of downward-facing UWB radars mounted on the ceiling. While this deployment strategy offers advantages such as wider coverage and reduced occlusion from furniture and walls, because the ceiling-mounted radars require only a small installation area on the ceiling to enable sensing across the entire room. It also introduces several technical challenges that must be addressed. The key challenges associated with this configuration can be summarized as follows:

1. **Extended Target and Weak Target Problem:** In a ceiling-mounted, downward-facing configuration, different parts of the human body reflect radar signals from different heights, resulting in multiple reflection peaks for a single person. For example, the head is closer to the radar than the feet, leading to distinguishable returns from the head, torso, and lower limbs. This causes a single person to appear as multiple targets in the radar data, which may lead to the generation of multiple tracks for one walking person. Similar observations were reported in [18] when a ceiling-mounted mmWave radar was used.

Another challenge arises from the reduced radar cross section (RCS) of the human body when the radar is mounted on the ceiling. Compared to wall-mounted configurations, the top-down view provides a smaller reflective surface area, which results in weaker radar signal returns. In contrast, a wall-mounted radar faces the entire side profile of the human body, yielding a larger reflective surface and stronger signal returns. This phenomenon leads to the weak target problem, whereby the received signal strength is hard to detect and track robustly [19]. Combined with the extended target effect discussed earlier, these factors pose significant challenges for achieving reliable single-person tracking using ceiling-mounted UWB radars.

2. **Limited Opening Angle and Viewpoint Variation:** The downward-facing, ceiling-mounted UWB radar has a limited field of view, typically constrained by the main lobe opening angle of the radar antennas. For instance, the Novelda X4 and X7 modules have a maximum opening angle of approximately 120° , which makes it difficult to cover the entire room [20] [21]. As illustrated in Figure 1.4, the triangle icons represent the UWB radars, and the dashed lines denote the boundaries of the maximum opening angle $\theta_{\max}$.

When a person is located near a corner of the room (Position 1), the head may fall outside the main lobe and into the side lobes, resulting in reduced signal strength or even missed detection. In such cases, the lower limbs may become the dominant reflection source, as indicated by the blue cross in Figure 1.4. In contrast, when the person is directly below the radar (Position 2), the entire body is within the field of view, and the head typically generates the strongest reflection (also shown as a blue cross).

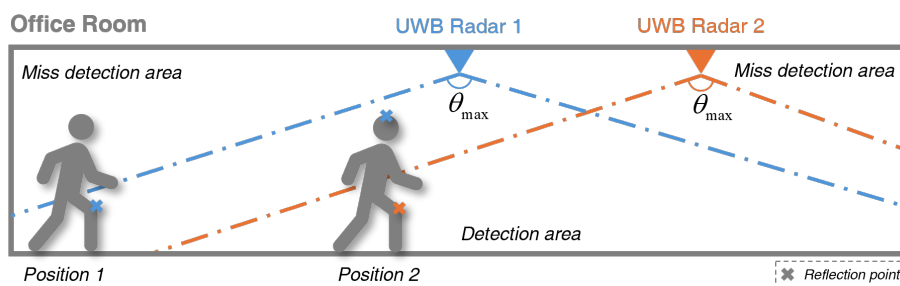


Figure 1.4: Illustration of limited opening angle issue for the ceiling mounted UWB radars.

This variation in the location of the main reflection point can lead to inconsistencies across multiple radars. For example, at target Position 2, UWB radar 1 may detect the head as the dominant reflection (blue cross), while UWB radar 2, due to its limited viewing angle, may only detect the lower limbs (orange cross). Besides, although both radars can detect the reflection from the person's head, slight discrepancies may exist between

their estimated head positions due to small measurement errors. These viewpoint-dependent differences can result in conflicting localization measurements for the same target, posing significant challenges to accurate and consistent multi-radar data fusion and further localization process.

3. **Limited Number of UWB Radars:** In general, there is an interest in minimizing the number of radars deployed to ensure acceptable performances. In the specific case of this thesis, due to equipment availability limitations, only three mono-static UWB radars are used for the data collection. For the selected radars, the distance between each radar and the target based on signal propagation time can be estimated. The estimated distances define circular regions centered at each radar location.

Hence, each dashed circle in Figure 1.5 represents an estimated distance from a radar to the target. Since the estimated range is not perfectly accurate, the actual target position is assumed to lie within the intersection region of the circular rings, where each ring corresponds to the possible range of the estimated distance for a given radar. The possible target position is then determined as the common intersection of these rings, illustrated by the blue shaded area around target point P in Figure 1.5.

In two-dimensional space, at least three radars are required for localization. Using only two circles results in two possible intersection points, which creates ambiguity in determining the true position. However, in real-world measurements, noise and inaccuracies may cause one radar to produce wrong distance estimates. In such cases, the three circles may not intersect at a common point, making localization unreliable.

Compared to other studies that typically use four or more mono-static UWB radars to improve robustness and fault tolerance in full-room two-dimensional localization tasks, the three-radar configuration used in this thesis offers minimal redundancy [7] [22]. Some studies have successfully deployed only two radars, but these are often placed on the same side of a wall and used to localize targets confined to one side of the monitored space. In such cases, the geometric ambiguity caused by two-circle intersection is naturally avoided [23][24]. However, this setup is not applicable to the scenario in this thesis, where ceiling-mounted radars are required to track human movement throughout the entire room. This limited configuration increases the system's sensitivity to measurement errors and introduces significant challenges for accurate and reliable localization. A more detailed explanation of TOA localization will be provided in section 2.2.2.

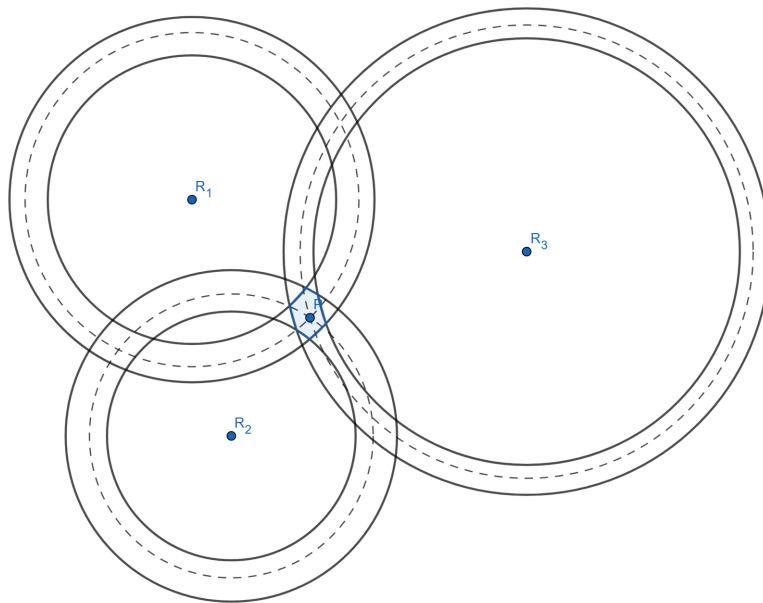


Figure 1.5: Illustration of TOA localization using three radars in a 2D plane.

1.2.4. Research Question

Based on the analysis and discussion in this section, the main research question of this thesis can be formulated as:

How to develop a processing pipeline to track a single walking person and simultaneously perform activity recognition including walking, standing, and sitting by using three downward-facing, ceiling-mounted UWB monostatic radars in a smart office?

Stemming from the above main question, several sub-questions are formulated to guide the investigation steps in this thesis:

1. How can the challenges of extended targets and weak radar reflections be addressed in single-person tracking, and is it possible to exploit the extended target characteristics to improve tracking performance?
2. How can the issue of viewpoint variation across multiple ceiling-mounted radars be mitigated, and can this variation be utilized to enhance tracking accuracy?
3. What classification approaches can be employed to accurately recognize human activities such as walking, standing, and sitting when little data is available for training?

1.3. Project Contribution

The main contributions of this thesis are listed as follows:

- **Proposed and validated the effectiveness of ceiling-mounted UWB radars for human tracking and activity recognition:** This research systematically demonstrates the feasibility of using multiple ceiling-mounted UWB radars for indoor human localization and activity recognition, achieving both high tracking accuracy and high activity classification accuracy.
- **Developed a novel 3D grid-based method:** The proposed method maps the Channel Impulse Response (CIR) to a 3D grid and applies compensation mechanisms to generate an accurate 2D likelihood map, which forms the basis for subsequent trajectory tracking and activity classification. In particular, the method explicitly considers the extended-target characteristics of a walking human, enabling the system to exploit multiple reflection points and thereby enhance localization robustness and accuracy.
- **Achieved high precision trajectory tracking and activity recognition:** By combining a particle filter for trajectory estimation with parallel height estimation for activity classification, the system achieved excellent performance in a real-world office environment using three Novelda X4 UWB radars. Compared to the baseline method, the proposed approach reduces the root mean square error (RMSE) by 52.6% to 0.23 m, improves the relative position error (RPE) by 63.3%, and achieves an activity recognition accuracy of 90%.
- **Validated system robustness in complex environments:** The method demonstrated good stability and reliability even under non-line-of-sight (NLOS) conditions, in weak-signal regions, and in limited opening-angle areas with viewpoint variation issues, such as room edges.

1.4. Thesis Structure

The rest of this thesis is structured as follows. Chapter 2 presents the technical background and reviews related researches. Chapters 3 and 4 introduce the proposed algorithms for single-person tracking and activity recognition, respectively, along with detailed processing pipelines. Chapter 5 presents the experimental setup and evaluates the performance of the proposed methods using real-life data collected in an office room. Finally, Chapter 6 summarizes the findings of the thesis and discusses potential directions for future research.

2

Technical Background

In this chapter, the following fundamental topics will be discussed: the basic principles of UWB technology, localization techniques, monostatic radar configurations, and signal propagation characteristics. Additionally, the related work from the scientific literature including tracking and activity recognition will be introduced.

2.1. UWB Technology

In this section, the fundamentals of UWB technology and UWB localization will be introduced, followed by a description of the Channel Impulse Response (CIR), which serves as the main data format used for signal processing in this thesis.

2.1.1. UWB Introduction

UWB is a radio technology that transmits short pulses (in the order of sub-nanoseconds or picoseconds) over a wide bandwidth (3.1 GHz to 10.6 GHz) compared with other radio technology (such as Wi-Fi and Bluetooth), offering high precision in distance and positioning [7] [25] [26]. Figure 2.1 shows the wide bandwidth of UWB compared with other radio technologies. Due to their large bandwidth, cost-effectiveness, and robust performance, UWB radio chips are useful for applications such as localization, communication, and radar, with their key advantage being high-range resolution, making them particularly suitable for indoor positioning, short-range communication, and object detection [7] [27].

The propagation of UWB signals in mobile and indoor environments is complex, influenced by environmental factors and signal properties. Due to the complex radio channel of UWB, the multipath and fading effects can lead to significant signal degradation. UWB propagation channels also cause substantial changes in pulse shape, including phase, amplitude, and frequency dispersion, resulting in significant signal distortion [7]. Therefore, advanced signal-processing techniques are essential to mitigate these effects and enhance UWB system performance.

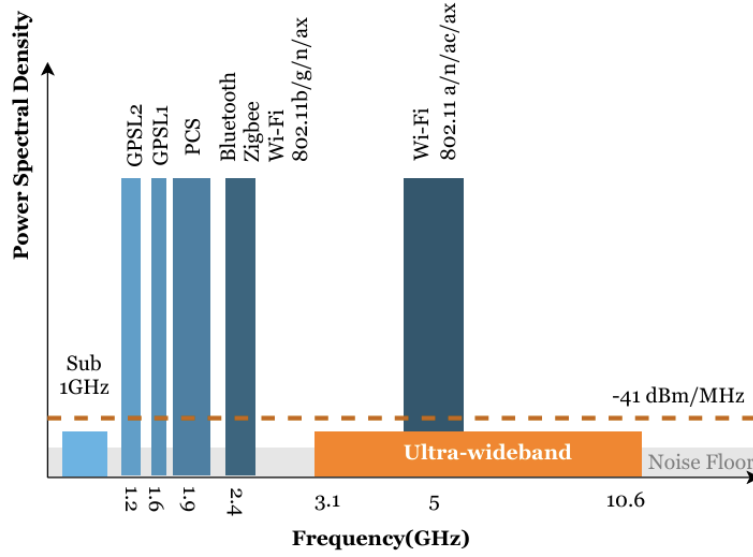


Figure 2.1: Comparison of the power spectral density and bandwidth of UWB radio technology versus other technologies, from [7].

2.1.2. UWB Localization

UWB positioning technology can be divided into active and passive positioning [28]. Active positioning requires the target to carry a UWB tag. UWB base stations receive the UWB signal emitted by the tag through Time Difference of Arrival measurements across multiple base stations, and the system can precisely calculate the tag's location. This method provides high positioning accuracy and strong real-time tracking. However, each tracked object needs a tag, which increases both device costs and operational complexity. So, its application may be limited or unfeasible in scenarios where deploying tags for all targets is impractical or prohibitive.

Passive positioning operates without requiring the target to carry any device. Instead, multiple UWB anchors in the environment transmit and receive UWB signals, measuring changes in the Channel Impulse Response (CIR) to determine position. This approach aligns with radar principles, where electromagnetic waves are transmitted and their echoes analyzed to locate objects, which can be considered as a form of UWB radar [19]. When an object is present, UWB signals reflect off its surface, causing changes in the CIR. By analyzing these changes, the system can detect the object's location and movements. Similar to radar, this method leverages signal reflection and multipath propagation, making it ideal for tracking individuals or stationary objects where active devices are impractical. Its benefits include reduced equipment cost, lower power consumption, and enhanced privacy, making it suitable for applications such as care for older or vulnerable individuals, health monitoring, and smart security systems requiring precise, contactless tracking.

2.1.3. Channel Impulse Response

As described in Section 2.1.2, the CIR between transmitter (Tx) and receiver (Rx) can help the radar system to detect the location of the target. The UWB radio signal consists of a series of pulses, which is denoted as CIR information. The CIR is informative of the different paths a signal takes when traveling from Tx to Rx [29]. The CIR can be modeled as an impulse response $h(t)$ in Equation 2.1, which contains two main components: deterministic multipath components (MPC) and diffuse MPC modeled as Additive Gaussian White Noise [7] [30]. The first term represents deterministic MPC, accounting for UWB signal transmission across superimposed multiple paths with delays τ_i , amplitudes A_i and total path number L . In an indoor environment, deterministic MPC is always caused by the wave reflection from walls or other objects in addition to the desired target(s). The second term, representing diffuse MPC, captures channel disruptions modeled as Additive White Gaussian Noise (AWGN), $v(t)$, for accurate channel

behavior description. Therefore, the unwanted signal components affecting the received CIR can be classified as MPC and AWGN in the UWB signal model. These two contributions will be suppressed in section 3.3.

$$h(t) = \sum_{i=1}^L A_i \delta(t - \tau_i) + v(t) \quad (2.1)$$

Figure 2.2 illustrates the difference of CIR samples transmitted from Tx to Rx under two scenarios: Background only CIR (orange line) and CIR with a target (blue line). The CIR samples of this figure are plotted using experimental data from measurements of the office room used in this thesis and using Nolvelda X4 UWB radar and related real position of the target person. The background CIR represents the case where no object is present in front of the radar, whereas the CIR with a target indicates the presence of an object in the field of view of the radar. By comparing the amplitudes of these two scenarios, the time delay of the direct signal between Tx and Rx (the first path of the CIR measurement) and the time delay caused by the reflection from the target object can be identified as peak in target position. For simplicity, the time delay caused by the target's reflection is denoted as τ_{TP} . Additionally, fluctuation signals occurring after τ_{TP} are often attributed to MPC and AWGN.

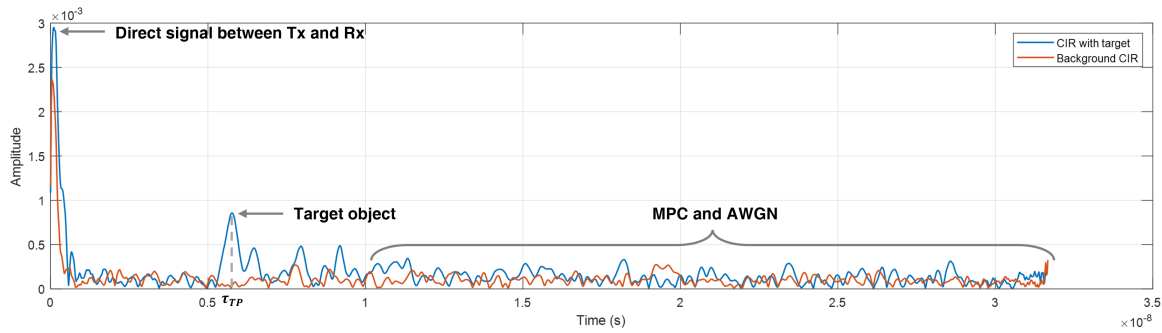


Figure 2.2: Illustration of a CIR sample collected with a UWB radar in two scenarios. Scenario 1 (orange line): no target in front of the radar (background only). Scenario 2 (blue line): one target in front of the radar (CIR with target).

In the signal processing pipeline, the CIR collected over multiple time frames can be stacked together into a 2D matrix to form a range time plot, as illustrated in Figure 2.3.

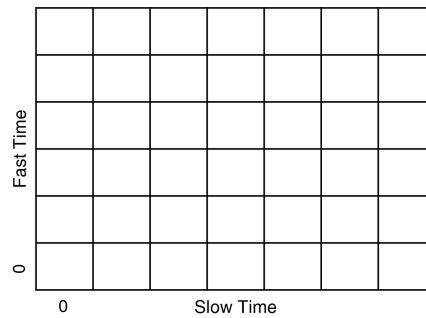


Figure 2.3: Illustration of range time plot generated by stacking together multiple CIR samples collected over time.

In this matrix representation, each column corresponds to a CIR captured at a specific slow time frame (the time index when the CIR was recorded). Each row represents a fast-time sample, which corresponds to a specific time delay within the CIR. This visualization provides an intuitive way to observe temporal evolution and motion-related changes in signal reflections, and is typically the starting point of many radar processing algorithms in the literature.

2.2. Monostatic UWB Radar

This section introduces monostatic UWB radar and its localization principle using time delay measurements in ceiling-mounted setups.

2.2.1. Definition

Radar configurations generally require at least two antennas—one for transmitting the UWB pulses and one for receiving them. Based on antenna placement, radar systems are typically classified into three categories: monostatic, bistatic, and multistatic [31]. Figure 2.4 illustrates the differences among these three configurations.

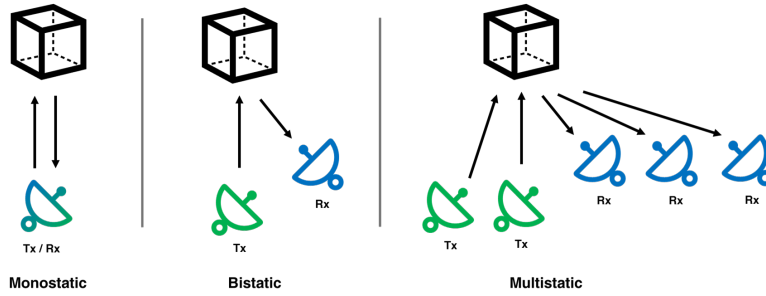


Figure 2.4: Example of three different radar configurations: monostatic, bistatic, and multistatic. The multistatic radar configuration can be different compared with the example one shown in the plot, but there should be at least three nodes.

In a monostatic radar system, the transmitting and receiving antennas are co-located, often sharing the same physical antenna. This setup simplifies hardware design and ensures precise synchronization between transmission and reception processes. Bistatic radar, on the other hand, separates the transmitter (Tx) and receiver (Rx) by a considerable distance, which can provide flexibility in system layout and is particularly useful in occlusion situations or passive sensing scenarios. Multistatic radar involves multiple transmitters and/or receivers placed at different locations, typically separated by distances large compared to the antenna size [32]. This configuration enables spatial diversity but significantly increases system complexity.

Compared to bistatic and multistatic setups, monostatic radar offers the advantages of a fully known transmitted waveform, inherent time synchronization between Tx and Rx, and the potential for phase coherence within each node [33] [34]. These characteristics lead to a simplified system architecture with lower geometric complexity, making monostatic radar especially suitable for practical ceiling-mounted deployment, such as integration into office lighting system [33].

2.2.2. Localization using the Monostatic UWB Radar

In monostatic UWB radar localization, the estimation of a target position can be performed based on the time delay τ_{TP} defined in previous sections. Figure 2.5 illustrates the concept using a ceiling-mounted radar setup, where blue points A, B, and C represent the positions of three radars, and the red point indicates the target.

In this geometry, τ_{TP} denotes the time delay for the UWB pulses to go from the transmitter to the target and back to receiver. The corresponding total traveled distance is therefore given by $c \cdot \tau_{TP}$, where c is the speed of light. The geometric relationship can be expressed in Equation 2.2, where $\mathbf{P}_T$ is target position, $\mathbf{P}_{R_m}$ is the radar m position and r is the distance from radar to the target.

$$r = \|\mathbf{P}_T - \mathbf{P}_{R_m}\|_2 = c \cdot \tau_{TP_i} \quad (2.2)$$

In a three-dimensional space, assuming all radars are mounted on the ceiling at the same level, which forms a horizontal plane, each τ_{TP} measurement defines a hemisphere centered at the corresponding radar location. Since the target is located below the ceiling in an indoor environment, it must lie on the surface of that hemisphere. As shown in Figure 2.5, using multiple monostatic radars results in multiple hemispheres, represented in different colors, and the intersection of these hemispherical surfaces corresponds to the estimated target position.

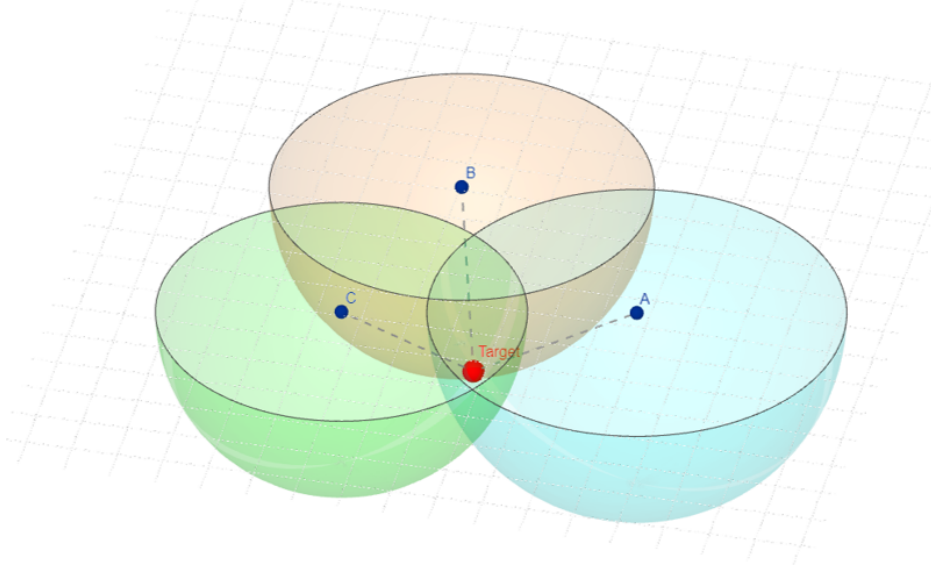


Figure 2.5: Illustration of localization process by using multiple (i.e., 3) monostatic UWB radar nodes.

2.3. Related Work in the Literature

In this section, related works relevant to this thesis are reviewed and categorized into two main areas: human tracking and activity recognition, which correspond to the two core components of this research.

2.3.1. Tracking

Since, to the best of the author's knowledge, no existing studies have explored human tracking using downward-facing, ceiling-mounted monostatic UWB radars, the literature review is extended to cover two related directions:

1. Human tracking using wall-mounted UWB radars (including all radar configurations)
2. Ceiling-mounted radars based on other technologies for indoor tracking applications.

Wall-Mounted UWB Radar for Human Tracking

Target tracking frameworks are generally categorized into two approaches: Detect-Before-Track (DBT) and Track-Before-Detect (TBD) [35]. In a typical DBT pipeline, radar data is processed on a per-frame basis, and target detection is first performed using a fixed threshold—commonly based on a Constant False Alarm Rate (CFAR) detector [36]. Once a target is detected, its position and motion information are passed to a tracking algorithm for state estimation. In contrast, the TBD approach adopts a fundamentally different strategy. It skips the explicit detection step and directly operates on raw or low-threshold radar measurements, jointly performing detection and tracking within a unified framework over multiple frames [36] [37]. In other word, DBT is efficient and simple but prone to missed detections in low SNR or weak target scenarios, while TBD improves robustness through spatiotemporal integration at the cost of higher complexity.

In the following review, related works on human tracking using wall-mounted UWB radar systems are grouped

according to these two categories for analysis.

1. Detect-Before-Track:

As previously described, the DBT approach follows a typical processing pipeline, which is illustrated in Figure 2.6. Although the specific implementations may vary across studies, especially in the choice of algorithms for each processing block, the overall structure remains consistent. In general, targets are first detected independently in each radar node if multiple of them are present, and the resulting detections are then combined to perform localization and tracking.

In [11], the authors employ a multistatic UWB radar system installed approximately 1 meter above the floor and propose a two-stage algorithm for human tracking. In their pipeline, a cell-averaging constant false alarm rate (CA-CFAR) detector is first used to identify human targets. Since the radar is not ceiling-mounted, the human body is considered as an extended object, and the resulting detections are spatially scattered. To address this, an averaging filter is applied to estimate the center of the scattered points. A Kalman filter (KF) is then used to track the estimated center in each radar individually, providing a more accurate range estimation. After obtaining range measurements from all radars, a Taylor series positioning algorithm is applied to estimate the target's 2D position. Finally, a second KF is used to track the target in the 2D space, resulting in a smoothed trajectory achieving a Root Mean Squared Error (RMSE) of 0.05-0.1 m in the simulation data. The results using a real dataset also demonstrate that incorporating an individual KF for each radar reduces tracking error compared to the pipeline that skips this step.

[29] also adopts a DBT-based approach to track a single person using a multistatic UWB radar system comprising four UWB units. In their pipeline, CIR variance is recursively calculated as a pre-processing step to enhance range resolution. Then, a leading-edge detection algorithm is employed to estimate the target's range from each radar individually. Finally, a particle filter (PF) is applied to jointly perform localization and tracking, achieving a RMSE of 0.33 meters.

In [15], the authors propose a multi-person tracking DBT framework based on a multistatic UWB radar setup. To enhance target detectability in cluttered environments, the antennas are mounted near the ceiling at a height of 2.5 meters and oriented horizontally. This configuration helps mitigate mutual shadowing among people, thereby increasing spatial diversity. The signal processing pipeline begins with background subtraction, followed by target detection using CFAR. Subsequently, a TOA-based matching algorithm is employed to associate range measurements across receivers and distinguish multiple individuals. To address spatial bias introduced by the elevated antenna placement, a height compensation algorithm is applied. The final 2D positions are estimated using ellipse-based localization, and multi-target tracking is performed with a Kalman filter. Compared to systems with lower antenna placements and earlier TOA association methods, the proposed approach improves tracking completeness and accuracy, achieving a mean localization error of approximately 0.26 m and a correct localization rate of 72% for multi-person tracking.

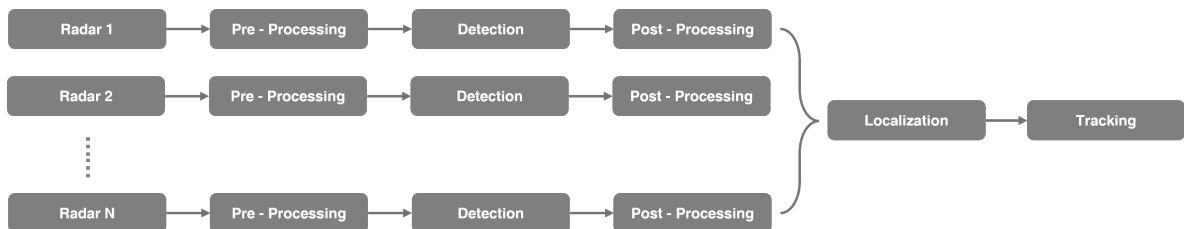


Figure 2.6: Typical pipeline for a Detect-Before-Track approach.

2. Track-Before-Detect:

Unlike DBT, which detects targets independently in each radar, TBD jointly processes signals from all radars.

This approach is particularly effective in handling weak target scenarios, as it accumulates radar echoes across all radars to enhance tracking accuracy [10].

[38] proposed a variance-based temporal-spatial (VATS) mapping method for human tracking using a multistatic UWB radar system, based on the public dataset from [29]. The initial step in their pipeline follows the recursive variance computation method introduced in [29], where the variance of the CIR is calculated for each Tx-Rx pair to enhance range resolution. Unlike traditional methods that rely on leading-edge detection to extract range measurements per radar pair, VATS directly explores the spatial correlation between the CIR variance matrix and the potential target location. Specifically, the experimental area is discretized into a 100×100 2D grid, where each grid cell corresponds to a physical position in the room. For each CIR measurement, the CIR variance values from all radar pairs are projected onto the grid accordingly. This results in a radiotomographic imaging (RTI) matrix, where higher values indicate a higher likelihood of the target's presence. This likelihood map is then used as the measurement input to a PF. The PF assigns higher weights to particles located in high-value regions of the RTI map, and the weighted average of these particles is taken as the estimated position of the human target. Compared to the method in [29], which uses leading-edge detection and PF directly on range estimates, the VATS method achieves improved performance, with a median localization error of approximately 0.16 m.

In [10], the authors propose a dynamic programming-based TBD method for both single- and multi-person tracking using a UWB radar network. After clutter suppression on the CIR data from each radar, all CIR values from every time frame are mapped onto a common 2D grid, where the x- and y-axes correspond to the physical dimensions of the experimental area. In this representation, high values indicate likely target locations. By stacking these 2D maps over time frame, a 3D spatiotemporal grid is constructed, with the z-axis representing time frames. Within this 3D grid, clustering and morphological opening are applied to enhance and isolate potential target scatter points. These points are then used to generate and associate short trajectories, allowing for the tracking of both single and multiple individuals. By jointly analyzing multiple scan frames using 3D image processing techniques such as region growing and morphological filtering, the system enhances weak target trajectories while suppressing noise and clutter artifacts. After identifying high-score regions, the system further refines them using 2D smoothing and local maxima detection to generate measurement points. These measurements are grouped into short track segments, which are later associated through a multiple hypothesis tracking framework. Outlier removal and trajectory smoothing are applied in the final stage to generate clean and continuous tracks. In contrast to other TBD approaches that utilize conventional tracking methods such as particle filtering (PF) [38], the method proposed in [10] operates directly on raw waveform data. It constructs an image-like score map through a multi-sensor voting mechanism, and then applies image processing-based tracking techniques to extract the target trajectory. According to the result, this approach achieves a mean tracking error around 0.2 to 0.4 meters.

The key feature of the studies discussed in this section are summarized in Table 2.1.

Table 2.1: Summary of Wall-Mounted UWB Radar-Based Human Tracking Approaches.

Reference	Type	Initial Processing	Tracking	Results
[11]	DBT	CFAR	KF	RMSE: 0.05m-0.1m
[29]	DBT	Leading-edge	PF	RMSE: 0.33m
[15]	DBT	CFAR	KF	Mean: 0.26m
[38]	TBD	Grid Mapping	PF	Median: 0.16m
[10]	TBD	Grid Mapping + Clustering	Morph. Filtering	Mean: 0.2–0.4m

Ceiling-Mounted Non-UWB Radars for Indoor Tracking

When the scope of the literature review is extended from UWB to other radar technologies for ceiling-mounted tracking, several additional studies can also be reviewed. The mm-Wave radar, equipped with multiple antennas (i.e., multiple transmitters and receivers within a single device, i.e. MIMO), can measure CIR across multiple antenna pairs. This allows the possibility to estimate both elevation and azimuth angles of the target. By combining angle-of-arrival (AoA) information with range estimates derived from CIR, a single mm-Wave radar is potentially capable of localizing human targets within an indoor environment.

For example, [18] and [39] both utilize ceiling-mounted mm-Wave radars for indoor human tracking. In [18], a downward-facing mm-Wave radar is installed at a height of 3 meters to track multiple people. The authors apply a CFAR detector and the Minimum Variance Distortionless Response (MVDR) algorithm to convert CIR data into spatial point clouds. These point clouds are then processed using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to distinguish between people. An alpha-extended Kalman filter is subsequently used to track group targets, followed by a track management algorithm that handles challenges such as track splitting, crossover, and near-static targets. In [39], the authors mount a mmWave radar on a motor installed on the ceiling. This mechanism enables rotation in both elevation and azimuth, significantly expanding the radar's effective coverage area. In detail, the radar first extracts 4D point clouds (space + velocity) through CFAR, range-Doppler processing and beamforming, followed by human localization via centroid estimation or micro-motion analysis. A rotation decision module then evaluates the target's view angle and adjusts the radar orientation using a calibrated pan-tilt transformation to ensure optimal sensing performance across tasks. Based on these processes, the system can dynamically reorient the radar toward the moving target, enabling robust tracking in larger spaces and supporting multiple end applications such as vital sign monitoring and fall detection.

The mentioned studies in this section are summarized in Table 2.2.

Table 2.2: Summary of Ceiling-Mounted Non-UWB Radar Based Human Tracking Approaches.

Reference	Radar Setup	Initial Processing	Tracking	Highlights
[18]	Ceiling-mounted mmWave radar	CFAR + MVDR + DBSCAN	Alpha-KF + Track Management	Crowd tracking
[39]	Ceiling-mounted mmWave radar with motor	CFAR + Angel Estimation	Directional Steering + Centroid estimation	Expands coverage area, Multi applications

2.3.2. Activity Recognition

Based on the problem definition of this thesis, activity recognition should be performed along the human walking trajectory. This implies that multiple activity types may occur sequentially, and human activities form a continuous stream in which one motion follows another. Such a setting can be regarded as a radar-based continuous activity recognition problem, as opposed to the traditional setting where one radar data segment corresponds to a single activity class [40]. In [41], the authors address this continuous recognition challenge by employing a sliding window approach. Specifically, they first apply Short-Time Fourier Transform (STFT) to the radar data and then use a Convolutional Neural Network (CNN) and ResNet to classify activities within each window. Their system, which uses five monostatic UWB radars, successfully recognizes nine different human activities in a continuous sequence with best accuracy 95.65%.

Not only in radar-based continuous activity recognition, but also in non-continuous activity recognition tasks, machine learning (ML) methods dominate the current research landscape. These approaches typically involve training

a model (i.e., a neural network) on preprocessed radar datasets and validating it on a separate test set [7, 40].

For instance, [12] demonstrates that a ceiling-mounted monostatic UWB radar can detect human falls with an accuracy of 93.97% using CNN. In another study, [42] demonstrates an extended target intensity maps and velocity distribution reconstruction method. The authors firstly map the radar signal from each node into a common 3D Cartesian voxel grid, where each voxel accumulates the reflection intensity across Doppler channels. Then, a dominant velocity vector is reconstructed per voxel by solving an overdetermined inverse problem using Doppler projections from multiple radar nodes. These fused intensity and velocity fields are subsequently used as input to CNN-BiLSTM for human activity classification. Experimental results on a public UWB radar dataset from [43] involving 14 subjects and nine activity classes show that the proposed fusion method achieves a test accuracy of 87%. These findings suggest that the extended target phenomenon, often regarded as a challenge in radar-based tracking, can instead be exploited as a useful feature for activity recognition.

However, due to time constraints, it is not feasible to collect a sufficiently large dataset to support ML based activity recognition with deep neural networks in this thesis. Given that the range of activities to be classified is limited to walking, standing, and sitting, several non machine learning approaches can be considered. For instance, [44] demonstrates that the Moving Target Indicator (MTI) can be used to detect human motion, effectively distinguishing between dynamic movements and static postures.

[45] and [46] both propose height-based thresholding methods for radar-based activity classification achieving accuracy 76% and 93.7%, respectively. These methods leverage temporal changes in estimated human height to distinguish between activities such as standing, sitting, and lying. It is worth highlighting that in [46], the authors deploy a single ceiling-mounted mmWave radar to achieve both human tracking and activity classification. By focusing on the vertical dimension of the point cloud, their system estimates the subject's height in real-time and applies empirically determined thresholds to infer posture transitions. The top-view configuration reduces occlusion and produces more centralized and reliable point cloud data, particularly around the head and upper body, which enhances both tracking robustness and posture estimation accuracy. Their experimental results demonstrate high sensitivity and precision across standing, sitting, and lying postures, validating the effectiveness of non-ML, height-based classification in compact indoor environments.

The key features of the research paper mentioned in this section are listed in Table 2.3.

Table 2.3: Summary of Radar-Based Activity Recognition Approaches.

Reference	Method	Activities	ML-based	Result
[41]	CNN / ResNet	9 continuous activities	Yes	Accuracy: 95.65%
[12]	CNN	Fall Detection	Yes	Accuracy: 93.97%
[42]	CNN-BiLSTM	9 activities	Yes	Accuracy: 87%
[44]	MTI Detection	Move / Static	No	N/A
[45]	Height-based thresholds	Sit / Stand / Lying	No	Accuracy: 76%
[46]				Accuracy: 93.7%

2.4. Summary

This chapter has introduced the fundamental principles of UWB radar-based human localization and activity recognition. It began by describing the characteristics and propagation behavior of UWB signals, followed by a detailed explanation of UWB localization approaches, including both active and passive (radar-based) methods. The concept

of CIR was introduced as the core data representation used throughout the signal processing pipeline. The monostatic UWB radar configuration, with its simplified hardware and inherent synchronization benefits, was described in the context of ceiling-mounted indoor tracking.

Finally, a comprehensive review of related work was conducted, covering DBT and TBD tracking frameworks with wall-mounted UWB radars, ceiling-mounted non-UWB radar systems, as well as both ML-based and non-ML-based activity recognition approaches. The literature indicates that TBD methods, by combining all available radar information before the detection stage, offer a potential solution to the weak target problem, while human-height-based activity classification provides a feasible option in non-ML scenarios. However, the challenges of extended targets and viewpoint variation remain unresolved in these studies and thus require further research in this thesis.

In summary, this technical background serve as the basis for the system design and algorithm development presented in the following chapters.

Localization and Tracking

In this chapter, the localization and tracking pipeline of the proposed 3D grid-based method is presented, including system modeling, detailed grid mapping, compensation, and tracking procedures. In addition, the baseline method used for performance comparison is also described.

3.1. System Model

In this section, the mathematical model of the proposed pipeline is presented, including the method for exploiting the extended target characteristics of a human to enable tracking using ceiling-mounted UWB radar.

3.1.1. Extended Target Modeling

According to the problem statement, the monostatic UWB radar is mounted on the ceiling, which introduces the extended target issue as discussed in section 1.2.3. When an person with normal body walks below the radar, his body acts as an extended target from the radar's top-down perspective. As shown in Figure 3.1, different parts of the human body (head, torso, and lower limbs) contribute multiple reflection peaks that are detected by the radar. These reflections are illustrated as red dots in Figure 3.1. Furthermore, as the person walks, the precise locations of these reflection points vary over time and are not fixed relative to the body (so-called 'non-rigid body').

In [47], ceiling-mounted acoustic localization devices are designed to track the human head. Indeed, the head is often the closest body part to the ceiling-mounted UWB radar and generally produces the strongest signal reflection. However, treating the entire human target as the head only neglects useful reflections from other body parts, such as the torso and legs, that can be useful for later activity classification. This limitation becomes especially relevant in scenarios with restricted radar opening angles, where the head may fall outside the radar's coverage (see Section 1.2.3 for details). Therefore, it can be beneficial to also incorporate reflections from other body regions.

To make full use of the extended target characteristics, a *change in observation perspective* is required. When a normal person walks upright with both feet, without exhibiting significant vertical oscillations, his body is aligned vertically with respect to the floor. This causes the radar reflections to be distributed approximately along the vertical (z) axis. While the side view shown in Figure 3.1 highlights this vertical spread, a top-down perspective, as illustrated in Figure 3.2, reveals that these reflection points cluster spatially around the actual position of the person. The center of this cluster can thus be used to estimate the person's location in the horizontal plane (xy plane), which

is the expected output of the localization system.

Based on this idea, we assume that a normal person is more likely to be located at a horizontal position (x, y) if multiple reflection points are consistently observed along the vertical (z) axis at that location and there is no extra moving non-human target in the room. In other words, a high concentration of vertically aligned reflections suggests a higher likelihood of human presence at the corresponding (x, y) position. Thus, when the radar's detection space is discretized into multiple horizontal planes, the human body tends to produce reflections across several of these planes at the same (x, y) coordinate. Therefore, locations where high reflection probabilities are observed across all I horizontal layers in a specific (x, y) location are more likely to correspond to the actual position of the person.

To formalize this physically-grounded assumption, we introduce a joint likelihood formulation to quantify the probability of target-induced reflections occurring along a vertical line at a specific (x, y) location. Specifically, let $\mathbf{R}_{x,y,z_i}$ denote the reflection probability caused by the target at spatial coordinate (x, y, z_i) , where z_i denotes the horizontal layer i in the 3D space. The probability of human presence at location (x, y) can then be expressed as the joint distribution showed in Equation 3.1.

$$P(\mathbf{R}_{x,y,z_1}, \mathbf{R}_{x,y,z_2}, \dots, \mathbf{R}_{x,y,z_n}) = \prod_{i=1}^I \mathbf{R}(x, y, z_i) \quad (3.1)$$

This distribution reflects the likelihood that reflections are observed at all horizontal layers corresponding to a given (x, y) position. The human location estimate (x^*, y^*) is defined as the position with the maximum joint probability. Assuming independence of reflections across different layers, the expression can be found in Equation 3.2.

$$(x^*, y^*) = \arg \max_{(x,y)} P(\mathbf{R}_{x,y,z_1}, \mathbf{R}_{x,y,z_2}, \dots, \mathbf{R}_{x,y,z_n}) = \arg \max_{(x,y)} \prod_{i=1}^I \mathbf{R}(x, y, z_i) \quad (3.2)$$

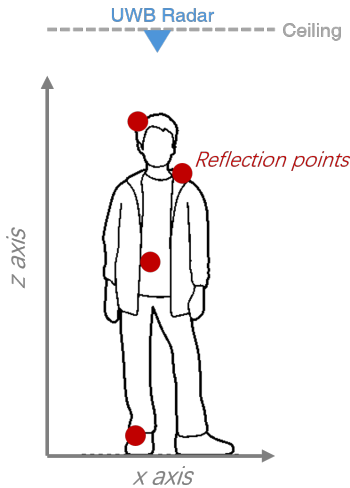


Figure 3.1: Visualization of the human body as an extended target from the side view.

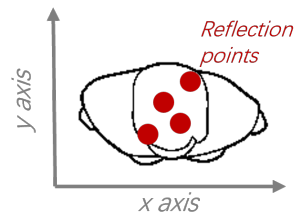


Figure 3.2: Visualization of the human body as an extended target from the top view.

3.1.2. Target Reflection Probability

From the previous analysis, it is essential to determine the probability of reflection caused by the target at a specific position (x, y, z_i) , denoted by $\mathbf{R}(x, y, z_i)$. Typically, the target reflection in the CIR obtained from a single UWB

radar appears as observable peaks. However, nonzero amplitudes in the CIR can also occur due to environmental multipath effects and noise, even without a true target at the given range. Consequently, the CIR amplitude at range r , denoted as $\mathbf{c}(r)$, can be interpreted as proportional to the probability of a target reflection occurring at range r , expressed as $P_{\text{reflection}}(r)$. This relationship is formulated in Equation 3.3. Hence, a higher value of $\mathbf{c}(r)$ indicates a higher probability of the target being present at range r .

$$P_{\text{reflection}}(r) \propto \mathbf{c}(r) \quad (3.3)$$

Considering the deployment of M multiple UWB radars, each radar m provides its own CIR, reflecting the possibility of target existence at its corresponding range r_m . According to the monostatic radar localization principle detailed in Section 2.2.2, the intersection points of hemispheres derived from multiple radars represent potential target locations. Specifically, the intersection point in space with a high reflection probability from all radars can be regarded as the target's estimated location. Thus, the joint reflection probability at an arbitrary position (x, y, z_i) can be expressed as the joint likelihood of reflections from all radars, given by Equation 3.4, where the distance $r_{m,i}$ from the position (x, y, z_i) to radar m is computed in Equation 3.5.

$$\mathbf{R}(x, y, z_i) = \prod_{m=1}^M P_{\text{reflection},m}(r_{m,i}) \propto \prod_{m=1}^M \mathbf{c}_m(r_{m,i}) \quad (3.4)$$

$$r_{m,i} = \|(x, y, z_i) - (x_m, y_m, z_m)\|_2 \quad (3.5)$$

Therefore, the likelihood function $L(x, y)$ of human presence at a horizontal position (x, y) can be defined in Equation 3.6. Here, i indexes the horizontal layers along the vertical (z) axis, with a total of I layers, and m indexes the multiple deployed UWB radars, with a total of M radars. $r_{m,i}$ is the distance between the point (x, y, z_i) and the m -th radar is given in Equation 3.5. According to Equation 3.6, the likelihood function combines the reflection probabilities from all radars across multiple horizontal layers and multiple radars. These probabilities are assumed to be proportional to the CIR amplitudes at the corresponding ranges. Thus, the estimated human location (x^*, y^*) is obtained by maximizing the likelihood function $\mathbf{L}(x, y)$ which is expressed in Equation 3.7.

$$\mathbf{L}(x, y) = \prod_{i=1}^I \mathbf{R}(x, y, z_i) = \prod_{i=1}^I \prod_{m=1}^M P_{\text{reflection},m}(r_{m,i}) \propto \prod_{i=1}^I \prod_{m=1}^M \mathbf{c}_m(r_{m,i}) \quad (3.6)$$

$$(x^*, y^*) = \arg \max_{(x,y)} \mathbf{L}(x, y) \quad (3.7)$$

Based on the modeling presented above, the target's position can be estimated by leveraging the extended target characteristics of a walking human in an indoor office environment. Accordingly, the problem shifts to two key questions:

1. How to accurately estimate the target reflection probability $\mathbf{R}(x, y, z_i)$ at each spatial location?
2. How to robustly identify the maximum point of the resulting likelihood function for precise localization?

3.2. Overall Pipeline of the Proposed 3D Grid-Based Method

Based on the problem statement, a radar signal processing pipeline should be designed to jointly perform human tracking and activity recognition. In this thesis, a 3D grid-based framework is proposed for achieving these tasks using ceiling-mounted monostatic UWB radar. The overall pipeline of the proposed method is illustrated in Figure 3.3, which comprises three main components:

- The first two components are introduced in detail in this chapter, while the third component, activity recognition, is discussed in Chapter 4.



Next, the *localization and tracking module* maps the preprocessed CIRs into a 3D voxel grid using the 3D grid mapping algorithm to formulate the target reflection probability $\mathbf{R}(x, y, z_i)$, generating M CIR grids (one per radar) at each slow-time index t . For each grid, layer compensation is first applied to remove artifacts and enhance target-related features. The CIR grids from the M radars at the same slow-time index are then fused into a single 3D grid based on Equation 3.4 to obtain target reflection probability grid $\mathbf{R}_t(x, y, z_i)$. To further improve the estimation accuracy, Position Dilution of Precision (PDOP) compensation is applied to $\mathbf{R}(x, y, z_i)$ in subsection 3.6.3. Finally, according to Equation 3.6, the compensated target reflection probability grid is projected onto a 2D likelihood map that represents the probability of human presence at each horizontal location (x, y) . The position corresponding to the maximum likelihood is estimated and tracked using a Particle Filter (PF), following a tracking strategy similar to that in [38]. Among these procedures, the 3D grid mapping algorithm is described in detail in Section 3.4. Since layer compensation and PDOP compensation are additional steps aimed at improving estimation accuracy, the CIR grid fusion and 2D projection processes are discussed first in Section 3.5, followed by the two compensation procedures in Section 3.6. The PF tracking method is presented in Section 3.7.

Simultaneously, the *activity recognition module* operates on the CIR output after clutter suppression. Inspired by the MTI detection concept described in [44], this module detects motion status (walking vs. static). Only periods labeled as walking are input to the PF for tracking. For static segments, which correspond to either standing or sitting activities, the system applies a height estimation algorithm to differentiate between these two classes. This approach is motivated by the observation that human height changes significantly (approximately 20% of human height) during sit-to-stand or stand-to-sit transitions [46]. If such a height variation is detected, the activity is classified to sitting activity. Through this integrated framework, both spatial tracking along with temporal activity classification are achieved without reliance on machine learning models.

3.3. Signal Preprocessing

To eliminate the influence of clutter caused by MPC in the CIR, an adaptive clutter algorithm based on the method shown in [48] is applied to the data. Here, the term ‘adaptive’ means that the application ratio $\mathbf{a}_t(k)$ is adaptive to change in different slow-time index compared to the fixed value in [49]. The process of adaptive clutter suppression (ACS) can be represented by Equation 3.8 [48]. In Equation 3.8, t denotes the index of the slow-time bin, k denotes the index of the fast-time bin, $\mathbf{c}_t(k)$ is the received CIR measurement, $\mathbf{s}_t(k)$ is the clutter signal, and $\mathbf{y}_t(k)$ is the processed signal. $\tilde{\mathbf{c}}_t(k)$ and $\tilde{\mathbf{y}}_t(k)$ represent the normalized envelopes of $\mathbf{c}_t(k)$ and $\mathbf{y}_t(k)$, respectively. The envelope is obtained from the Hilbert transform of $\mathbf{c}_t(k)$ and $\mathbf{y}_t(k)$. The application ratio $\mathbf{a}_t(k)$ is updated based on the value of the envelope.

$$\begin{aligned} \mathbf{y}_t(k) &= \mathbf{c}_t(k) - \mathbf{s}_t(k) \\ \mathbf{d}_t(k) &= \frac{\min(\tilde{\mathbf{c}}_t(k), \tilde{\mathbf{y}}_t(k))}{\tilde{\mathbf{c}}_t(k)} \\ \mathbf{a}_t(k) &= a_{\min} + (a_{\max} - a_{\min}) \cdot \mathbf{d}_t(k) \\ \mathbf{c}_{t+1}(k) &= \mathbf{a}_t(k) \cdot \mathbf{s}_t(k) + (1 - \mathbf{a}_t(k)) \cdot \mathbf{c}_t(k) \end{aligned} \quad (3.8)$$

A high application ratio improves signal-to-noise ratio (SNR) but delays accurate clutter signal capture, risking inaccurate target movement representation. Conversely, a low application ratio quickly extracts moving targets but suppresses both clutter and target signals, potentially leading to insufficient SNR and missed detections [48]. The variable $\mathbf{d}_t(k)$ is used to determine the application ratio for the next scan, using the envelope of two signals. If there is motion at a certain delay bin k_0 , $\tilde{\mathbf{y}}_t(k_0)$ would be greater than $\tilde{\mathbf{c}}_t(k_0)$ at that point. Therefore, $\mathbf{d}_t(k_0)$ would be one and $\mathbf{a}_t(k_0)$ would have the value $a_{\max}$ to maximize SNR. Conversely, $\mathbf{d}_t(k)$ has a lower application ratio according to the ratio of $\tilde{\mathbf{c}}_t(k)$ and $\tilde{\mathbf{y}}_t(k)$ at points with little or temporary motion [48]. By using this method, $\mathbf{a}_t(k)$ can be adjusted based on different circumstances. During the iterative period, some slow-time bins may contain all zeros due to measurement errors. In these zero slow-time bins, the ACS method is skipped, and the value in the zero slow-time bins is set equal to the last slow-time bin with a valid value. By balancing sufficient SNR and signal delays, $a_{\max} = 0.75$ and $a_{\min} = 0.25$ are selected in this thesis and applying the ACS along with normalization, the example filtered data¹ is presented in Figure 3.4.

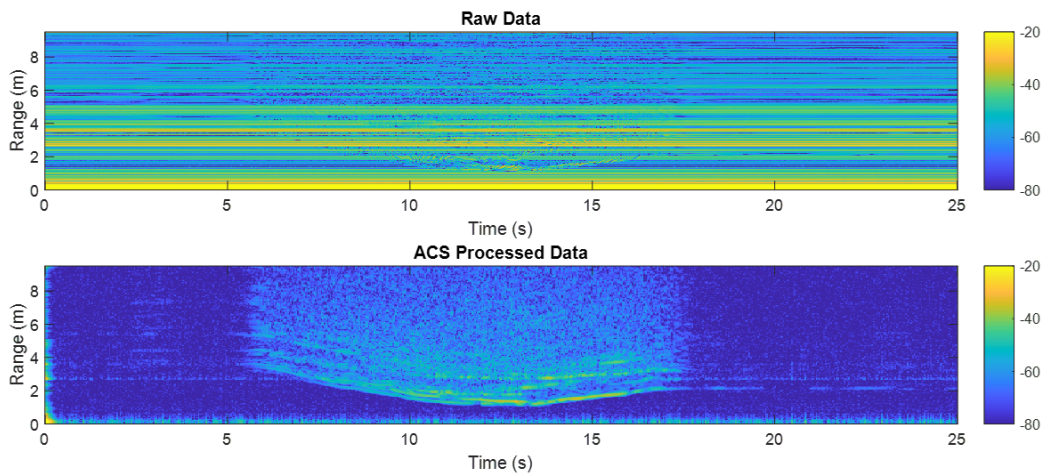


Figure 3.4: An example result of raw data vs ACS using the dataset from test ID *VH_T_W_P1_H1*.

Compared with the original range-time plot with massive MPC, nearly all clutter is discarded and the wave pattern

¹The data used to illustrate this ACS example is obtained from Radar 1 during an experiment. The corresponding test ID for this dataset is *VH_T_W_P1_H1*. A detailed description of the experimental setup will be provided later in Chapter 5.

due to target movement is easier to distinguish. Based on the signal preprocessing stage, a cleaner signal with dominant MPC suppressed is obtained. This pre-processed signal provides a solid foundation for subsequent processing stages.

3.4. CIR 3D Grid Mapping

To compute the target reflection probability grid $R(x, y, z_i)$ at each horizontal layer z_i for all corresponding (x, y) positions, the processed CIR data from each time frame and radar must be mapped into a 3D voxel grid. This grid serves as a discretized representation of the physical room space, enabling the CIR data to be spatially visualized and interpreted in three dimensions.

To map the CIR data into a 3D CIR grid, the gridding strategy is adapted from the method proposed in [10], which originally mapped CIR measurements onto a 2D grid. In this thesis, the original method is extended and enhanced to support 3D spatial mapping. This improvement allows the CIR data to be projected into a voxel-based representation, capturing not only the horizontal (x, y) distribution, but also the vertical dimension z , thereby enabling a more comprehensive spatial interpretation of the radar reflections.

The pipeline for mapping the processed CIR data of each radar at every slow-time frame into a 3D voxel grid is illustrated in Figure 3.5. This strategy consists of three main steps, further detailed in the following sub-sections: CIR score calculation, voxel range determination, and CIR grid construction. First, the CIR score is computed by applying a power-law normalization to the raw CIR amplitudes. This normalization suppresses high amplitude variations and stabilizes the distribution of CIR range bins. Next, the voxel range determination utilizes known room geometry, radar position and grid size parameters to determine which CIR range bins correspond to specific voxels in the 3D space. Finally, the CIR grid construction step maps the CIR scores with their corresponding voxels, resulting in a voxelized representation of the spatial distribution of CIR for each radar and time frame.

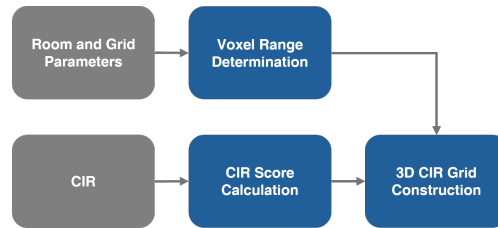


Figure 3.5: Pipeline of mapping CIR onto a 3D grid.

3.4.1. CIR Score Calculation

To enhance target-related reflections, a power-law normalized CIR score is computed following the formulation proposed in [10], as shown in Equation 3.9. Here, $\mathbf{c}_{m,t}(k)$ denotes the CIR amplitude of the k -th fast-time bin (i.e., range bin) from radar m at slow-time frame t . $\mathbf{V}_{m,t}$ is the resulting CIR score, and N_c represents the total number of CIR range bins. The parameter $\alpha \in (0, 1)$ is a power-law compression factor that reduces the influence of high amplitude noise while increasing weaker reflections. In this thesis, $\alpha = 0.5$ is selected to improve weak signal peak amplitude. This normalization enables a more uniform and stable representation of reflection intensity across different CIR range bins.

$$\mathbf{V}_{m,t}(k) = \frac{(\mathbf{c}_{m,t}(k))^\alpha}{\frac{1}{N_c} \sum_{l=1}^{N_c} (\mathbf{c}_{m,t}(l))^\alpha} \quad (3.9)$$

3.4.2. Voxel Range Determination

The room with physical dimensions $\mathbf{R}_{\text{room}} = (R_x, R_y, R_z)$, is discretized into a 3D grid of $\mathbf{N}_{\text{grid}} = (N_x, N_y, N_z)$ voxels, resulting in spatial resolutions of $\Delta x = \frac{R_x}{N_x}$, $\Delta y = \frac{R_y}{N_y}$, and $\Delta z = \frac{R_z}{N_z}$ along the respective axes. To map the CIR score into the 3D voxel grid, it is essential to ensure resolution alignment between the CIR range measurements and the voxelized space. Each CIR range bin corresponds to a certain propagation distance from the radar, and the voxel grid discretizes the room into spatial units. Therefore, voxels that are equidistant from the radar share the same range bin index. Proper mapping requires determining which voxels fall within the propagation range represented by each CIR bin, so that the corresponding CIR score can be accurately assigned to the appropriate voxel locations.

However, the raw CIR typically provides much finer range resolution than needed for human tracking tasks. For instance, the Novelda X4 UWB radar offers a range bin resolution of approximately 0.64 cm, which far exceeds the requirement for indoor tracking [20]. Therefore, instead of directly assigning each CIR bin to a voxel, a downsampling strategy proposed in [10] is adopted to align the CIR range bins with the voxel grid defined by the chosen Δx , Δy , and Δz .

To downsample the CIR for voxel-based mapping, it is necessary to determine which range bins in the CIR correspond to each voxel in the 3D grid. This relation can be determined by the minimum and maximum propagation distances between the radar and each voxel, which correspond to the range interval that a particular voxel occupies with respect to the radar. Figure 3.6 illustrates this concept of ‘voxel range’. The black cube in the figure represents the voxel with side lengths defined by the grid steps ($\Delta x, \Delta y, \Delta z$). Assuming the radar is located outside the voxel, the minimum and maximum distances from the radar to the voxel can be found on the voxel cube surface. These two distances define the range interval associated with the voxel and are used to determine which CIR range bins contribute to this voxel.

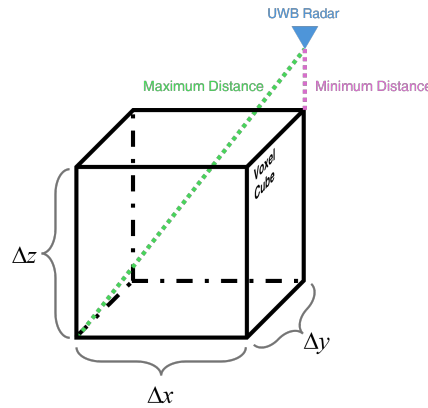


Figure 3.6: Illustration of the voxel size of the 3D grid and its range with respect to the UWB radar.

In detail, for the radar m located at position $\mathbf{p}_{r_m} = (x_{r_m}, y_{r_m}, z_{r_m})^\top$, and a voxel indexed by (n_x, n_y, n_z) , whose spatial extent is bounded in Equation 3.10. Then, the minimum Euclidean distance $\min D_{n_x, n_y, n_z}^m$ from the radar to this voxel is computed in Equation 3.11, where $\min D_{n_x, n_y, n_z}^{(r)} = 0$ if the radar position lies inside the voxel cube. Similarly, the maximum Euclidean distance ($\max D_{n_x, n_y, n_z}^m$) from the radar to the voxel is calculated in Equation 3.12 by identifying the furthest corner.

$$\begin{aligned}
 x_{n_x}^{\min} &= (n_x - 1)\Delta x, & x_{n_x}^{\max} &= n_x \Delta x \\
 y_{n_y}^{\min} &= (n_y - 1)\Delta y, & y_{n_y}^{\max} &= n_y \Delta y \\
 z_{n_z}^{\min} &= (n_z - 1)\Delta z, & z_{n_z}^{\max} &= n_z \Delta z,
 \end{aligned} \tag{3.10}$$

$$\begin{aligned}
d_x^{\min} &= \max(0, \max(x_{n_x}^{\min} - x_{r_m}, x_{r_m} - x_{n_x}^{\max})) \\
d_y^{\min} &= \max(0, \max(y_{n_y}^{\min} - y_{r_m}, y_{r_m} - y_{n_y}^{\max})) \\
d_z^{\min} &= \max(0, \max(z_{n_z}^{\min} - z_{r_m}, z_{r_m} - z_{n_z}^{\max})) \\
\min D_{n_x, n_y, n_z}^m &= \sqrt{(d_x^{\min})^2 + (d_y^{\min})^2 + (d_z^{\min})^2},
\end{aligned} \tag{3.11}$$

$$\begin{aligned}
d_x^{\max} &= \max(|x_{r_m} - x_{n_x}^{\min}|, |x_{r_m} - x_{n_x}^{\max}|) \\
d_y^{\max} &= \max(|y_{r_m} - y_{n_y}^{\min}|, |y_{r_m} - y_{n_y}^{\max}|) \\
d_z^{\max} &= \max(|z_{r_m} - z_{n_z}^{\min}|, |z_{r_m} - z_{n_z}^{\max}|) \\
\max D_{n_x, n_y, n_z}^m &= \sqrt{(d_x^{\max})^2 + (d_y^{\max})^2 + (d_z^{\max})^2}.
\end{aligned} \tag{3.12}$$

Based on these calculations, the voxel range for each voxel at index (n_x, n_y, n_z) in the 3D grid, with respect to radar m , can be defined as the interval $[\min D_{n_x, n_y, n_z}^m, \max D_{n_x, n_y, n_z}^m]$. This distance interval determines which range bins in the CIR are spatially associated with the given voxel, enabling the mapping of CIR scores to corresponding voxel locations in the grid.

3.4.3. CIR Grid Construction

After determining the voxel range for each radar m , defined as the interval $[\min D_{n_x, n_y, n_z}^m, \max D_{n_x, n_y, n_z}^m]$ for every voxel in the 3D grid of size (N_x, N_y, N_z) , the CIR grid construction method proposed in [10] is applied. In detail, for each voxel indexed by (n_x, n_y, n_z) , the maximum value of the CIR score within the corresponding voxel range interval is selected and assigned to that voxel. This process forms the 3D CIR grid $G_{m,t}$ for radar m at slow-time index t , as expressed in Equation 3.13.

$$G_{m,t}(n_x, n_y, n_z) = \max_{k \in [\min D_{n_x, n_y, n_z}^m, \max D_{n_x, n_y, n_z}^m]} V_{m,t}(k) \tag{3.13}$$

This maximum-value assignment strategy is similar to the max pooling operation in a CNN, as it captures the most prominent feature within a given interval. Since high amplitude values in the CIR are typically indicative of target reflections, this voxel-wise maximum extraction effectively preserves target reflection information. Compared to simpler approaches that assign CIR values based on the voxel center's closest distance to a single range bin, this method provides a more robust mapping by aggregating strong reflection features over a bounded spatial interval.

3.4.4. CIR Grid Configuration

Based on the proposed 3D grid mapping method, the CIR data can be effectively assigned to each voxel in the 3D CIR grid while preserving essential reflection features from the target. In this thesis, the 3D voxel grid is defined with dimensions $100 \times 100 \times 8$ for an experimental office room measuring $3.5 \text{ m} \times 6.2 \text{ m} \times 2.7 \text{ m}$. For human tracking in the horizontal (2D) plane, the resulting spatial resolutions along the x - and y -axes are approximately 3.5 cm and 6.2 cm respectively, which are sufficient to meet human tracking requirements.

The choice of the grid size in the vertical (z) dimension is influenced by three main considerations:

1. According to the assumption presented in Section 3.1, reflection points caused by a target are primarily distributed vertically (i.e., aligned with the z -axis). However, due to viewpoint variations across multiple radars (as discussed in Section 1.2.3), exact matching of reflection peaks from different radars to a single spatial

point cannot be guaranteed. A rougher grid resolution in the vertical dimension thus provides redundancy, allowing more accurate capture of reflection features despite slight discrepancies among radars.

2. However, the grid resolution in the vertical dimension cannot be excessively rough, because the vertical distribution of reflection points must be captured adequately in multiple horizontal layer to facilitate the construction of a 2D likelihood map.
3. The vertical grid resolution must also support accurate human height detection for activity classification purposes (detailed in Chapter 4). As described in [46], a height reduction exceeding approximately 20% typically indicates a sitting posture. With a vertical grid size of 8 layers, the resulting vertical resolution is approximately 33.75 cm per voxel. This resolution allows reliable detection of height changes for individuals taller than approximately 1.6875 m. Note that the 20% threshold is an empirical guideline; actual detection sensitivity also depends on practical considerations such as individual postures and chair height, which may affect classification accuracy for significantly shorter individuals sitting on very high chair.

The top voxel layer at $n_z = 8$ corresponds to the height interval 2.3625 m – 2.7 m, which is above normal human height and thus unlikely to contain valid human reflections. Thus, this layer is excluded from following grid computations throughout this thesis. Therefore, in the remaining sections, the voxel grid dimensions will be referenced as $100 \times 100 \times 7$, corresponding to a room size of 3.5 m $\times$ 6.2 m $\times$ 2.3625 m. The second top layer ($n_z = 7$, height interval 2.025 m – 2.3625 m), although higher than normal human height, is retained as a buffer zone. This layer accommodates potential target reflections from very tall persons and provides a transitional zone between the typically occupied layers ($n_z \leq 6$) and the empty top layer ($n_z = 8$).

Based on this grid configuration, the 3D grid of CIR score values for each radar at different slow-time frames can be constructed. Figure 3.7 illustrates the 2D slices $G_{m,t}^{n_z}$ (with n_z denoting each layer of $G_{m,t}$) extracted from the CIR grid $G_{m,t}$ for an experimental data sample with test ID *VH_T_W_P1_H1*, obtained from UWB radar 1 at slow-time index $t = 390$.

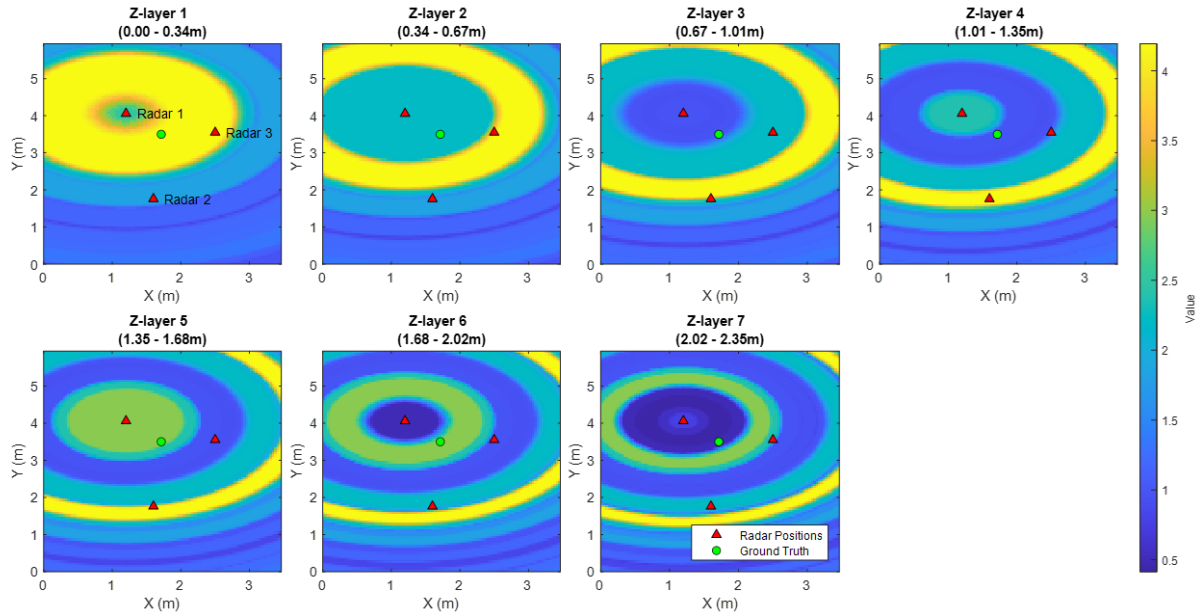


Figure 3.7: An example of 2D slices (along the vertical z -dimension) from the CIR grid is shown for the experimental data sample with test ID *VH_T_W_P1_H1*, obtained from UWB radar 1 at slow-time index $t = 390$.

As the original 3D CIR grid is difficult to visualize in a printed format, all seven z -layer slices are displayed. Each layer is represented as a 100×100 matrix. Green dots indicate the ground truth positions; the explanation of ground

truth positions will be discussed in detail in Chapter 5. Red triangles denote the positions of the three UWB radars, whose exact locations will also be introduced in Chapter 5. The color scale in the plots represents the amplitude of the CIR score. From Figure 3.7, it can be observed that regions of equal value form concentric rings around each radar position, and the ground truth locations mainly lie on the high value rings across nearly all layers.

3.5. CIR Grid Fusion and 2D likelihood Projection

Since the two compensation procedures (i.e., layer compensation and PDOP compensation) shown in Figure 3.3 are additional steps intended to improve the accuracy of target localization, they are skipped in this section. Instead, the focus here is on completing the key steps of likelihood map construction, including the CIR grid fusion and 2D likelihood projection.

3.5.1. CIR Grid Fusion

After the CIR grid $G_{m,t}$ is constructed in Section 3.4, Equation 3.4 serves as the key equation for fusing grids from multiple radars. In practice, the M individual CIR grids are multiplied element-wise and then normalized within each z -layer to ensure that the values in the fused layer represent the target reflection probability. This procedure is expressed in Equation 3.14, where $\tilde{\mathbf{R}}_t$ denotes the reflection grid from M radars, in which higher voxel values correspond to a higher probability of the target's presence in the entire grid at slow-time index t . $\mathbf{R}_t$ represents the target reflection probability grid at slow-time index t representing the target's presence probability in every z -layer, and $\mathbf{P}$ is the PDOP compensation factor tensor, which will be discussed later in Section 3.6.3.

In this section, this compensation step is skipped, and thus tensor $\mathbf{P}$ is set to a tensor containing all '1' values. Through this process, each z -layer from the M radars is fused, and the normalization expressed in the second equation ensures that the resulting values remain within the range $[0, 1]$, allowing them to be directly interpreted as the probability of target presence in that layer.

$$\begin{aligned}\tilde{\mathbf{R}}_t(n_x, n_y, n_z) &= \prod_{m=1}^M G_{m,t}(n_x, n_y, n_z) \\ \mathbf{R}'_t(n_x, n_y, n_z) &= \mathbf{P}(n_x, n_y, n_z) \cdot \tilde{\mathbf{R}}_t(n_x, n_y, n_z) \\ \mathbf{R}_t(n_x, n_y, n_z) &= \frac{\mathbf{R}'_t(n_x, n_y, n_z) - \min_{n_x, n_y} \mathbf{R}'_t(n_x, n_y, n_z)}{\max_{n_x, n_y} \mathbf{R}'_t(n_x, n_y, n_z) - \min_{n_x, n_y} \mathbf{R}'_t(n_x, n_y, n_z)}\end{aligned}\quad (3.14)$$

Figure 3.8 shows the target reflection probability grid $\mathbf{R}_t$ for the test data sample with ID *VH_T_W_P1_H1*, obtained at slow-time index $t = 390$. The green dot indicates the target ground truth, while the red triangles mark the positions of the UWB radars. It can be observed that the probability values around the target are high in most layers, except for the buffer layer ($n_z = 7$). The relatively low probability values in this buffer layer may introduce errors in the following 2D likelihood projection process; this issue will be addressed through compensation in Section 3.6.1.

3.5.2. 2D Likelihood Map Construction

Based on the target reflection probability grid $\mathbf{R}_t$, the 2D likelihood map can be constructed using the joint distribution formulation in Equation 3.6. In practice, the 2D likelihood map $\mathbf{L}_t$ for each slow-time index t is computed according to Equation 3.15. The value of $\mathbf{L}_t$ is proportional to the likelihood of the target being present at a given location on the map.

$$\mathbf{L}_t(n_x, n_y) = \prod_{n_z=1}^{N_z} \mathbf{R}_t(n_x, n_y, n_z) \quad (3.15)$$

After performing the 2D likelihood map construction on the grid shown in Figure 3.8, the resulting likelihood map is presented in Figure 3.9. Due to the low probability values in the buffer layer around the target position in Figure 3.8, the corresponding values near the target in the 2D likelihood map are also relatively low after the element-wise product operation used in its construction.

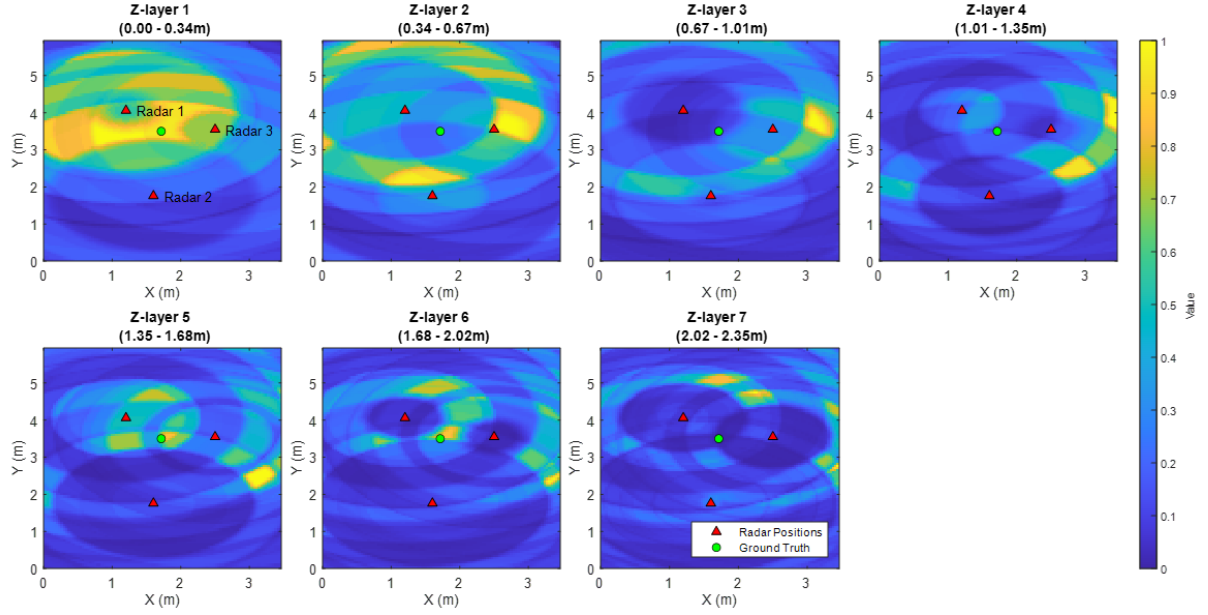


Figure 3.8: An example of 2D slices (in the vertical z -dimension) from the target reflection probability grid is shown for the experimental data sample with test ID $VH_T_W_P1_H1$, obtained at slow-time index $t = 390$.

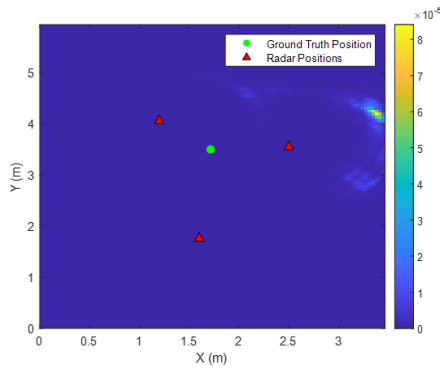


Figure 3.9: An example of the likelihood map is shown for the experimental data sample with test ID $VH_T_W_P1_H1$, obtained at slow-time index $t = 390$.

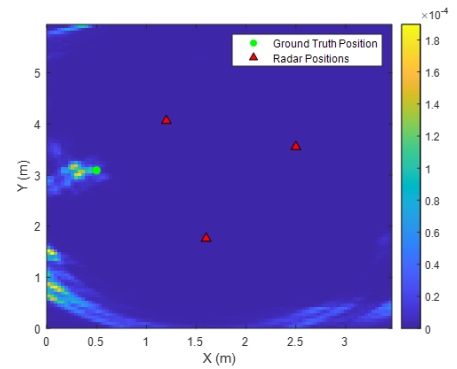


Figure 3.10: An example of the likelihood map is shown for the experimental data sample with test ID $VH_T_W_P1_H1$, obtained at slow-time index $t = 210$.

Interestingly, in the slow-time frame $t = 210$, where the target is located farther from all three radars compared to $t = 390$ (when the target is near the center of the room and closer to all radars), the likelihood map in Figure 3.10 shows a high likelihood value around the target. By examining the target reflection probability grid in Figure 3.11 for $t = 210$, it can be seen that the probability values around the target are consistently high across all layers.

This observation suggests that the likelihood map is generally capable of reflecting the target's position. However, when the target is positioned close to the radars, the likelihood map appears to suffer from certain limitations that require compensation to improve estimation accuracy. The mathematical model behind this phenomenon, along with the proposed solutions, will be discussed in Section 3.6.1 and Section 3.6.2.

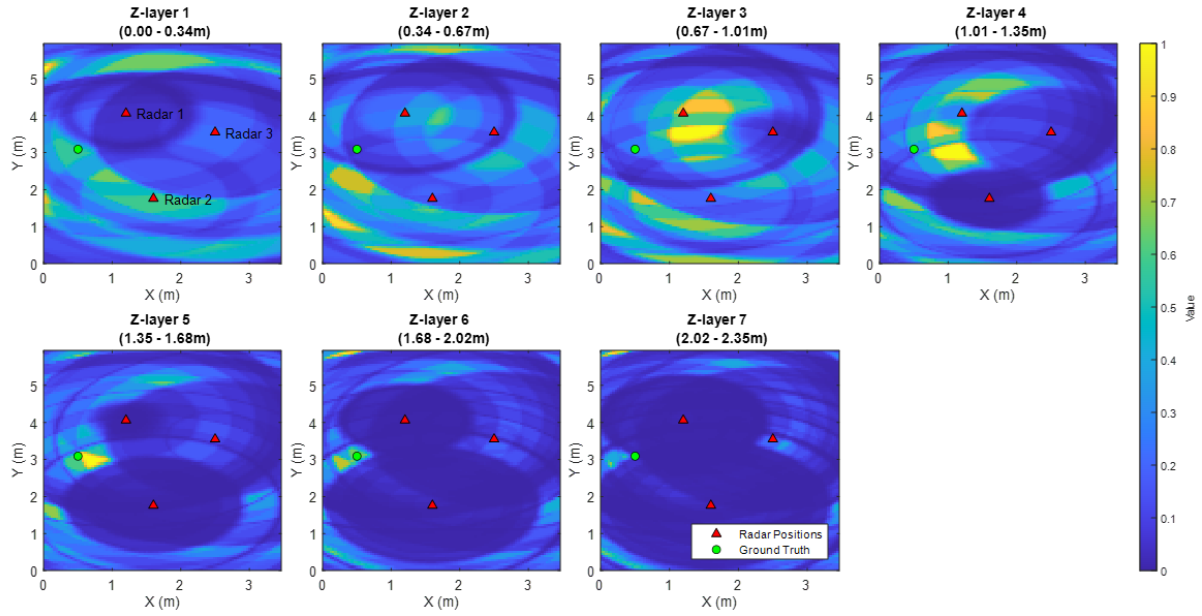


Figure 3.11: An example of 2D slices (in the vertical z -dimension) from the target reflection probability grid is shown for the experimental data sample with test ID $VH_T_W_P1_H1$, obtained at slow-time index $t = 210$.

3.6. 3D Grid Compensation

After the mapping process, 3D grids of CIR $G_{m,t}$ and target reflection probability grid R_t are constructed. However, these grids require further compensation to enhance target signals and suppress artifacts. In this thesis, three compensation procedures are applied:

1. Artifacts Compensation
2. Target Ring Enhancement
3. Position Dilution of Precision (PDOP) Compensation

The relationship among these compensation methods is illustrated in Figure 3.12. The first two procedures, Artifacts Compensation and Target Ring Enhancement, belong to the layer compensation stage in Figure 3.3. The former is applied only to the vertical buffer layer, whereas the latter is applied to all vertical layers, and both are performed on each CIR grid individually. In contrast, the PDOP compensation is applied to the fused CIR grid R_t .

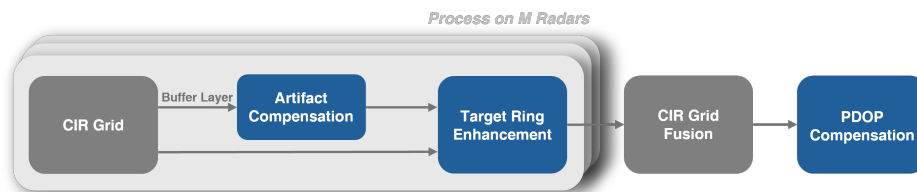


Figure 3.12: The relationship between the three 3D grid compensation methods applied in this thesis.

3.6.1. Artifacts Compensation

As described in Section 3.5, a low probability around the target location can be observed when the target is close to the radar(s). In this section, the geometric model of the artifacts is first introduced, followed by the artifact compensation method proposed to address this problem.

Geometric Model

This phenomenon occurs in the frequent experimental scenario when the human subject's height is below 2.02 m. In such cases, the head is located in the head layer ($n_z = 6$) rather than in the buffer layer ($n_z = 7$), meaning that no real target reflections exist in the buffer layer. Consequently, all high values observed in the buffer layer originate from artifacts projected from the head layer. Figure 3.13 illustrates the geometric model behind this phenomenon. In the figure, the red line represents the human body, with points P_1 , P_2 , and P_3 denoting the head, torso, and feet, respectively. For clarity, only four layers are depicted: the buffer layer, head layer, torso layer, and foot layer. The head, torso, and foot layers are named according to their alignment with the horizontal planes corresponding to the respective body parts, while the buffer layer is highlighted in orange and the others in blue. Point R denotes the UWB radar mounted on the ceiling. Due to the mapping method, all voxel points at the same distance from the radar share the same CIR value, forming a hemispherical surface. Even though no real target exists in the buffer layer, the high value reflection from the head at P_1 produces an artifact point G_1 such that $\|P_1 R\| = \|G_1 R\|$. Considering the human body's projection onto the buffer layer at point H , this artifact causes the target range ring to be larger than it should be.

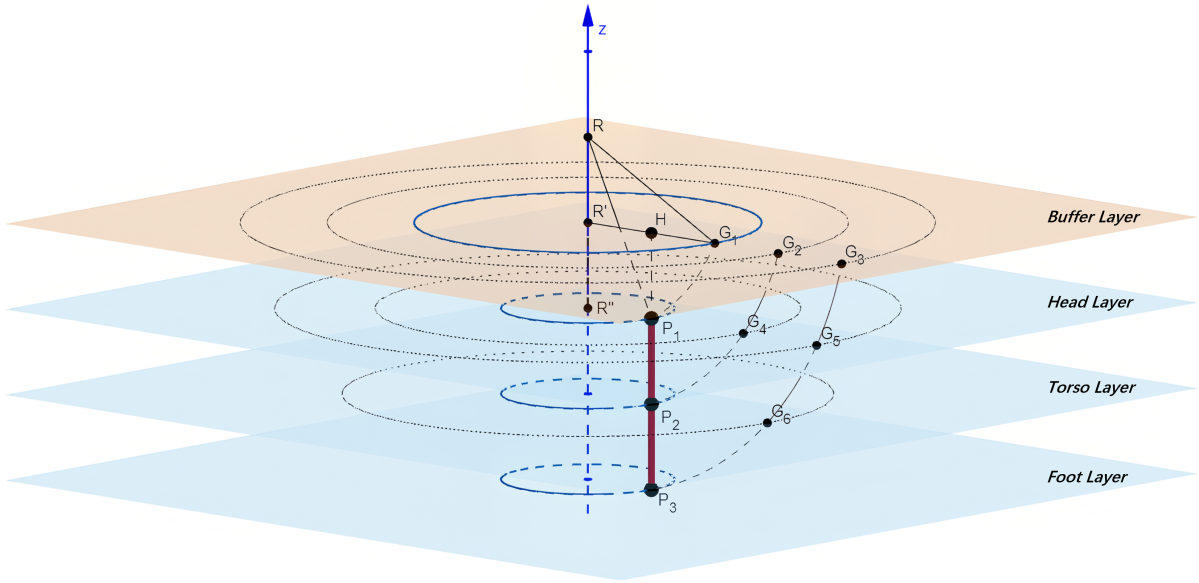


Figure 3.13: Illustration of the artifacts in the CIR grid across different vertical layers.

Furthermore, because the head is the closest body part to the radar, it corresponds to the first peak in the CIR. The range bins before this first peak have very low values, as no target reflections are present. Consequently, in the buffer layer, the region within the solid blue circle of radius $\|G_1 R'\|$ centered at R' exhibits low values, where R' is the projection of the radar onto the buffer layer. As a result, the human projection point H has a low target reflection probability in the buffer layer. This phenomenon can be observed in the $n_z = 6$ and $n_z = 7$ layers of Figure 3.7, where $n_z = 6$ corresponds specifically to the head layer and $n_z = 7$ to the buffer layer. In the buffer layer, the ground truth point exhibits a low value, and the target ring appears enlarged due to artifacts.

During the 2D likelihood map construction, the joint distribution is computed by performing an element-wise

product across all vertical z -layers. This process requires that the target position maintain high probability values in all layers; otherwise, a low probability in any layer will reduce the final likelihood at the target location. In this case, the low probability at point H in the buffer layer results in a reduced likelihood in the 2D likelihood map at the true target position, ultimately leading to incorrect localization as shown in Figure 3.9.

Proposed Solution via Radial Shrink Method

To address this problem, the radial shrink method is applied as the core strategy for artifact compensation. The main idea is to quantify the extent to which artifacts enlarge the target ring, measured by the distance $\|G_1 H\|$, and to shift the value from the artifact point to its corresponding true target projection point on the buffer layer of the CIR grid $G_{m,t}$. For each (n_x, n_y) point on the head layer, a corresponding artifact point is generated on the buffer layer. The distance between the artifact point and its associated target projection point is computed and used to reassign the value accordingly, thereby performing the radial shrink. Figure 3.14 illustrates the procedure of the artifact compensation process: the artifact distance is first calculated along the radial line, and then the artifact distance is used to shift the values along that radial line on the buffer layer.

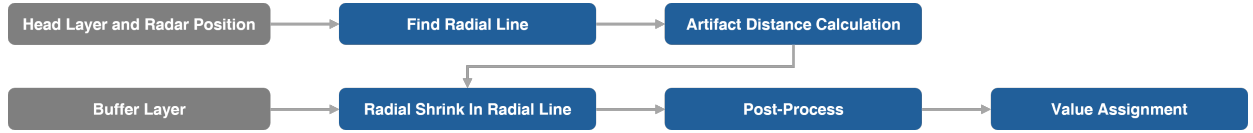


Figure 3.14: Pipeline of the Artifact Compensation Process.

To simplify the computation of the distance between an artifact point and its associated target projection point (denoted as the ‘artifact distance’), the calculation can be restricted to points along the radial line of the grid. This is because the hemispherical target range in each horizontal layer forms a circular ring centered at the projected radar position, with equal values along the ring. This implies that the values on the longest radial line within the layer can represent all points in the grid. Based on this idea, it is unnecessary to compute the artifact distance for every voxel in the layer, thereby reducing the total number of computations from $N_x \cdot N_y$ to a smaller amount.

Figure 3.15 illustrates the radial-line concept, focusing on two key layers: the buffer layer and the head layer.

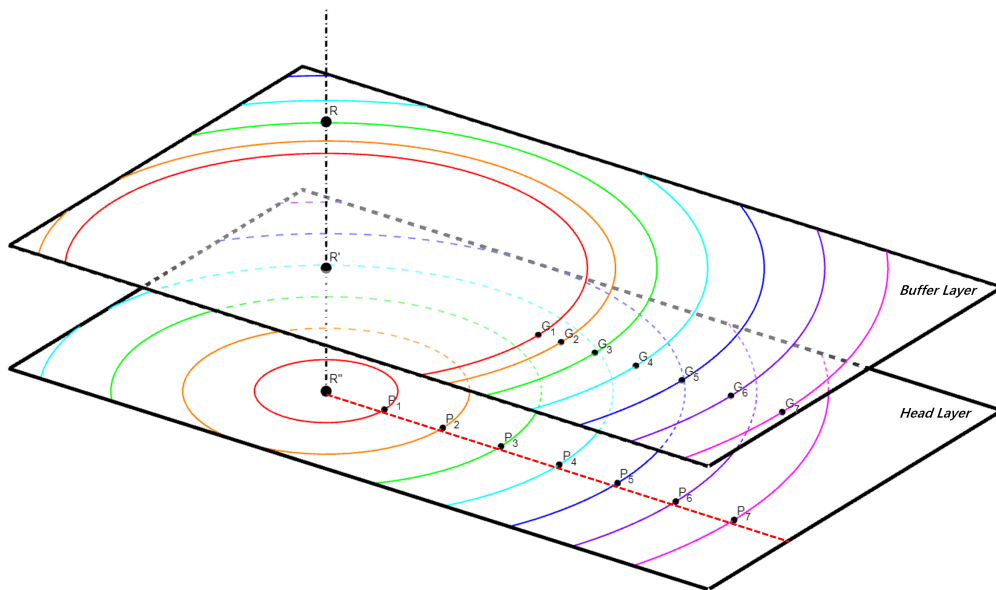


Figure 3.15: Illustration of the radial shrink strategy for artifact compensation.

In this figure, the red dashed line represents the radial line along which calculations are performed. Point R denotes the radar position, while R' and R'' represent its projections onto the buffer and head layers, respectively. Points P_1 to P_7 are grid points along the radial line, and their corresponding artifact points are G_1 to G_7 . Different colors in the figure indicate regions of equal values forming circular rings in both vertical layers. By performing computations only along the radial line, the values for the entire grid layer can be covered, except for a few corner regions. These corner points are less likely to contain a human target, and their artifact distances are generally small (as discussed later), so the values in these regions are left unchanged.

According to the radial line shown as a red dashed line in Figure 3.15, its position is calculated by comparing how far the radar position extends toward the room boundaries along the X and Y axes, then choosing the axis with the longest available length. A straight line of grid cells (i.e., the radial line) is then traced along this chosen axis, passing through the radar position and extending all the way to the farthest boundary of the room. All artifact distances are measured along this line, ensuring the maximum possible artifact distance is captured.

Once the radial line is determined, the artifact distance can be calculated using Equation 3.16, where $\mathbf{d}_{\text{artifact}}(i)$ denotes the artifact distance at index i along the radial line, which geometrically corresponds to the distance $\|G_i H_i\|$ in Figure 3.16. The variables h_t , h_g , and h_r represent the height of the target (i.e., the head layer), the height of the artifact (i.e., the buffer layer), and the height of the radar, respectively.

In this formulation, the furthest point on the radial line (with respect to the radar position) is designated as the origin, denoted as O' and O'' in Figure 3.16, corresponding to the buffer and head layers, respectively. The variable $\mathbf{l}_t(i)$ represents the horizontal distance from the target point P_i on the head layer to the origin O'' , i.e., $\|P_i O''\|$. Similarly, l_r denotes the horizontal distance from the radar projection point R'' to the origin O'' , i.e., $\|R'' O''\|$. Using this setup, the position of the artifact point G_i corresponding to P_i can be computed as $(\mathbf{l}_t(i) - \mathbf{d}_{\text{artifact}}(i), h_g)$. This formulation allows the calculation of both the artifact position and its corresponding artifact distance in physical size for every point P_i along the radial line.

$$\mathbf{d}_{\text{artifact}}(i) = \sqrt{(\mathbf{l}_t(i) - l_r)^2 + (h_t - h_r)^2 - (h_g - h_r)^2} - l_r + \mathbf{l}_t(i) \quad (3.16)$$

Next, the artifact distance $\mathbf{d}_{\text{artifact}}(i)$ and artifact point G_i calculated in physical units are mapped to the grid scale. This process yields a voxel-based artifact distance denoted as $\tilde{\mathbf{d}}_{\text{artifact}}(i)$. This represents the voxel-based distance between the artifact point and its associated target projection point. By converting the artifact distance from physical units to voxel indices, the following shifting operation becomes easy to compute. Besides, it is observed that the artifact distance function is an increasing function: the closer the target point is to the radar, the greater the corresponding artifact distance becomes. Conversely, when the target is farther from the radar, the artifact distance decreases, indicating a reduced impact from artifacts. This also implies that certain corner points not covered by the selected radial line are unlikely to introduce significant errors into the system.

Based on the artifact distance $\tilde{\mathbf{d}}_{\text{artifact}}(i)$ expressed in voxel units, the values in the buffer layer are shifted closer to the radar along the radial line, as defined in Equation 3.17. Here, $\mathbf{s}$ denotes the values from the buffer layer of the CIR grid $\mathbf{G}_{m,t}$ located along the radial line corresponding to artifact positions. All such values on the buffer layer are shifted accordingly. After the shifting operation, a smoothing process is applied to ensure continuity along the radial line. Specifically, a moving mean filter and a maximum filter (implemented using the `ordfilt2` function in MATLAB) are used to smooth the shifted values and fill in unshifted gaps, resulting in a more continuous profile along the radial line. Finally, the shifted radial line on the buffer layer is utilized to generate all values within the entire buffer-layer grid. In detail, each voxel in the buffer layer is assigned the value of its corresponding radial-line point based on having the same voxel-based distance from the radar.

$$\mathbf{s}(i + \tilde{\mathbf{d}}_{\text{artifact}}(i)) \leftarrow \mathbf{s}(i) \quad (3.17)$$

Based on these procedures, the artifacts in the buffer layer of the CIR grid $G_{m,t}$ can be effectively compensated. Figure 3.17 illustrates the comparison of results before and after artifact compensation in $G_{m,t}$, R_t , and L_t . The example is based on the experimental test sample with ID *VH_T_W_P1_H1*, using radar $m = 1$ and slow-time index $t = 390$. From the comparison of $G_{m,t}$ and R_t at the buffer layer ($n_z = 7$), it can be observed that the ring in $G_{m,t}$ is radially shrunk, and the values near the target become higher, resulting in higher values in L_t around the actual target location. This artifact compensation procedure ensures that the joint probability in the 2D likelihood map yields a more accurate result.

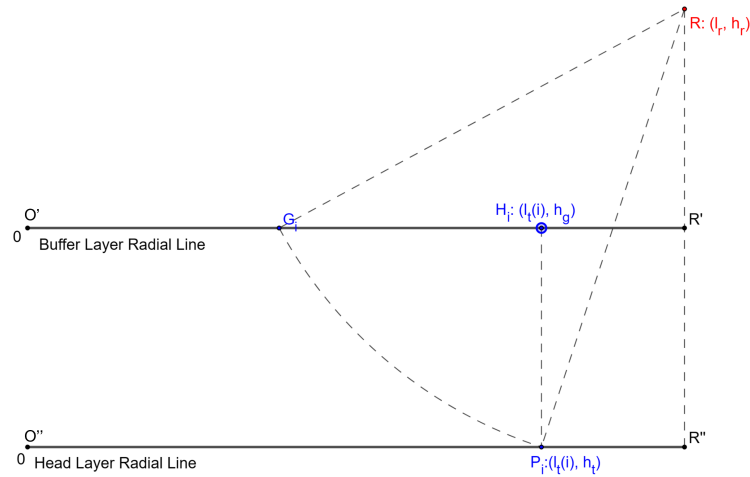


Figure 3.16: Illustration of artifact distance calculation described in this section.

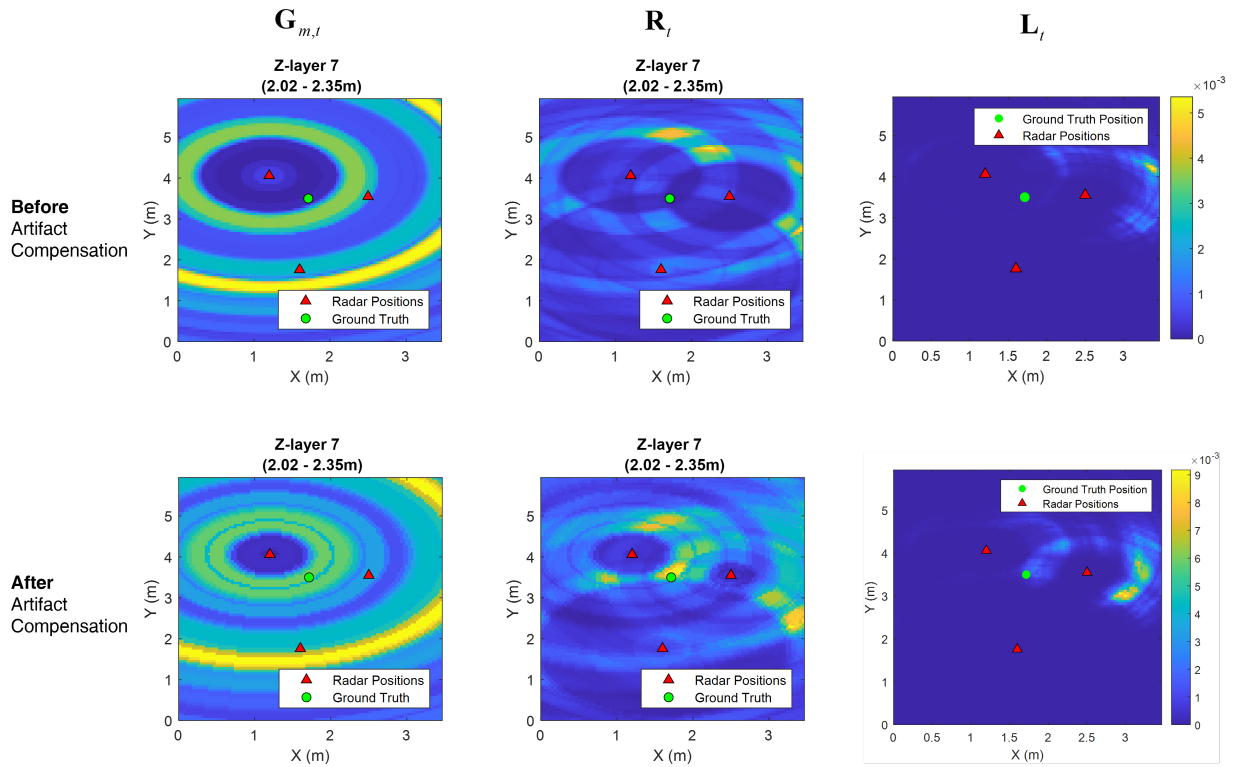


Figure 3.17: Comparison of the results before and after artifact compensation for 3D CIR grid $G_{m,t}$, target reflection probability grid R_t and 2D likelihood map L_t .

3.6.2. Target Ring Enhancement

Although artifact compensation in the buffer layer improves the joint distribution results near the radar, some false alarm points can still be observed in the 2D likelihood map L_t after artifact compensation, as shown in Figure 3.17. In this example, although the ground truth target is located around (1.7 m, 3.5 m), and the artifact compensation result indeed shows increased likelihood values around that area, while some other high likelihood regions also appear around (3 m, 3 m). These may lead to false alarms in the likelihood map, ultimately causing errors in the following tracking process. Therefore, it is necessary to suppress such false alarms while further enhancing the true target region in the likelihood map. Similar to the previous structure, the geometric model causing this issue will first be described, followed by the introduction of the proposed method to address it.

Geometric Model

The geometric reason behind this issue is that strong reflections from the foot layer or torso layer may generate upward-propagated artifacts in higher layers, resulting in false alarm points. As illustrated in Figure 3.13, a real target located in the foot layer (e.g., point P_3) can produce artifact reflections at points G_3 , G_5 , and G_6 in the buffer layer, head layer, and torso layer, respectively. When the signal strength of the foot target in the grid is very high, and even exceeds that of other body parts, the artifacts it generates in the higher layers may appear stronger than the true target.

As shown in Figure 3.7, a high value target ring is observed in the foot layer ($n_z = 1$), which generates artifacts in all higher layers. This may be because in some position, the RCS of the lower limb is larger than the head. These artifacts result in additional high value rings appearing in those layers. For example, in the head layer ($n_z = 6$), although a ring structure exists around the true target location with moderate values, another ring originating from the foot layer artifact is also present and exhibits even higher values. This phenomenon does not occur in just a single radar but appears across all radars, as illustrated in Figure 3.18.

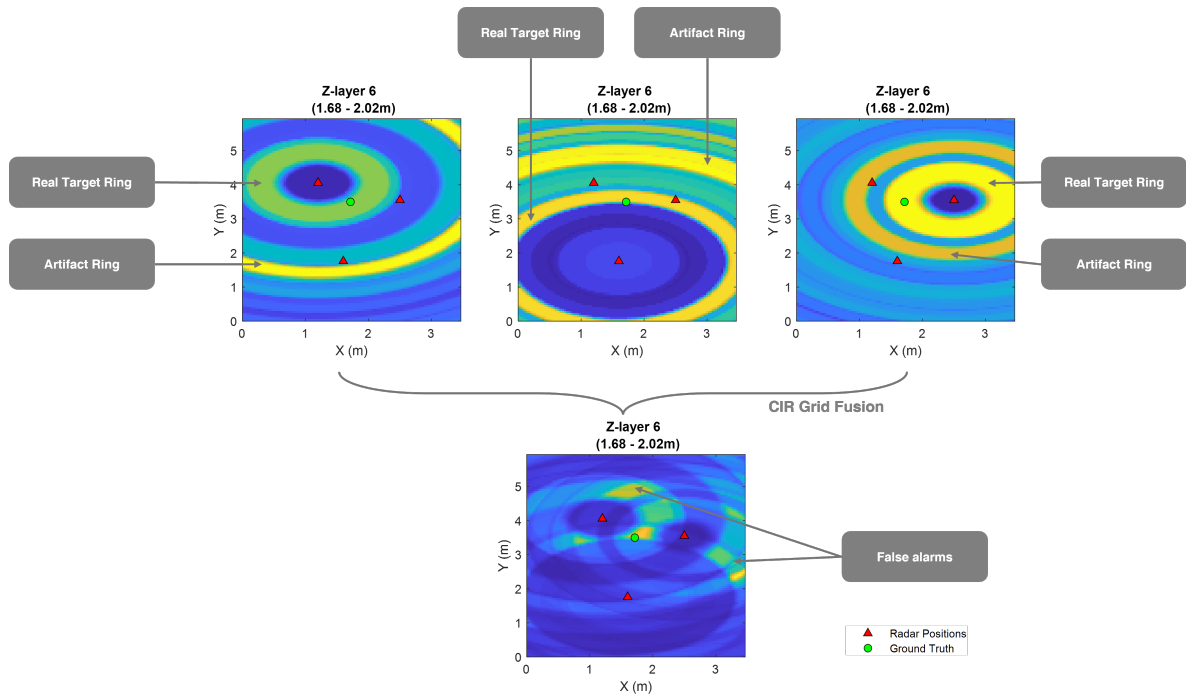


Figure 3.18: Illustration of artifact rings across all radars and the resulting false alarms at slow-time $t = 390$ for experimental test sample with ID $VH_TW_P1_H1$.

In Figure 3.18, artifact rings are present in the CIR grid $G_{m,t}$ of all radars and lead to false alarm regions in the

corresponding fused $\mathbf{R}_t$ after the CIR grid fusion process. This artifact-induced ring can interfere with the identification of the actual target in the higher layer causing false alarms in the likelihood map and affect the accuracy of the following tracking process.

Proposed Adaptive Peak Enhancement and Suppression Method

To reduce the impact of false alarm regions, the general idea is to enhance the real target ring and suppress the artifact rings in each layer, for each radar, and at every slow-time index. This is achieved through a proposed adaptive peak enhancement and suppression method, which serves as a key strategy in this compensation progress. Figure 3.19 illustrates the overall pipeline of the proposed method. Following the previous analysis, the values in each layer of the CIR grid $\mathbf{G}_{m,t}$ can be represented by the corresponding values along radial lines. The peaks along these radial lines correspond to either the target ring or artifact rings. By computing enhancement and suppression ratios for these identified peaks, a ratio matrix is constructed for each layer. This matrix is then applied to the original CIR grid $\mathbf{G}_{m,t}$ to adjust the values, thus enhancing the true target ring and suppressing artifact rings, finally leading to improved likelihood estimation performance.

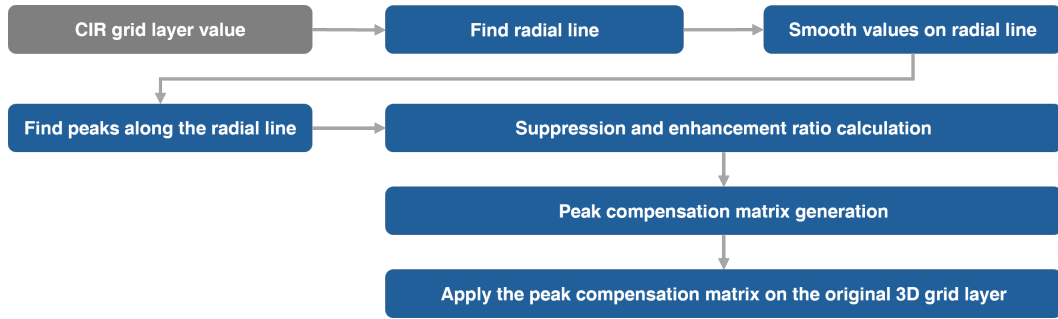


Figure 3.19: Pipeline of the Target Ring enhancement process.

In each z -layer of $\mathbf{G}_{m,t}$, the radial line is first determined using the method described in Section 3.6.1. By analyzing the values along this radial line, the computational complexity is reduced, as it is unnecessary to process every point within the layer. In this representation, the target ring and artifact ring, both with high amplitudes in the CIR grid layer, appear as peaks along the radial line. To detect these peaks, the `findpeaks` function in MATLAB is applied to the values along the radial line, preceded by a moving-mean filter as a preprocessing step to suppress small fluctuation. This function outputs the locations, widths, and prominence of the detected peaks [50]. To simplify the procedure, only the two highest peaks are considered as potential target and artifact peaks, since only high-value peaks can significantly affect the joint distribution result in the CIR grid fusion and the construction of the 2D likelihood map. Moreover, in most cases, the signal strength of the target reflection is not particularly low. To ensure the robust result for peak detection, specific constraints are provided as input arguments to `findpeaks` to ensure that peaks which are too close together or too narrow (caused by abnormal values) are excluded. By balancing the trade-off between noise suppression and true peak preservation, the parameter settings for peak detection are listed in Table 3.1.

Based on these settings, peaks along the radial line in each layer of $\mathbf{G}_{m,t}$ can be identified. An example of peak detection for $n_z = 6$ in the 3D grid $\mathbf{G}_{m,t}$, with radar $m = 1$ and slow-time index $t = 390$, is shown in Figure 3.20. As defined in Section 3.6.1, the furthest point on the radial line is labeled as index 0, and the radar position corresponds to the largest index value on the radial line. From this figure, two main peaks can be observed: a smaller peak located closer to the radar, and a larger peak located farther away. Since artifact rings originate from lower layers, their projected radius in the higher layers is larger. Therefore, the target ring in a specific layer's sectional circle will have a smaller radius than the corresponding artifact ring. Thus, when two peaks are present along the radial

line, the peak closer to the radar can be regarded as the target peak, while the one farther away is regarded as the artifact peak. This feature enables the following enhancement of target peaks and suppression of artifact peaks.

Table 3.1: Input arguments in the findpeaks function used within the processing pipeline for Target Ring enhancement.

Parameter	Value	Description
NPeaks	2	Maximum number of peaks to detect
MinPeakWidth	1	Minimum peak width
MinPeakDistance	1	Minimum distance between peaks
MinPeakProminence	0.1	Minimum peak prominence

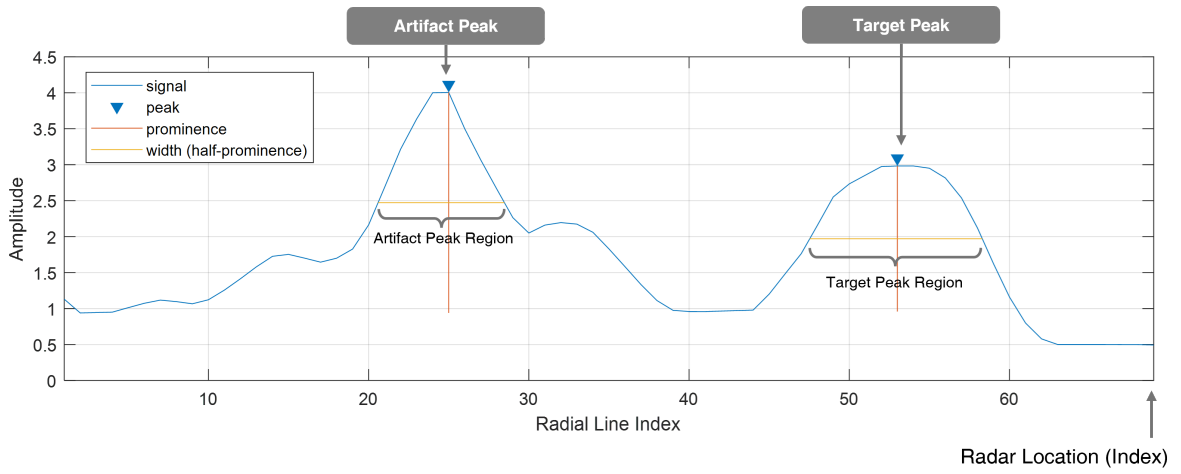


Figure 3.20: Example of peaks along the radial line at $n_z = 6$ in the CIR grid $G_{m,t}$, where radar $m = 1$ and the slow-time index is $t = 390$, from experimental test sample with ID *VH_TW_P1_H1*.

To enhance the target peak and suppress the artifact peak, the proposed approach applies separate multiplicative factors to these two peaks within their respective peak regions, as defined by the peak widths obtained from the findpeaks function. The peak regions are illustrated in Figure 3.20. The adaptive enhancement and suppression ratios are computed using Equation 3.18, where f_1 and f_2 denote the enhancement ratio for the target peak and the suppression ratio for the artifact peak, respectively. p_1 and p_2 represent the amplitudes of the target and artifact peaks, l_1 and l_2 are their corresponding peak location indices along the radial line, and l_r is the index of the radar location on the radial line.

$$\begin{aligned}
 f_1 &= \beta^{1+\lambda(1-\frac{1}{\gamma})}, \\
 f_2 &= \beta^{-\lambda[1+\lambda(1-\frac{1}{\gamma})]}, \\
 \beta &= \frac{p_2}{p_1}, \quad \gamma = \frac{l_r - l_2}{l_r - l_1}, \quad \lambda > 1.
 \end{aligned} \tag{3.18}$$

This formulation is based on two main considerations:

1. When the artifact peak is significantly larger and located farther from the radar than the target peak, the enhancement ratio f_1 should be increased, and the suppression ratio f_2 should be decreased accordingly.
2. To ensure that the artifact peak remains weaker than the target peak even when their original amplitudes are similar, the suppression strength must exceed the enhancement strength ($f_2 < 1$, $f_1 > 1$). This is achieved

by setting $\lambda > 1$, which guarantees that the artifact peak is reduced below the target peak.

If λ is set too low, the enhancement and suppression effects will be insufficient. Inversely, if λ is too large, the too low suppression ratio may cause missed detections in following processing. Based on these considerations, a balanced value of $\lambda = 2.5$ is selected in this thesis.

To unify the enhancement and suppression operations across all radial lines, an enhancement coefficient vector $\mathbf{e} \in \mathbb{R}^{l_r}$ is formulated according to Equation 3.19. In this formulation, the enhancement ratio f_1 and suppression ratio f_2 are assigned to their respective index intervals. Here, l_r denotes the radar index, which is also the length of the radial line, $[i_{t,1}, i_{t,2}]$ represents the index range corresponding to the half peak width of the target peak, and $[i_{a,1}, i_{a,2}]$ represents the index range corresponding to the half peak width of the artifact peak.

$$\mathbf{e}(i) = \begin{cases} f_1, & i \in [i_{t,1}, i_{t,2}], \\ f_2, & i \in [i_{a,1}, i_{a,2}], \\ 1, & \text{otherwise.} \end{cases} \quad (3.19)$$

Then, each value in e is mapped to an enhancement coefficient matrix $\mathbf{E} \in \mathbb{R}^{N_x \times N_y}$ by assigning the same value to all grid points that have the same distance from the radar position as the corresponding index in e . The compensated grid layer is then obtained using Equation 3.20, where $\mathbf{E}_{m,t}^{n_z}$ denotes the enhancement coefficient matrix for radar m , slow-time index t , and z-layer n_z , and $\mathbf{G}_{m,t}^{n_z}$ denotes the CIR grid layer from CIR grid at radar m , slow-time index t and z-layer n_z . The operator $\circ$ represents element-wise product.

$$\mathbf{G}_{m,t}^{n_z} = \mathbf{E}_{m,t}^{n_z} \circ \mathbf{G}_{m,t}^{n_z}, \quad (3.20)$$

However, not all z-layers in the CIR grids are suitable for this compensation. The enhancement and suppression process is applied only when the following conditions are satisfied:

1. **Artifact peak should not be lower than target peak:** If the target peak is already dominant, neither enhancement of the target nor suppression of the artifact is necessary.
2. **Target peak amplitude should be sufficiently large:** If the target peak amplitude is too low, it may indicate that there is no target reflection in this layer, or the target is too far from the radar (such that the artifact lies outside the room), or even that the peak originates from noise. In such cases, compensation is unnecessary.
3. **Target peak should not be located too close to the radar:** When the target is too close, strong reflections from the head are expected. In such cases, strong artifact signals are more likely to represent noise, which does not require compensation.

Since only the top two peaks are selected as target peak and artifact peak, there is a possibility that both peaks correspond to artifact when the true target signal is very weak. This may lead to false target ring enhancements. However, as the system processes N_z layers and incorporates data from M radars, the probability of all detections being incorrect simultaneously is low. Furthermore, even if such errors occur, the following tracking procedure combines the prior distribution with the estimated 2D likelihood map, which helps to correct these mistakes during the final trajectory estimation.

Based on the previous steps and considerations, the target ring can be enhanced while the artifact ring is suppressed. An example of results after applying the target ring enhancement procedure is shown in Figure 3.21. This figure compares the results before and after target ring enhancement (with prior artifact compensation applied) for the $n_z = 6$ layer of the CIR grid $\mathbf{G}_{m,t}$, the $n_z = 6$ layer of the target reflection probability grid $\mathbf{R}_t$, and the 2D likelihood map $\mathbf{L}_t$ at slow-time frame $t = 390$. The $\mathbf{G}_{m,t}$ shown in this figure corresponds to radar $m = 1$, and the data is taken from the experimental test sample with ID *VH_T_W_P1_H1*.

From the comparison, it can be observed that in the CIR grid $G_{m,t}$, the target ring is enhanced while the artifact ring is almost completely removed. In R_t , the artifact region around (3 m, 3 m) is suppressed compared to the result before target ring enhancement. In the 2D likelihood map L_t , the same artifact region is suppressed and the target area is enhanced, leading to a more accurate likelihood estimation.

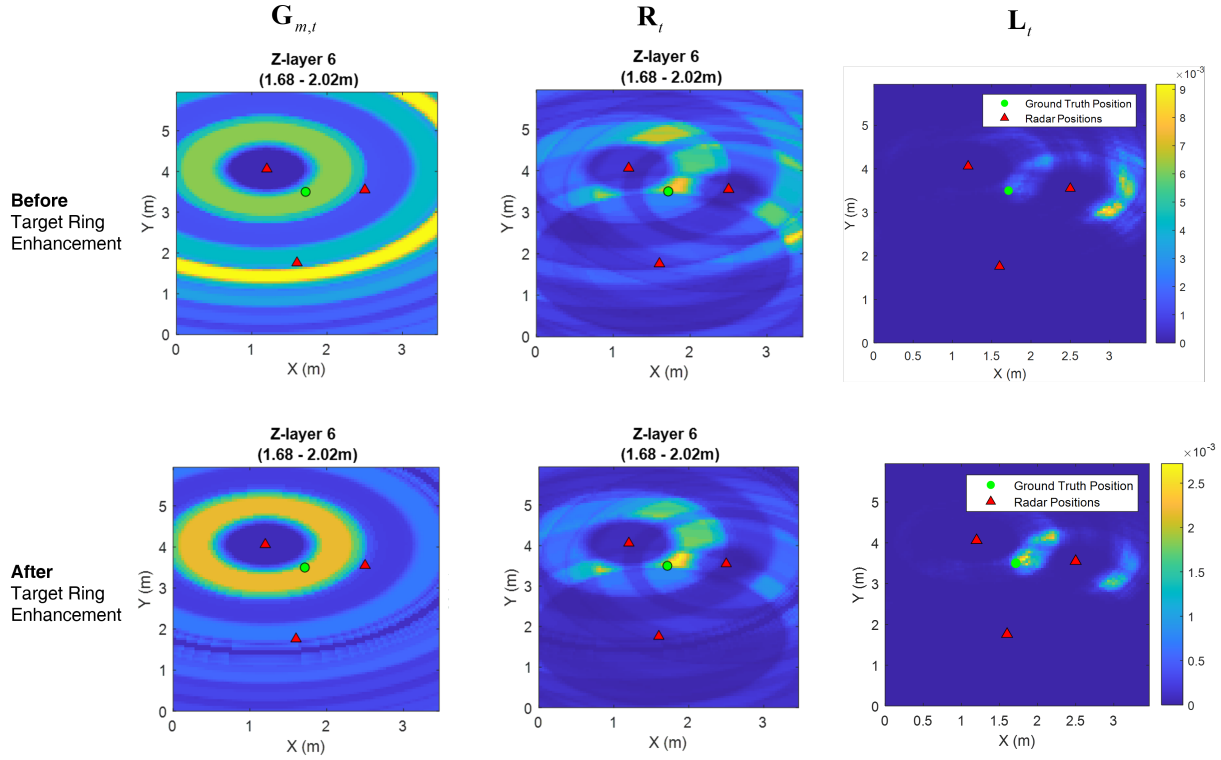


Figure 3.21: Comparison of the result before and after Target Ring enhancement for 3D CIR grid $G_{m,t}$, target reflection probability grid R_t and 2D likelihood map L_t .

3.6.3. PDOP Compensation

Based on the previous compensation steps, most artifacts in the 3D CIR grid have been suppressed. However, not every voxel in the grid contributes equally to the construction of the 2D likelihood map. For example, some voxels are located too far from the radar or have large opening angles relative to the radar position. Such voxels may reduce the accuracy of localization and consequently degrade tracking performance.

To quantify which voxels may reduce localization accuracy, the Position Dilution of Precision (PDOP) metric is introduced in this thesis. PDOP is a variant of the Geometric Dilution of Precision (GDOP) metric that excludes the time-delay component. GDOP, which is widely used in the Global Positioning System (GPS), quantifies how the geometry of sensor (or satellite) placement affects the accuracy of position estimation at a given point in space. When the time-delay component is ignored, GDOP becomes equivalent to PDOP. A lower PDOP value at a position indicates higher localization accuracy [51].

Figure 3.22 illustrates how the geometry of satellite placement influences localization accuracy. Because the estimated ranges in the system are not perfectly accurate and are subject to measurement variance, the feasible region defined by these ranges is not the simple intersection of several circles, but rather the intersection of several circular rings. If the satellites are positioned close together, as in Figure 3.22(a), the intersection region (red area) becomes larger, making it more difficult to pinpoint the true position and thus leading to poor localization accuracy. Conversely, if the satellites are positioned more evenly, as in Figure 3.22(b), the intersection area of the rings becomes smaller, resulting in better localization accuracy. In other words, if the volume of the polyhedron formed

by connecting all satellites to the user is larger, the achievable positioning accuracy will be higher [51].

The traditional application of PDOP is as a metric for selecting an optimal satellite geometry [51]. PDOP considers only the geometric effect, which is the spatial relationship between a point in space and the sensor system, on the dilution of precision. Because of this property, PDOP (or GDOP, which is similar to PDOP but also accounts for time delay) can also be applied in radar systems, for example, to determine the optimal placement of receivers [52].

In this thesis, however, the use of PDOP differs from the traditional approach. Instead of guiding the optimal placement of radar sensors, PDOP is used to quantify which points in a room, given a fixed radar configuration, have better localization accuracy with lower PDOP. The PDOP value is then used as a kind of confidence weight to indicate which voxels in the grid have higher localization accuracy and should therefore be assigned greater trust in the estimation process.

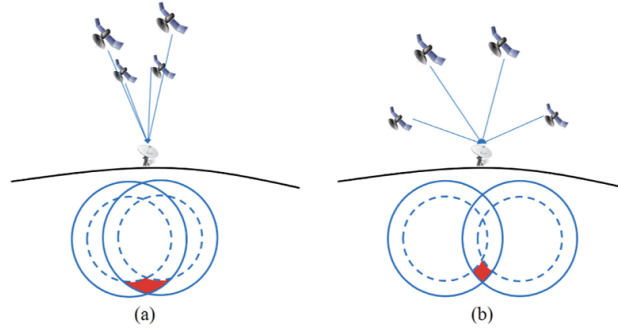


Figure 3.22: Illustration of the geometry of satellite placement affecting PDOP (from [53]): (a) Satellites placed close together result in poor localization accuracy and a higher PDOP value; (b) Satellites placed evenly result in better localization accuracy and a lower PDOP value. The solid circles represent the true distances from the satellites to the user and intersect at the receiver's position. The dotted circles, which are parallel to the solid circles, represent the erroneous measurements obtained by the receiver. The red region indicates the uncertainty area.

The PDOP value for each voxel index (n_x, n_y, n_z) in the discretized 3D grid is computed according to Equation 3.21, where $\text{PDOP}(n_x, n_y, n_z)$ denotes the PDOP value in the PDOP tensor for voxel (n_x, n_y, n_z) [51]. The inverse of the geometry matrix $\mathbf{A}_{n_x, n_y, n_z}$ quantifies the amplification of measurement noise caused purely by the radar-target geometry at that voxel. A smaller PDOP value indicates that the voxel lies in a geometrically favorable region, where ranging errors propagate less into the position estimation. In contrast, a larger PDOP value suggests that the geometry between the voxel and radar positions results in greater dilution of precision, increasing the impact of measurement noise. The matrix $\mathbf{A}_{n_x, n_y, n_z}$ is defined in Equation 3.22, where $\mathbf{H}_{n_x, n_y, n_z}$ is a Jacobian matrix of voxel index (n_x, n_y, n_z) . The matrix $\mathbf{H}_{n_x, n_y, n_z}$ is given in Equation 3.23, where $\mathbf{p}_{r_m}$ denotes the position of radar m (defined in Equation 3.24), and $\mathbf{p}_{n_x, n_y, n_z}$ is the physical location of voxel (n_x, n_y, n_z) in the room (also defined in Equation 3.24) with a grid step size of $(\Delta x, \Delta y, \Delta z)$.

$$\text{PDOP}(n_x, n_y, n_z) = \sqrt{\text{trace}(\mathbf{A}_{n_x, n_y, n_z}^{-1})} = \sqrt{\frac{\text{trace}(\text{adj}(\mathbf{A}_{n_x, n_y, n_z}))}{\det(\mathbf{A}_{n_x, n_y, n_z})}} \quad (3.21)$$

$$\mathbf{A}_{n_x, n_y, n_z} = \mathbf{H}_{n_x, n_y, n_z}^\top \mathbf{H}_{n_x, n_y, n_z} \quad (3.22)$$

$$\mathbf{H}_{n_x, n_y, n_z} = \begin{bmatrix} \frac{\mathbf{p}_{r_1} - \mathbf{p}_{n_x, n_y, n_z}}{\|\mathbf{p}_{r_1} - \mathbf{p}_{n_x, n_y, n_z}\|} & \frac{\mathbf{p}_{r_2} - \mathbf{p}_{n_x, n_y, n_z}}{\|\mathbf{p}_{r_2} - \mathbf{p}_{n_x, n_y, n_z}\|} & \dots & \frac{\mathbf{p}_{r_M} - \mathbf{p}_{n_x, n_y, n_z}}{\|\mathbf{p}_{r_M} - \mathbf{p}_{n_x, n_y, n_z}\|} \end{bmatrix}^\top \quad (3.23)$$

$$\mathbf{p}_{n_x, n_y, n_z} = (\Delta x \cdot n_x, \Delta y \cdot n_y, \Delta z \cdot n_z)^\top, \quad \mathbf{p}_{r_m} = (x_{r_m}, y_{r_m}, z_{r_m})^\top \quad (3.24)$$

Based on the calculation, the PDOP tensor $\mathbf{PDOP}$, which has the same grid size as $\mathbf{G}_{m,t}$ or $\mathbf{R}_t$, can be obtained. Each element $\mathbf{PDOP}(n_x, n_y, n_z)$ represents the PDOP value of voxel (n_x, n_y, n_z) in the room grid. The PDOP is a dimensionless quantity that reflects the geometric contribution of radar placement to localization accuracy. Then, the PDOP values are normalized according to Equation 3.25, ensuring numerical stability and making them suitable for use as confidence weights. The normalized PDOP tensor is denoted as $\mathbf{O}$. Since a lower PDOP corresponds to higher positioning accuracy, the inverse of the PDOP is used so that higher accuracy regions receive higher confidence. To further emphasize this weighting effect, the square of the PDOP value is taken before inversion. This choice is motivated by the theoretical relationship that the variance of the estimated position is proportional to the square of the PDOP multiplied by the variance of the measurement error, as shown in Equation 3.26 [54]. From an intuitive perspective, using the squared PDOP in the denominator amplifies the weight difference between high-accuracy and low-accuracy voxels: regions with low PDOP are boosted more significantly, while regions with high PDOP are suppressed more strongly.

$$\mathbf{O}(n_x, n_y, n_z) = \frac{\mathbf{PDOP}(n_x, n_y, n_z)^{-2}}{\max(\mathbf{PDOP}(n_x, n_y, n_z)^{-2})} \quad (3.25)$$

$$\sigma_{\text{pos}}^2 = \mathbf{PDOP}^2 \cdot \sigma_{\text{meas}}^2 \quad (3.26)$$

Figure 3.23 shows all vertical layers of the normalized PDOP tensor $\mathbf{O}$ based on the radar positions of the experimental setup², where the red triangles indicate the positions of the UWB radars. From this figure, it can be observed that after normalization, the maximum normalized PDOP value is 1. When the voxel position is closer to the ceiling-mounted radars and located in the central area of the room (especially within the polygon formed by the radar positions), the normalized PDOP is higher, indicating better positioning accuracy. Conversely, when the voxel is located higher but outside the central area (closer to the room corners), the normalized PDOP value decreases, indicating poorer localization accuracy.

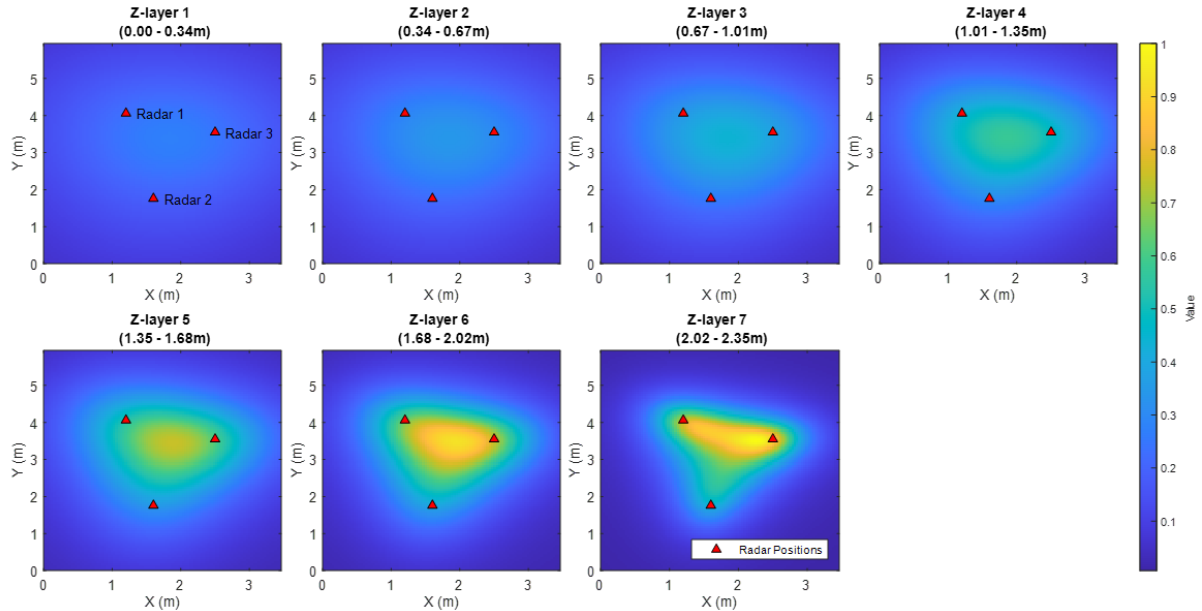


Figure 3.23: Example of normalized PDOP values for horizontal layers of the experiment setup.

Furthermore, if the voxel is inside the central area, a lower height corresponds to a lower normalized PDOP. In

²The experimental setup will be fully described in Chapter 5

contrast, if the voxel is located in the corner area, a lower height corresponds to a higher normalized PDOP. This suggests that when a person is in the central area of the room, the head position will have better localization accuracy, while when the person is near the room corners, the lower limbs may achieve better localization accuracy. In addition, when the person is in the corner area, the head position will also be affected by the opening-angle issue discussed in Section 1.2.3, which further reduces localization accuracy.

To merge PDOP compensation into the 2D likelihood map construction as a measure of voxel confidence, the key challenge is to determine the confidence of position accuracy along the z -axis of the 3D grid $\mathbf{R}_t$, since the 2D likelihood map is computed from the joint distribution of all z -layers. Based on the previous analysis, when a person is outside the central area defined by the polygon formed by the radar positions, the lower limbs generally exhibit better localization accuracy than the head. Therefore, in such regions, the PDOP factor should assign higher confidence to the lower limb voxels. In contrast, when the person is inside the central area, although the head has a better normalized PDOP, the PDOP of the lower limbs is still higher than that of the corner regions within the same area. Moreover, the signal strength from both the lower limbs and the head in the central area is strong and is not significantly affected by the opening-angle problem observed for the head outside the central area. As a result, voxel columns in the central area should have similar confidence values for both the head and the lower limbs.

Based on this reasoning, two PDOP confidence strategies are defined, depending on whether the person is inside or outside the central area. The transition between these two strategies must be smooth to ensure continuous results at the boundary of the center area. The boundary and transition are defined by Equation 3.27, which specifies the PDOP weight factor $\mathbf{W}(n_x, n_y)$. There are two key components in Equation 3.27: (1) defining the central area, and (2) ensuring a smooth transition between the two strategies. The central area is defined using the mean normalized PDOP of the entire tensor, denoted by μ in Equation 3.27. If the mean normalized PDOP along the z -axis for a voxel, $\frac{1}{N_z} \sum_{n_z=1}^{N_z} \mathbf{O}(n_x, n_y, n_z)$, is larger than μ , that voxel is classified as inside the central area; otherwise, it is classified as outside. To achieve a smooth transition, a sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ is employed together with a smoothness factor ψ . A larger ψ results in a smoother transition, whereas a smaller ψ produces a sharper boundary. Based on this method, $\mathbf{W}(n_x, n_y)$ approaches zero when the voxel is further outside the central area, and approaches one when the voxel is further inside.

$$\mathbf{W}(n_x, n_y) = \sigma \left(\frac{\frac{1}{N_z} \sum_{n_z=1}^{N_z} \mathbf{O}(n_x, n_y, n_z) - \mu}{\psi} \right), \quad (3.27)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad \mu = \frac{1}{N_x N_y N_z} \sum_{n_x=1}^{N_x} \sum_{n_y=1}^{N_y} \sum_{n_z=1}^{N_z} \mathbf{O}(n_x, n_y, n_z)$$

Finally, this weight is applied in Equation 3.28 to normalize $\mathbf{O}$ along z axis to get PDOP compensation factor tensor $\mathbf{P}(n_x, n_y, n_z)$ representing PDOP confidence in the space. This ensures that in areas outside the central region, the PDOP compensation factor is normalized according to geometry, while in areas inside the central region, the PDOP compensation factor remains close to one. After the PDOP compensation factor is formulated, the PDOP compensation factor tensor $\mathbf{P}(n_x, n_y, n_z)$ is applied in Equation 3.14 to assign a confidence value to each voxel, thus improving the estimation accuracy of the likelihood map.

$$\mathbf{P}(n_x, n_y, n_z) = \frac{\mathbf{O}(n_x, n_y, n_z)}{(1 - \mathbf{W}(n_x, n_y)) \cdot \max(\mathbf{O}(n_x, n_y, n_z)) + \mathbf{W}(n_x, n_y) \cdot \mathbf{O}(n_x, n_y, n_z)} \quad (3.28)$$

By setting the smooth factor $\psi = 0.1$ in Equation 3.27, an example of the PDOP compensation factor tensor for the experimental radar setup is shown in Figure 3.24. From this figure, it can be observed that the PDOP compensation

factor formulation follows the proposed two strategies: in the central area, the confidence factor is close to one, while in the outside-central area, it is normalized along the z -axis with a smooth transition.

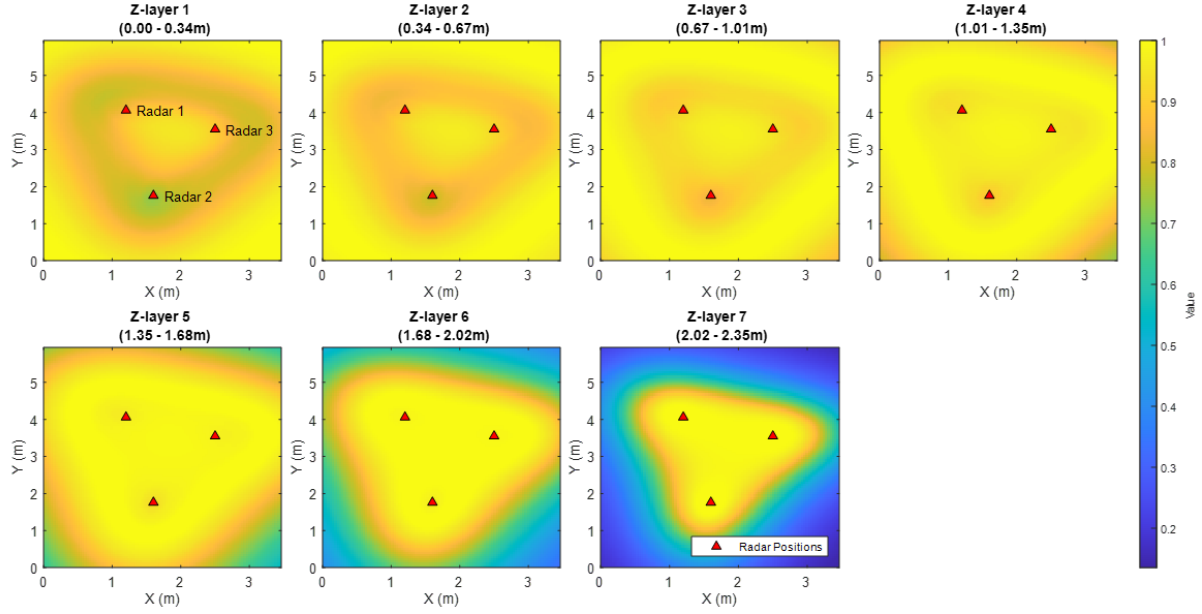


Figure 3.24: Example of the PDOP compensation factor tensor for horizontal layers of the experimental setup.

After the proposed PDOP compensation technique is applied to the target reflection probability grid R_t , the compensated 2D likelihood map L_t can then be obtained. Figure 3.25 presents an example of the PDOP compensation result, following the prior layer compensation step. The 2D likelihood map L_t in this figure corresponds to the slow-time index $t = 390$, and the data is taken from the test experimental data sample with ID *VH_TW_P1_H1*. From this example, it can be observed that the ground truth target area exhibits higher likelihood values, while likelihood values associated with artifacts outside the central area, such as around (3 m, 3 m) are degraded.

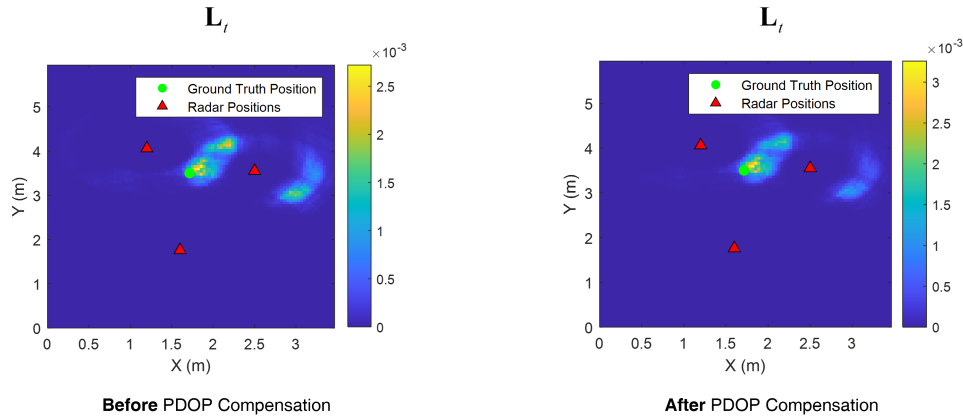


Figure 3.25: Comparison of the result before and after PDOP compensation of 2D likelihood map.

3.7. Particle Filter Tracking

After the 3D grid compensation, the 2D likelihood map L_t contains values proportional to the probability of target occupancy at each grid location. The target position can be estimated by computing $\arg \max L_t$, but residual artifacts still persist in some regions (e.g., around (3 m, 3 m) in Figure 3.25), which may lead to incorrect position estimates.

To robustly exploit temporal continuity, a particle filter (PF) is applied, using the normalized 2D likelihood map as the measurement to estimate the target location, similar to the approach in [38]. The PF employs the likelihood map as the measurement model and a constant-velocity dynamic model as the state model. Through prediction, weighting, and resampling, the PF estimates the posterior distribution of the target state, producing smoother and more reliable trajectories than frame-wise maximization.

3.7.1. Particle Filter Overview

A particle filter (PF) is employed to track the target position from the compensated 2D likelihood map $\mathbf{L}_t$. In contrast to generic PF formulations that require an explicit measurement function, the normalized likelihood map $\tilde{\mathbf{L}}_t$ is directly used as the measurement likelihood $p(z_t|\mathbf{x}_t)$, where (x_t, y_t) denote the particle's continuous grid coordinates. A constant velocity state model is adopted, and systematic resampling followed by post-resampling roughening is applied to maintain particle diversity.

At each time step, the algorithm is executed as follows:

1. **Prediction:** N_p particles are propagated according to the constant velocity model, with process noise injected into the velocity components. In this thesis, $N_p = 25000$ is selected by balancing the estimation result and computational load.
2. **Measurement update:** Each particle's position is mapped to the nearest grid cell (n_x, n_y) , and $\tilde{\mathbf{L}}_t(n_x, n_y)$ is directly used as the likelihood to update the particle weight.
3. **Resampling and roughening:** Weights are normalized to one, and the effective sample size N_{ess} is computed. When N_{ess} is smaller than the threshold, systematic resampling is performed and small Gaussian perturbations (defined in Equation 3.39 and 3.40) are added to position and velocity to mitigate sample impoverishment.
4. **State estimation:** The weighted mean particle position is taken as the state estimate $\hat{\mathbf{p}}_t$. When the subsequent activity classifier indicates a non-walking frame, the estimate is held constant at $\hat{\mathbf{p}}_{t-1}$.

By directly exploiting the target occupancy probability encoded in $\mathbf{L}_t$ and enforcing temporal smoothness through the motion prior, smoother and reliable trajectories can be produced.

3.7.2. State Model

Because indoor walking in an office environment (without running or sudden movements) can be approximated as motion with nearly constant velocity and small fluctuations, the PF adopts a constant-velocity kinematic model, with process noise applied only to the velocity components.

The state vector $\mathbf{x}_t$ is defined in Equation 3.29, where x_t and y_t are the grid-space positions, and $v_{x,t}$ and $v_{y,t}$ are the velocities along the x - and y -axes, respectively. The system dynamics are given in Equation 3.30, where $\mathbf{F}$ is the state transition matrix defined in Equation 3.31, and Δt denotes the time step. The process noise term $\mathbf{q}_t$ has non-zero entries only in the velocity components, represented by $\nu_{x,t}$ and $\nu_{y,t}$, which are independent zero-mean Gaussian noise terms applied to the x - and y -velocities, respectively. These noise terms have variance $\sigma_v^2 \Delta t$ and model the small random variations in speed. In this thesis, $\sigma_v = 0.5$ is selected by balancing the model's ability to adapt to changes in velocity while maintaining a smooth trajectory.

$$\mathbf{x}_t = [x_t, y_t, v_{x,t}, v_{y,t}]^\top \quad (3.29)$$

$$\mathbf{x}_t = \mathbf{F} \mathbf{x}_{t-1} + \mathbf{q}_t, \quad (3.30)$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.31)$$

$$\mathbf{q}_t = [0, 0, \nu_{x,t}, \nu_{y,t}]^\top, \quad \begin{bmatrix} \nu_{x,t} \\ \nu_{y,t} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \Delta t \mathbf{I}) \quad (3.32)$$

$$\mathbf{Q} = \text{diag}(0, 0, \sigma_v^2 \Delta t, \sigma_v^2 \Delta t) \quad (3.33)$$

3.7.3. Measurement Model

In the particle filter, particle weights are updated according to the measurement likelihood. Since particles are initialized within a predefined region around the initial estimate and subsequently propagated within the bounded search space according to the dynamic model, their spatial distribution naturally overlaps the grid of the likelihood map. Under this setting, the likelihood value at the grid cell containing a particle can be directly interpreted as the particle's measurement likelihood. This formulation removes the need for an explicit sensor measurement function, as the normalized likelihood map inherently encodes the target-occupancy probability over the grid.

At each slow-time index t , the UWB signal processing pipeline produces a 2D likelihood map $\mathbf{L}_t$ from the construction and compensation steps described in earlier sections. To ensure numerical stability, $\mathbf{L}_t$ is normalized for every slow-time index to the range $[0, 1]$, yielding the normalized likelihood map $\tilde{\mathbf{L}}_t$ as defined in Equation 3.34:

$$\tilde{\mathbf{L}}_t = \frac{\mathbf{L}_t - \min(\mathbf{L}_t)}{\max(\mathbf{L}_t) - \min(\mathbf{L}_t) + \varepsilon}, \quad \tilde{\mathbf{L}}_t \in [0, 1] \quad (3.34)$$

For each particle i , its position $(x_t^{(i)}, y_t^{(i)})$ is rounded to the nearest grid indices $(n_x^{(i)}, n_y^{(i)})$. The normalized likelihood value at that grid cell is assigned to the unnormalized particle weight $\tilde{w}_t^{(i)}$ according to Equation 3.35, where ε is a small positive constant to avoid zero weight. Then, the particle weights are normalized across all N_p particles to form a probability mass function based on Equation 3.36. Finally, the estimated position $\hat{\mathbf{p}}_t$ is computed as the weighted mean of the particle positions based on Equation 3.37.

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \cdot (\tilde{\mathbf{L}}_t(n_x^{(i)}, n_y^{(i)}) + \varepsilon) \quad (3.35)$$

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N_p} \tilde{w}_t^{(j)}} \quad (3.36)$$

$$\hat{\mathbf{p}}_t = \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \end{bmatrix} = \sum_{i=1}^{N_p} w_t^{(i)} \begin{bmatrix} x_t^{(i)} \\ y_t^{(i)} \end{bmatrix} \quad (3.37)$$

3.7.4. Resample and Roughening

After the measurement update, particle weights tend to concentrate in a small number of high-likelihood regions. While this concentration can improve instant estimation accuracy, it also leads to the particle degeneracy problem, where most particle weights approach zero and only a few particles contribute to the state estimate [55]. This reduces the long-term stability and adaptability of the filter.

To address this issue, the Effective Sample Size (ESS) is used to quantify the degree of particle degeneracy, denoted as N_{ess} in Equation 3.38 [56]. When N_{ess} falls below a predefined threshold ($N_{\text{ess}} < \zeta N_p$), systematic resampling is performed ($\zeta = 0.75$ is used in this thesis). In this method, particles are drawn according to their cumulative weight distribution, so that high weight particles are replicated and low weight particles are eliminated, producing a new particle set evenly distributed over high-likelihood regions [57].

$$N_{\text{ess}} = \frac{1}{\sum_{i=1}^{N_p} \left(w_t^{(i)}\right)^2} \quad (3.38)$$

However, frequent resampling can cause the sample poverty problem, in which particle diversity decreases over time. To mitigate this issue, a roughening step is applied after each resampling, injecting independent Gaussian noise into both the particle positions and velocities according to Equation 3.39. In this equation, σ_{pr} and σ_{vr} are the standard deviations of the position and velocity perturbations for roughening, respectively, computed from Equation 3.40 [58].

$$\mathbf{x}_t^{(i)} \leftarrow \mathbf{x}_t^{(i)} + \mathcal{N}(0, \sigma_{pr}^2 \mathbf{I}), \quad \mathbf{v}_t^{(i)} \leftarrow \mathbf{v}_t^{(i)} + \mathcal{N}(0, \sigma_{vr}^2 \mathbf{I}), \quad (3.39)$$

$$\begin{aligned} \sigma_{pr} &= K_p E_p N_p^{-\frac{1}{d}} \\ \sigma_{vr} &= K_v E_v N_p^{-\frac{1}{d}} \end{aligned} \quad (3.40)$$

In Equation 3.40, E_p and E_v denote the lengths of the intervals between the maximum and minimum samples of the position and velocity components before roughening. d is the dimension of the state space ($d = 4$ in this thesis, consisting of two position components and two velocity components). N_p is the number of particles, and K_p and K_v are constant tuning parameters. Larger values of K_p and K_v may excessively blur the distribution, while smaller values may lead to tight clusters of particles [58]. In this thesis, $K_p = 1.5$ and $K_v = 0.75$ are chosen as a trade-off between maintaining particle diversity and preserving the underlying distribution. This roughening process enhances diversity without substantially distorting the posterior distribution.

3.7.5. Initialization

The previous section discussed the particle filter in detail. However, the PF requires an initial state as the starting point. The initial state vector $\mathbf{x}_0$ is defined in Equation 3.41, where the initial velocities in both the x - and y -directions are assumed to be zero. The initial position (x_0, y_0) is obtained from Equation 3.42 by identifying the grid index with the maximum value in the sum of $\tilde{\mathbf{L}}_t$ over the first second, where $n_x^{(0)}$ and $n_y^{(0)}$ denote the initial position in grid indices. Here, $\tilde{\mathbf{L}}_t$ represents the normalized likelihood map, Δt is the time step, and Δx and Δy are the spatial resolutions of the likelihood grid along the x - and y -axes, respectively. By combining the likelihood information over the first second, a stable estimate of the initial position can be obtained.

$$\mathbf{x}_0 = [x_0, y_0, 0, 0]^\top \quad (3.41)$$

$$\begin{aligned} (n_x^{(0)}, n_y^{(0)}) &= \arg \max_{n_x, n_y} \sum_{t=1}^{\frac{1}{\Delta t}} \tilde{\mathbf{L}}_t(n_x, n_y), \\ x_0 &= n_x^{(0)} \cdot \Delta x, \quad y_0 = n_y^{(0)} \cdot \Delta y \end{aligned} \quad (3.42)$$

3.8. Baseline Method

In the previous sections of this chapter, a detailed introduction to the proposed 3D grid-based localization method has been provided. To evaluate its advantages and disadvantages, a baseline method is adopted for performance comparison. The baseline method follows the DBT approach to track a human in the room, following the main DBT structure described in [11], with several modifications to adapt the procedure to the ceiling-mounted UWB radar scenario of this thesis.

The method in [11] is selected because it presents a traditional DBT-based approach for indoor UWB radar localization. Its processing pipeline follows the conventional DBT framework: CFAR detection is first applied to identify the target, the target range is then extracted, and the processed range measurements are used for localization. Finally, a Kalman filter is applied to generate the target's trajectory.

In contrast, the proposed method adopts a TBD approach, directly utilizing the original CIR data to track the person without intermediate detection steps. Therefore, the DBT method in [11] serves as a representative and practical benchmark for comparison, as it reflects a widely used DBT processing pipeline and is relatively straightforward to implement. In this section, the overall pipeline of the baseline method is first introduced, followed by a detailed description of each individual processing block.

3.8.1. Overall Pipeline of the Baseline Method

Figure 3.26 illustrates the processing pipeline of the baseline method. The pipeline starts with clutter suppression applied to the CIR from each radar. The clutter suppression procedure and parameters are identical to those described in Section 3.3, ensuring comparable signal preprocessing conditions between the baseline and proposed methods. After clutter suppression, the processed CIR is passed through a CFAR detector to extract signals of interest. A leading-edge detection step is then applied to identify the first peak, which is taken as the target range. The target range from each radar is subsequently smoothed using a Kalman filter (KF). Next, Taylor localization is performed to estimate the target's position in space based on the filtered ranges. The resulting position estimates are further tracked in the 2D plane using another KF to obtain the final trajectory.

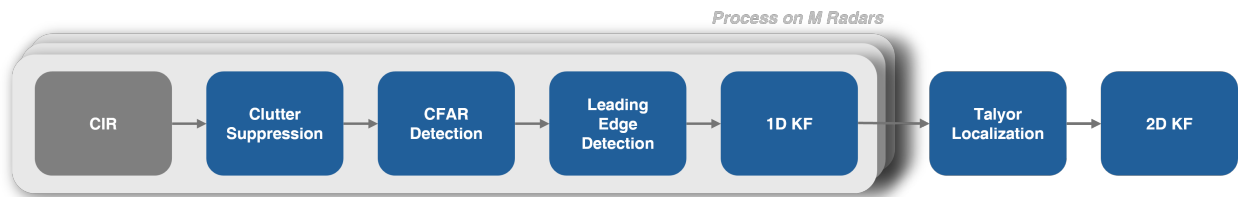


Figure 3.26: Overall pipeline of the baseline method considered for performance comparison.

Compared to the original pipeline in [11], two modifications are introduced to adapt to the ceiling-mounted radar scenario:

1. **Clutter suppression:** It is added to mitigate static reflections from furniture and walls;
2. **Leading-edge detection:** In the original method of [11], where radars are wall-mounted and the human body is treated as an extended target, the centroid of CFAR-detected scatterers is used as an approximation of the target range. However, this centroid-based approach can include short-lived or spurious scatterers, which are subsequently removed through the track management procedure described in [11].

Following the method in [29], a leading-edge detection approach is applied so that the measured range corresponds to the first significant rising edge in the radar return. In the ceiling-mounted radar configuration,

this rising edge of the first peak is typically produced by the subject's head, which is both the closest point to the radar and kinematically stable along the z -axis during walking. Besides, the leading-edge detection approach in [29] is combined with segment computation to filter out small CFAR detections caused by false alarms. Therefore, the track maintenance procedure in [11] is omitted, as the same objective of accurately determination of the target range is achieved by the leading-edge detection approach. In summary, the original CFAR centroid estimation and track maintenance are replaced by the leading-edge detection in the modified baseline method. This simplification is appropriate under the single-target assumption and has been found to yield stable results in the experiment.

3.8.2. CFAR Detection

Based on [11], CA-CFAR is applied to detect the presence of a target within noise and clutter from the processed CIR. CA-CFAR, derived from the Neyman–Pearson criterion, adaptively estimates the detection threshold under the assumption that the probability density function (PDF) of the noise is known [11]. In this process, a sliding window moves along the range profile, with the Cell Under Test (CUT) located at the center and surrounded by guard cells and training cells [59]. The general steps of the CA-CFAR procedure are as follows [59]:

1. **Noise power estimation:** The noise power $\hat{P}_{\text{noise}}(k)$ is estimated from the training cells using Equation 3.43, where Ω_k denotes the set of indices corresponding to the training cells for the k -th CUT, excluding the CUT itself and its guard cells. Here, N is the total number of training cells, and P_i is the CIR signal power in the i -th range bin.

$$\hat{P}_{\text{noise}}(k) = \frac{1}{N} \sum_{i \in \Omega_k} P_i \quad (3.43)$$

2. **Threshold calculation:** The detection threshold $T(k)$ is then computed from the estimated noise power according to the desired probability of false alarm P_{FA} , as given in Equation 3.44.

$$T(k) = N \cdot (P_{\text{FA}}^{-1/N} - 1) \quad (3.44)$$

3. **Target detection decision:** Finally, the CUT power $P_{\text{CUT}}(k, t)$ (signal power in the CUT range bin of slow-time t) is compared with the detection threshold and the estimated noise power to determine whether a target is present based on Equation 3.45. After the detection decision, if a target is detected based on this criterion, the corresponding range bin index k is labeled as one; otherwise, it is labeled as zero. Then, the detection results for all range bins k over each slow-time index t form the binary detection matrix $\mathbf{D}_{\text{det}}(k, t)$.

$$\mathbf{D}_{\text{det}}(k, t) = \begin{cases} 1, & P_{\text{CUT}}(k, t) > T(k, t) \hat{P}_{\text{noise}}(k, t) \\ 0, & \text{otherwise} \end{cases} \quad (3.45)$$

To implement CA-CFAR, the CFARDetector function in MATLAB is employed by specifying the size of the training cells, the size of the guard cells, and the desired probability of false alarm. Following the guidelines in [11], the guard cell length should exceed 1 m , and the training cell length should be approximately 1–2 m , considering the physical size of a typical person. Given the range bin resolution of the experimental device (detailed later in Chapter 5), the guard cell size in the baseline method is set to 100 indices, while the training cell size is set to 250 indices. To balance the trade-off between missing weak targets and generating excessive false alarms, the probability of false alarm P_{FA} is set to 0.05, which ensures sufficient sensitivity for detecting low-power signals while maintaining an acceptable false detection rate for stable tracking.

3.8.3. Leading Edge Detection

In the previous steps, the detection matrix $\mathbf{D}_{\text{det}}(k, t)$ from the CA-CFAR contains the detection result of every range bin in the CIR at every slow-time index. In this step, under the single-person detection assumption, the range bin (i.e., the leading-edge index) corresponding to a nonzero element in the detection matrix is identified for each slow-time index within the dynamic segment. However, during this leading edge estimation process, two issues arise:

1. No nonzero points exist in the detection matrix for certain slow-time indices, resulting in missing target range data for those indices.
2. Clutter-related nonzero points may appear before the leading edge in the range-time index.

To address these issues, for the first case, interpolation is applied to estimate the target range for slow-time indices without detection. For the second issue, a segment computing method (Equation 3.46) is applied, as proposed in [29]. In this method, $\hat{\tau}_{\text{TP}}$ denotes the estimated target range, while τ_l is a candidate detection range. The processing is performed on the slow time index column of the detection matrix $\mathbf{D}_{\text{det}}(:, t)$ at each slow-time index t as follows. First, the matrix is scanned from near to far range bins to locate the first nonzero detection point s_n , which serves as the starting index for a range window of length N_{win} . Next, the number of nonzero detections within this window is counted; if the count is greater than or equal to the threshold N_{seg} , τ_l is accepted as a valid segment start. Finally, among all valid candidates, the smallest (closest) τ_l is selected as the estimated target range $\hat{\tau}_{\text{TP}}$. This segment-based criterion ensures that the leading edge is derived from a sufficiently continuous set of detections, thereby reducing the influence of clutter. After obtaining $\hat{\tau}_{\text{TP}}$ over slow-time, a moving average filter is applied to smooth out small fluctuations. In this thesis, $N_{\text{win}} = 50$ and $N_{\text{seg}} = 10$ is selected by balancing the missed detection rate and false alarm probability to get a better target range estimation result denoted as $\hat{r}_t$ in every slow time index t .

$$\hat{\tau}_{\text{TP}} = \min \tau_l \quad \text{s.t.} \quad \sum_{n=l}^{l+N_{\text{win}}} [s_n > 0] \geq N_{\text{seg}} \quad (3.46)$$

3.8.4. Kalman Filter on Target Range

Following [11], a constant-velocity Kalman filter (KF) is applied to smooth the target range measurements from each radar. The constant-velocity state model is given in Equation 3.47, where r_t and $\dot{r}_t$ denote the target range and range rate at slow-time index t , respectively. Here, Δt is the slow-time sampling interval, and w_{t-1} represents zero-mean white Gaussian process noise with variance σ_w^2 .

$$\begin{bmatrix} r_t \\ \dot{r}_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \dot{r}_{t-1} \end{bmatrix} + \begin{bmatrix} \Delta t^2/2 \\ \Delta t \end{bmatrix} w_{t-1}. \quad (3.47)$$

The KF recursion consists of two main stages: the prediction step and the measurement update step [60]. In the prediction step (Equation 3.48), the prior state estimate $\hat{\mathbf{x}}_{t|t-1}$ and error covariance $\mathbf{P}_{t|t-1}$ are predicted from the previous state using the state transition matrix $\mathbf{F}$ (Equation 3.49) and process noise covariance $\mathbf{Q}$ (Equation 3.50). The state vector is defined as $\mathbf{x}_t = [r_t, \dot{r}_t]^\top$.

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{F} \hat{\mathbf{x}}_{t-1|t-1}, \quad \mathbf{P}_{t|t-1} = \mathbf{F} \mathbf{P}_{t-1|t-1} \mathbf{F}^\top + \mathbf{Q}. \quad (3.48)$$

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad (3.49)$$

$$\mathbf{Q} = \begin{bmatrix} \Delta t^4/4 & \Delta t^3/2 \\ \Delta t^3/2 & \Delta t^2 \end{bmatrix} \sigma_w^2 \quad (3.50)$$

According to the measurement update step, the new target range observation Z_t (value of the target range $\hat{r}_t$ at slow time index t) is used to refine the a priori prediction at slow-time index t . The residual y_t measures the difference between the observation and the predicted measurement, where $\mathbf{H}$ is the measurement matrix and $\mathbf{R}$ is the measurement noise covariance (Equation 3.51). The innovation covariance $\mathbf{S}_t$ quantifies the uncertainty of this residual. The Kalman gain $\mathbf{K}_t$ (Equation 3.52) determines the relative weight assigned to the new observation versus the prediction. Finally, the a posteriori state estimate $\hat{\mathbf{x}}_{t|t}$ and its covariance $\mathbf{P}_{t|t}$ are updated as in Equation 3.53.

$$\begin{aligned} y_t &= Z_t - \mathbf{H} \hat{\mathbf{x}}_{t|t-1}, \\ \mathbf{S}_t &= \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{R}, \\ \mathbf{H} &= [1 \quad 0] \end{aligned} \quad (3.51)$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}^T \mathbf{S}_t^{-1}. \quad (3.52)$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t y_t, \quad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}) \mathbf{P}_{t|t-1}. \quad (3.53)$$

Moreover, the initial state is initialized using the first and second target range measurements to estimate both the initial range and range rate. For the process noise, standard deviation is set as $\sigma_w = 0.2$, and the measurement noise is set as $\mathbf{R} = 0.01$. These values reflect a moderate confidence in the motion model, ensuring that the filter produces a smooth trajectory while still remaining responsive to actual measurement variations. By applying the Kalman filter to the target range from each radar, a smoothed target range sequence from radar m is obtained.

3.8.5. Taylor Series based Localization

Following [11], a Taylor series-based localization method is applied to convert the estimated target ranges $\hat{r}$ from each radar into the localization result. In [11], the target is localized in a 2D plane; however, in the ceiling-mounted radar configuration, the position of the human head is estimated instead. Since the human height is unknown and the head is not located on the same plane as the ceiling-mounted radars, the localization problem is no longer strictly 2D but instead corresponds to a form of 3D localization for the head position. The implementation of the Taylor series-based localization in this work generally follows the approach described in [61].

Since the radar configuration is monostatic, each radar constrains the target's position (x, y, z) to lie on a hemisphere defined by the estimated target range $\hat{r}$. This ellipsoid can be expressed as Equation 3.54, where $P_R = (x_R, y_R, z_R)$ is the positions of the radar, and $P_T = (x, y, z)$ is the target coordinate. The term e represents the residual error between the estimated and actual target location. A function $f_r(x, y, z)$ is defined in Equation 3.55 that represents the measured quantity after subtracting the error term.

$$\hat{r} = \|P_R - P_T\|_2 + e = \sqrt{(x - x_R)^2 + (y - y_R)^2 + (z - z_R)^2} + e \quad (3.54)$$

$$f_r(x, y, z) = \hat{r} - e = \sqrt{(x - x_R)^2 + (y - y_R)^2 + (z - z_R)^2} \quad (3.55)$$

To estimate (x, y, z) , $f_r(x, y, z)$ around an initial guess (x_v, y_v, z_v) is expanded using a Taylor series expansion. The generic Taylor series form is given by Equation 3.56, and applying it to f_r yields the approximation in Equation 3.57. In this expansion, δ_x , δ_y and δ_z represent small increments in x_v , y_v and z_v to iteratively refine the estimated position.

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n + R_n(x) \quad (3.56)$$

$$\begin{aligned}
f_r(x_v, y_v, z_v) + a_x \delta_x + a_y \delta_y + a_z \delta_z &\approx \hat{r} - e \\
a_x &= \left. \frac{\partial f_{R,i}(x, y, z)}{\partial x} \right|_{x=x_v, y=y_v, z=z_v} = \frac{x_v - x_R}{\sqrt{(x_v - x_R)^2 + (y_v - y_R)^2 + (z_v - z_R)^2}} \\
a_y &= \left. \frac{\partial f_{R,i}(x, y, z)}{\partial y} \right|_{x=x_v, y=y_v, z=z_v} = \frac{y_v - y_R}{\sqrt{(x_v - x_R)^2 + (y_v - y_R)^2 + (z_v - z_R)^2}} \\
a_z &= \left. \frac{\partial f_{R,i}(x, y, z)}{\partial z} \right|_{x=x_v, y=y_v, z=z_v} = \frac{z_v - z_R}{\sqrt{(x_v - x_R)^2 + (y_v - y_R)^2 + (z_v - z_R)^2}}
\end{aligned} \tag{3.57}$$

Since there are M radars in the considered setup, these equations are combined into a matrix form as shown in Equation 3.58, where the subscript notation $_m$ denotes values corresponding to the radar m . In this matrix formulation, δ encapsulates the incremental adjustments in (x_v, y_v, z_v) , and b reflects the difference between the estimated ranges $\hat{r}$ and the function value $f_r(x_v, y_v)$. Minimizing e ensures that the chosen position (x_v, y_v, z_v) best fits all radar range constraints.

$$A\delta = b - e, \quad A = \begin{bmatrix} a_{x_1} & a_{y_1} & a_{z_1} \\ a_{x_2} & a_{y_2} & a_{z_2} \\ \vdots & \vdots & \vdots \\ a_{x_m} & a_{y_m} & a_{z_m} \\ \vdots & \vdots & \vdots \\ a_{x_M} & a_{y_M} & a_{z_M} \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_x \\ \delta_y \\ \delta_z \end{bmatrix}, \quad b = \begin{bmatrix} \hat{r}_1 - f_{r_1} \\ \hat{r}_2 - f_{r_2} \\ \vdots \\ \hat{r}_m - f_{r_m} \\ \vdots \\ \hat{r}_M - f_{r_M} \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \\ \vdots \\ e_M \end{bmatrix} \tag{3.58}$$

To solve for δ , a least squares solution is employed, as in Equation 3.59. The target coordinates (x_v, y_v, z_v) are updated iteratively using Equation 3.60 until δ is below a small threshold ϵ , indicating convergence.

$$\delta = (A^\top A)^{-1} A^\top b \tag{3.59}$$

$$x_v \leftarrow x_v + \delta_x, \quad y_v \leftarrow y_v + \delta_y, \quad z_v \leftarrow z_v + \delta_z \tag{3.60}$$

The iterative procedure can be summarized as follows:

1. **Calculate** $f_r(x_v, y_v, z_v)$: For the first iteration, set $(x_v, y_v, z_v) = (0, 0)$. For following iterations, use the updated (x_v, y_v, z_v) from the previous iteration. When processing the next slow-time index, the initial guess (x_v, y_v, z_v) is set to the estimated coordinates obtained from the previous slow-time index.
2. **Compute A and b**: Evaluate A and b as defined in Equation 3.58.
3. **Solve for δ** : Use Equation 3.59 to obtain δ .
4. **Update** (x_v, y_v, z_v) : Apply Equation 3.60.
5. **Check for Convergence**: If $\|\delta\| < \epsilon$, stop the iterations; otherwise, continue.

Based on this method, the position of the target's head can be estimated in 3D space. However, since the final trajectory output is required in 2D, the z -coordinate is discarded, retaining only the x - and y -coordinates for generating the 2D trajectory sequence (denoted as $(\hat{x}(t), \hat{y}(t))$) and for the following 2D Kalman filter tracking.

3.8.6. Kalman Filter on 2D Trajectory

After the Taylor series-based positioning, a sequence of 2D position observations is obtained. The subsequent 2D Kalman tracking step aims to smooth the estimated trajectory and improve position accuracy. The 2D Kalman filter follows the same recursive formulation as the 1D case in subsection 3.8.4, using the prediction and update equations in Equation 3.48–Equation 3.53, but extending the state vector, measurement model, and initialization to the 2D case [11]. The state vector adopts a nearly constant velocity model as in Equation 3.61, where (x_t, y_t) are the target coordinates and $(v_{x,t}, v_{y,t})$ are the corresponding velocities. The state transition matrix $\mathbf{F}$ and process noise covariance $\mathbf{Q}$ are defined in Equation 3.62 and Equation 3.63, respectively, with σ_w^2 denoting the process noise variance. The measurement vector $\mathbf{Z}_t$ and measurement matrix $\mathbf{H}$ are given in Equation 3.64 and Equation 3.65, where $\mathbf{R}$ is the measurement noise covariance matrix. The initial state $\hat{\mathbf{x}}_{0|0}$ is computed from the first two position observations, as shown in Equation 3.66. To achieve an appropriate balance between the confidence in the motion model and the reliability of the measurements, the process noise standard deviation $\sigma_w = 0.1$ and the measurement noise standard deviation $\sigma_x = \sigma_y = 0.01$ are selected. This choice ensures that the Kalman filter maintains a smooth trajectory while remaining responsive to actual target movements.

$$\mathbf{x}_t = [x_t, y_t, v_{x,t}, v_{y,t}]^\top. \quad (3.61)$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (3.62)$$

$$\mathbf{Q} = \sigma_w^2 \begin{bmatrix} \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} & 0 \\ 0 & \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & 0 & \Delta t^2 & 0 \\ 0 & \frac{\Delta t^3}{2} & 0 & \Delta t^2 \end{bmatrix}. \quad (3.63)$$

$$\mathbf{Z}_t = [\hat{x}(t), \hat{y}(t)]^\top. \quad (3.64)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{R} = \text{diag}(\sigma_x^2, \sigma_y^2). \quad (3.65)$$

$$\hat{\mathbf{x}}_{0|0} = \begin{bmatrix} \hat{x}(1) \\ \hat{y}(1) \\ \frac{\hat{x}(2) - \hat{x}(1)}{\Delta t} \\ \frac{\hat{y}(2) - \hat{y}(1)}{\Delta t} \end{bmatrix}. \quad (3.66)$$

3.9. Summary

In this chapter, the proposed 3D grid-based method and the baseline method for tracking and localization are described. In the proposed 3D grid-based method, the extended target characteristic of a walking person is exploited to achieve more accurate position estimation. The CIR data from all radars are mapped and fused into a 3D grid, and compensation procedures are applied to ensure that the projection onto the 2D likelihood map is accurate. A particle filter is then applied to the 2D likelihood map to perform target tracking.

The baseline method following a similar pipeline of [11] is used for comparison. This method adopts a traditional DBT framework in which the target range from CIR in each radar is detected using CFAR and leading edge detection. The estimated ranges are then processed using Taylor series based localization to obtain the target position, and a Kalman filter is applied to produce a smoothed trajectory. In addition, the comparison between the proposed method and the baseline method will be evaluated in Chapter 5 based on real life data.

4

Activity Recognition

In this chapter, the proposed activity classification method is presented, addressing the tasks of walk–static classification and stand–sit classification. The proposed method utilizes signal energy and human height for activity recognition and can be applied effectively even when only limited data are available, compared to more data-hungry methods that employ machine learning.

4.1. General Introduction of Activity Recognition

As described in Section 3.2, the proposed 3D grid-based method is built upon mapping the CIR onto a 3D grid to enable both tracking and activity recognition (including walk, stand and sit). In the proposed pipeline shown in Figure 3.3, the tracking and activity recognition modules are integrated: the localization module employs the walk–static classification from the recognition module to determine whether the person is walking or in a static posture.

When the person is static, the particle filter for tracking is not executed; it operates only when the person is walking. Conversely, the recognition module requires the human position information to perform sit–stand classification, which is applied only when the person is static, based on the assumption that sitting and standing occur within the static category. Therefore, the recognition module consists of two main components: walk–static classification and sit–stand classification.

4.2. Walk–Static Classification

To distinguish between walking and static activities using a limited amount of data, the approach in [44] is adopted, which compares the CIR corresponding to human movement and to a static posture after applying an MTI filter. Since [44] does not provide detailed implementation details, a signal-energy-based walk–static classification method is proposed.

The overall pipeline shown in Figure 4.1 compares the signal energy during the movement period (when the person is walking) with that during a background period (CIR measured before the person starts walking), determines which period exhibits higher similarity to the background, and finally applies static period extension to generate the walk–static classification.

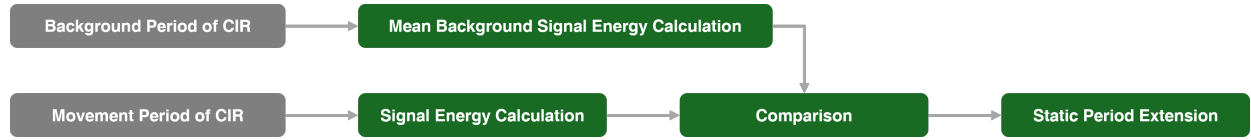


Figure 4.1: Pipeline of the proposed walk-static classification procedure.

4.2.1. Background Signal Energy

The adaptive clutter suppression method described in Section 3.3 suppresses static clutter and highlights moving objects, and can thus serve as the MTI stage in [44]. In the data collection process, the first several seconds of CIR data are recorded in an empty room, allowing the adaptation ratio in the clutter suppression filter to converge to a stable value. This initial recording period before any human motion is treated as the background period, and represents a static reference signal for later comparison.

As illustrated in Figure 3.4, an example test record shows human movement between approximately 6 s and 17 s; after the walking motion stops, several extra seconds are retained to ensure the complete capture of movement related data, while all other segments belong to the background period. It is observed that the CIR magnitude increases during walking, and this effect can be quantified using the signal energy. To compare the signal energy between background and movement period, the background signal energy is calculated at first. The background signal energy E_{bg} is computed as in Equation 4.1, where $c_{m,t}$ denotes the CIR after clutter suppression for radar m at slow-time index t , K is the number of range bins, T is the total number of slow-time indices in the background period, and M is the number of radars.

$$E_{bg} = \frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T \sum_{k=1}^K c_{m,t}(k)^2 \quad (4.1)$$

4.2.2. Movement Signal Energy

To compare the signal energy between the background and movement periods, the movement signal energy $E_s(t)$ is computed at each slow-time index t using Equation 4.2, where the CIR energy after clutter suppression is averaged across all M radars at time t . To reduce the impact of noise and outliers, the resulting $E_s(t)$ is then smoothed using a moving average filter.

$$E_s(t) = \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^K c_{m,t}(k)^2 \quad (4.2)$$

The smoothed movement energy is then compared with the background energy E_{bg} to determine whether the target is walking or static, resulting in the binary classification sequence $\mathbf{a}(t)$ defined in Equation 4.3. If $E_s(t)$ exceeds the background energy, the target is considered to be walking; otherwise, the energy is deemed similar to the background, indicating a static state. A thresholding factor $\eta = 2$ is introduced to make the decision more robust, accounting for slight energy fluctuations during static periods and for the fact that E_{bg} itself is computed as an average and may contain some noise values. Based on this comparison result, $\mathbf{a}(t)$ is a label set to 1 for walking period and 2 for static period. By using this method, walking and static states can be classified prior to the tracking procedure, thereby reducing unnecessary tracking computations for tracking in the static state.

$$\mathbf{a}(t) = \begin{cases} 1, & E_s(t) \geq \eta E_{bg} \\ 2, & \text{otherwise} \end{cases} \quad (4.3)$$

4.2.3. Static Period Extension

After a static period is detected over all slow-time indices, it is further extended to account for activity transitions. The static period is assumed to contain two activity types: standing and sitting. For standing, the transition between walking and standing is negligible. However, in the sitting case, during the transition from walk/stand to sit or sit to walk/stand, the person's height changes rapidly in approximately 0.5 s (based on real-life test recordings), which can significantly affect tracking accuracy. Since this transition is short in duration and typically occurs at a fixed location, it is reasonable to treat it as part of the static period by temporarily suspending tracking also during this transition period. In addition, the ground-truth activity annotation records the time when the person is fully seated and static. To ensure consistency with the ground truth, the extended period is restored in the final classification results so that it matches the annotation. Considering that the MTI filter used in this work requires time to adjust its adaptation ratio to static signal strength and to recover when returning to movement, an extra extension of 1 s is added before the static period and 0.5 s after the static period, as observed in the experimental data.

Since the specific static activity type (standing or sitting) is not known during detection, the same extension is applied to all static periods. This extension process is illustrated in Figure 4.2. In total, 1.5 s is added before the start of each static period and 1 s is added after its end. Based on this extension, the static period is expanded to cover possible transition phases, thereby improving tracking accuracy.

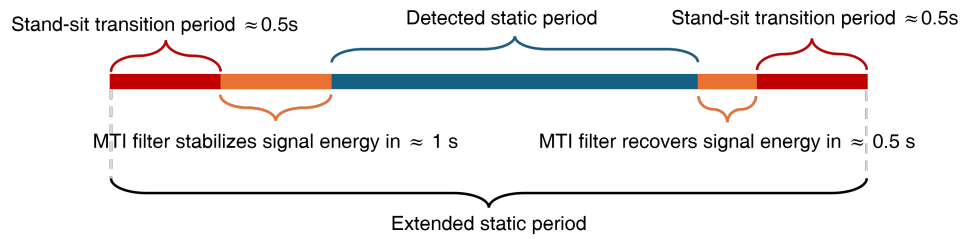


Figure 4.2: Illustration of static period extension to account for transition periods.

4.3. Stand-Sit Classification

Based on the strategies in [45] and [46], standing and sitting can be classified using the person's height. As reported in [46], a height decrease of approximately 20% can indicate that the person is sitting. Following this principle, the proposed classification approach determines the height change during the stand-sit transition period: if a large height drop is observed in the walk/stand-to-sit transition and a large height increase occurs in the sit-to-walk/stand transition, the person is classified as sitting during the related static period.

Figure 4.3 illustrates the pipeline of the proposed stand-sit classification procedure. In this procedure, the reflection grid $\tilde{\mathbf{R}}_t$ (describe in Equation 3.14) is cropped to the human body region in x - and y - axes at the estimated position obtained from the localization pipeline. The human height is then estimated at each slow-time index to form a height sequence. Finally, based on this height sequence, a height change decision is made by detecting significant height drops or increases to determine whether the person is standing or sitting.



Figure 4.3: Pipeline of the proposed stand-sit classification procedure.

4.3.1. Human Body Region Construction

To estimate the human height change, the human body region must first be determined within the reflection grid¹ $\tilde{\mathbf{R}}_t$. The value in $\tilde{\mathbf{R}}_t$ is proportional to the target presence probability at each voxel in the whole grid, with higher values indicating a higher likelihood of the target being present at that voxel. Since the head is the closest part of the human body to the radar, it should occupy the voxel with the highest reflection value. However, due to a limited number of artifacts in the grid after compensation, the global maximum voxel value is not always a reliable indicator of the head position for height estimation. Instead, a local maximum can be computed within the human body region based on the estimated human location.

From the localization pipeline, the approximate position of the person is obtained, which can be regarded as the centroid of the human body. It is assumed that a human body is not represented by a single vertical voxel column in the grid, but has a finite width. Moreover, the localization result may contain positioning errors. Thus, the human body region is extended from the estimated location as shown in Figure 4.4. Specifically, the body region is modeled as a cuboid with a width and length of 0.5 m, and a height covering all z -layers cropped from the reflection grid $\tilde{\mathbf{R}}_t$. This process yields the 3D human body region $\mathbf{B}_t$ for each slow-time index t with size $\tilde{N}_x, \tilde{N}_y, N_z$.

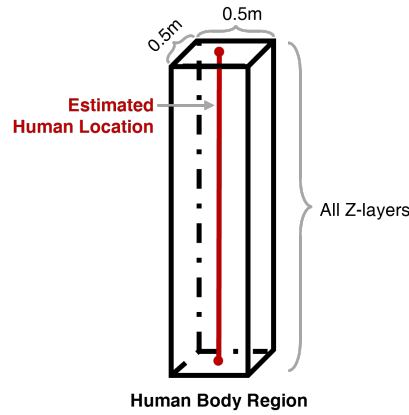


Figure 4.4: Illustration of the extended body region represented by a cuboid.

4.3.2. Human Height Estimation

Based on the cropped region from $\tilde{\mathbf{R}}_t$, the human body region $\mathbf{B}_t$ is determined. The human height is then estimated from the local maximum voxel point within this region. To obtain a more robust result, Equation 4.4 is applied to estimate the human height sequence $\mathbf{h}(t)$ by finding which n_z vertical layer contains maximal average voxel value in every z -layer in $\mathbf{B}_t$ at slow-time index t . The value of $\mathbf{h}(t)$ denotes the index along the z -dimension corresponding to the head position, representing the estimated human height in units of the z -index. In Equation 4.4, each voxel in the human body region $\mathbf{B}_t$ is denoted as $(\tilde{n}_x, \tilde{n}_y, n_z)$ and $\tilde{N}_x, \tilde{N}_y$ are the dimensions of $\mathbf{B}_t$ along the x - and y -axes, respectively. The estimated height sequence is then smoothed using a moving average filter to suppress noise and reduce the effect of occasional erroneous height estimates, thereby producing a more stable and reliable human height trend.

$$\mathbf{h}(t) = \arg \max_{n_z} \left(\frac{1}{\tilde{N}_x \tilde{N}_y} \sum_{\tilde{n}_y=1}^{\tilde{N}_y} \sum_{\tilde{n}_x=1}^{\tilde{N}_x} \mathbf{B}_t(\tilde{n}_x, \tilde{n}_y, n_z) \right) \quad (4.4)$$

¹This grid is defined in Equation 3.14 and differs from the target reflection probability grid $\mathbf{R}_t$, which represents the target presence probability in each layer.

4.3.3. Height Change Decision

Based on $\mathbf{h}(t)$, the human height sequence can be estimated. The height change is defined in Equation 4.5, where $\Delta h(t_s)$ denotes the change in height between the start and end time indices of a static period, with t_s representing either the start or end index for simplicity. Δt_s is the window length used to detect the height change. In this thesis, $\Delta t_s = 1$ s is selected, as the duration of a stand-up or sit-down transition is approximately 0.5 s adding with extra 0.5 s as buffer time for a healthy normal person.

$$\Delta h(t_s) = \mathbf{h}(t_s + \Delta t_s) - \mathbf{h}(t_s - \Delta t_s) \quad (4.5)$$

For each start and end point of a static period, the corresponding height change is calculated. If the height change is greater than +1 z -layer index, this indicates that the person's height has increased by at least $\Delta z = 33.75$ cm, which exceeds 20% of the total height for a person taller than 1.68 m. Conversely, if the height change is less than -1 z -layer index, this indicates that the person is sitting down. This relationship is expressed in Equation 4.6, where $\mathbf{s}(t_s)$ is the label of the sitting activity at time t_s : label '0' denotes no height change, label '1' denotes sitting down, and label '2' denotes standing up.

$$\mathbf{s}(t_s) = \begin{cases} 0, & |\Delta h(t_s)| < 1 \\ 1, & \Delta h(t_s) \leq -1 \\ 2, & \Delta h(t_s) \geq 1 \end{cases} \quad (4.6)$$

Based on the sitting activity label, the final activity type between standing and sitting can be determined. If the start index and end index of a static period are labeled as sit down and stand up, respectively, the entire static period is classified as sitting; otherwise, it is classified as standing. Using this criterion, the activity list $\mathbf{a}(t)$ is finalized in Equation 4.7, where different labels are assigned to different slow-time indices t .

$$\mathbf{a}(t) = \begin{cases} 1, & \text{walk} \\ 2, & \text{stand} \\ 3, & \text{sit} \end{cases} \quad (4.7)$$

4.4. Summary

Based on the proposed activity recognition method described in this chapter, the activity type can be classified with low data requirements, as the approach primarily relies on physical body features and thresholding. Compared to more advanced approaches using machine learning or neural networks, this characteristic makes the method suitable for scenarios with limited sensing data, enabling efficient deployment in practical environments without the need for complex training datasets. Experimental results for the validation of this approach are presented in the following chapter.

Experimental Setup and Evaluation

This chapter introduces the experiment setup including selected radar, experimental environment and test plan. Based on the setup, the tracking and activity evaluation result achieve up to RMSE 0.23 m and classification accuracy 98%, respectively.

5.1. Radar Configuration

In the previous chapters, the detailed methodology of tracking and classification of this thesis has been described. To validate the proposed approach in a real-world setting, it is necessary to present the specifications of the selected radar and the details of its deployment.

5.1.1. Radar Selection

At imec and TU Delft, two types of commercial UWB radars are available for use in this thesis: the Novelda XeThru X4 and the Novelda XeThru X7. In this section, a detailed comparison between these two radars is presented, including their specifications and data collection method.

Radar Specifications

The Novelda XeThru X4 is an impulse-radio UWB radar transceiver System-on-Chip (SoC) operating at center frequencies of 7.29 GHz or 8.748 GHz [62]. It integrates a transmitter, a direct RF-sampling receiver, a fully programmable system controller, and advanced power management in a compact chip-scale package. The receiver achieves a typical sampling rate of 23.328 GS/s with a maximum detection range of approximately 9.8 m [62][63]. The device supports a Serial Peripheral Interface (SPI) for raw radar data output, features low-power duty-cycle-controlled operation, and uses differential RF terminals for high sensitivity in both static and dynamic applications. The X4 has been widely applied in indoor localization, vital-sign monitoring, fall detection, and human activity recognition [63][64].

In this thesis, two radar module types containing the Novelda XeThru X4 chip are available: two X4M03 modules and one X4M300 module. Since the X4-based radar modules are no longer in production, these three modules are the only available units for experiments, yet the proposed method remains agnostic to the radar type. Both modules share the same specifications; the primary difference is that the X4M03 consists of two separate boards (an antenna

board and an SoC board), whereas the X4M300 integrates all components, including the chip and antennas, onto a single board [64][65]. The images of these two modules are shown in Figure 5.1. The distinction between the two radar modules is related application. While the Novelda X4M300 integrates advanced embedded algorithms for respiration and heartbeat detection, the fundamental sensing performance of the underlying X4 radar chip remains identical across both modules [64][65]. Thus they are collectively referred to as “X4” for simplicity. A detailed summary of the radar specifications is provided in Table 5.1.

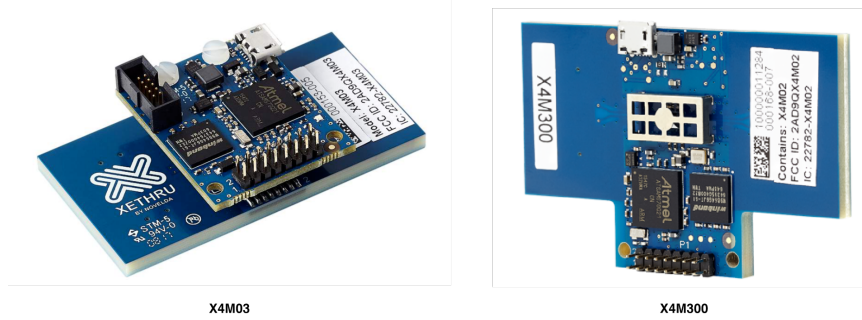


Figure 5.1: Image of Novelda X4M03 and Novelda X4M300 [64][65].

The Novelda XeThru X7 UWB chip is integrated into the X7F202 shield module, as shown in Figure 5.2, with four such modules available for this thesis. The X7F202 is a multichannel impulse-radio UWB radar transceiver module operating at a center frequency of 7.875 GHz [21]. It integrates $2 \times \text{Tx/Rx}$ with antennas, enabling digital beamforming to improve spatial resolution, detection accuracy, and Angle of Arrival (AoA) estimation [21].

The module incorporates a transmitter, a direct RF-sampling receiver, clock systems, decoupling capacitors, and a fully programmable system controller, along with advanced power management capable of reducing power consumption to below $100 \mu\text{W}$ in specific use cases. The receiver achieves a typical sampling rate of 2.1 GS/s with a maximum detection range of approximately 13.6 m, and supports low-power duty-cycle-controlled operation for long-term embedded deployments [21]. The X7F202 offers high sensitivity in both static and dynamic environments, making it suitable for applications such as low-power presence detection, smart building automation, consumer electronics sensing, and indoor human activity monitoring [21]. A detailed summary of the radar specifications is provided in Table 5.1.



Figure 5.2: Image of Novelda X7F202 [66].

Table 5.1 summarizes the key specifications of the X4 and X7 radars. The values are summarized and derived from [67, 68]. According to the transmit bandwidth B_{Tx} , the X4 provides nearly twice the bandwidth of the X7, resulting in range resolutions of 10 cm and 21.4 cm, respectively. The range resolution is defined in Equation 5.1, where c is the speed of light.

$$\Delta R = \frac{c}{2B_{\text{Tx}}} \quad (5.1)$$

For the range-bin size, which reflects the fast-time sampling rate, the expression is given in Equation 5.2, where f_s denotes the sampling rate. Since the X4 operates with a much higher sampling rate than the X7, its range-bin size is finer, potentially enabling the detection of smaller movements in human activity.

$$\Delta R_{\text{bin}} = \frac{c}{2f_s} \quad (5.2)$$

In addition, the X4 has a higher transmit peak power than the X7, leading to better received signal-to-noise ratio (SNR) [59]. The receiver noise figure of the X4 is also lower than that of the X7, further enhancing its SNR performance [69].

Although the X7 offers advanced functionalities such as dual Tx/Rx channels for AOA estimation and improved energy efficiency, as well as wider availability compared to the discontinued X4 modules [65, 21], the X4 still provides superior performance in terms of bandwidth, sampling frequency, and SNR, which are critical for signal quality. Therefore, despite the limited availability of only three X4 radars, they remain the preferred choice in this thesis, as they are sufficient to support the minimum number of units required for localization, i.e. 3.

Table 5.1: Specifications of Novelda X4 and X7 Radars [67] [68].

Radar Type	B_{Tx}	Tx Peak Power	Rx Noise Figure	f_s	Transceiver	AoA	Range Resolution	Range Bin Size
X4	1.5 GHz	6.4 dBm	6.3 dB	23.3 GHz	1Tx+1Rx	No	10 cm	0.64 cm
X7	0.7 GHz	2.5 dBm	8.3 dB	2.1 GHz	2TRx	Yes	21.4 cm	7.1 cm

Radar Data Collection

Because this thesis aims to have its validation work on data collected in real life, the ease of data acquisition must also be considered. Since multiple radars are required to operate simultaneously for localization and tracking, the data collection process needs to handle concurrent acquisition from multiple devices. In particular, for localization and tracking experiments where the target is continuously walking, accurate time synchronization between radars is essential to ensure data consistency and reliability.

For the *Novelda X4 radar*, which was released in 2017, the related data acquisition programs provided by the company and third-party developers are relatively mature and easy to access [70]. In this thesis, the radar data acquisition process is modified based on the MATLAB program presented in [71] to realize the data acquisition process for multiple Novelda X4 UWB radars on one host computer. In this program, data from three X4 radars are collected sequentially within a single host computer, as illustrated in Figure 5.3. Because the data reading speed in MATLAB is sufficiently fast with respect to the observed movements, this sequential acquisition can be regarded as quasi-simultaneous, thereby supporting multi-radar data collection for the experiments.



Figure 5.3: Illustration of Novelda X4 data acquisition program.

When it comes to the *Novelda X7 radar*, which was only released in 2024, the data acquisition programs are not fully developed by Novelda at the time of this thesis. For instance, the official development kit does not fully

support Linux or Raspberry Pi, and it also lacks the capability of reading and acquiring data from multiple radars in a single host computer, even without simultaneous acquisition requirements. This limitation arises because the development kit restricts to connect only one Novelda X7 radar on the communication ports (COMs) in a single host computer. To collect data from multiple X7 radars for further analysis, a virtual machine-based data acquisition pipeline is fully developed for Novelda X7 in this thesis, as illustrated in Figure 5.4. It is noteworthy that, at the time of this thesis, only two papers have been published that utilize the Novelda X7 radar for data collection, focusing on people counting and activity classification, respectively [72] [73]. Both of these works are limited to single radar experiments.

Thus, to the best of the author's knowledge, this thesis is the first to realize data collection with multiple Novelda X7 radars from one host computer at the same time .

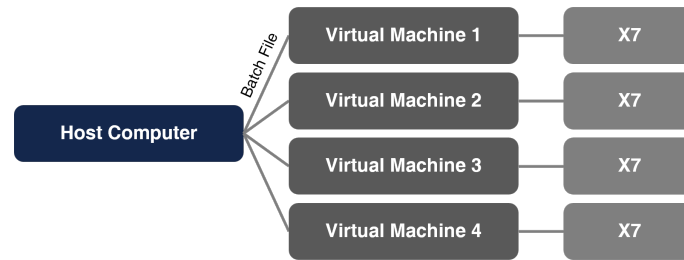


Figure 5.4: Illustration of Novelda X7 data acquisition pipeline.

In this pipeline (Figure 5.4), four virtual machines are installed on the host computer. Each virtual machine runs a Windows 11 operating system with the Novelda official development kit installed, and the virtualization platform is VirtualBox [74]. An individual COM port is assigned to each virtual machine, ensuring that the development kit inside each environment can only detect and communicate with a single radar. Following this strategy, four X7 radars are connected to four dedicated COM ports on the host computer, with each port wholly managed by one virtual machine.

The control between the host computer and the development kits is realized through batch files: once executed on the host, these batch files trigger the development kits inside all virtual machines to start data recording at the same time. To address synchronization issues across the four virtual machines and the host computer, Windows automatic time updates are disabled and the local clocks are manually reset when the data acquisition pipeline starts in every virtual machine. With this optimization, the four virtual machines are able to collect Novelda X7 radar data in nearly synchronized time, storing the CIR sequences for each slow-time index as .csv files in a shared folder on the host computer.

The Novelda X7 development kit is further modified to support the proposed multi radar data acquisition pipeline and to improve usability. The main modifications are summarized as follows:

1. **Data saving functionality:** The original development kit does not provide an option to store the collected radar data on the computer. Therefore, a data saving module is implemented, enabling each radar to store its acquired data into a dedicated .csv file.
2. **Radar configuration management:** In the original development kit, modifying radar parameters required embedding the configuration into the source code, recompiling the program, and then executing it. This process had to be repeated for every configuration update, which becomes highly impractical when four virtual machines with four development kits are used in parallel. To address this issue, the radar configuration is externalized into a separate configuration file. During runtime, each development kit automatically reads its

corresponding configuration file stored in the shared folder on the host computer to set the radar parameters automatically. By decoupling the compilation process from the parameter configuration process system operation, the maintenance is significantly simplified, making the pipeline more efficient and user-friendly.

Although the proposed pipeline enabled multiple Novelda X7 radars to be reliably operated in parallel with improved usability, the virtual machine-based pipeline introduces several disadvantages that significantly reduce data collection efficiency:

1. **High computational and storage requirements:** Since a total of five Windows 11 systems are running on the host computer (four in virtual machines and one on the host), the minimum RAM requirement reaches 20 GB (4 GB per Windows 11 instance) [75]. In addition, the storage demand is also substantial, with at least 188 GB required for the four virtual machines. Thus, both memory and storage costs are considerably high for the data acquisition process, especially when compared to the Novelda X4 data acquisition process, which only requires several MATLAB scripts running on a single host computer.
2. **System instability:** As the proposed pipeline relies entirely on virtual machines with independent instances of the development kit, robustness is limited. Various issues may occur during data collection, such as synchronization failures that lead to delay among radars, or system freezes caused by high computational load, which may require restarting or even resetting the virtual machines. In contrast, the Novelda X4 data acquisition program, implemented entirely in MATLAB on a single host computer, provides a much more stable and robust solution for data collection.

Therefore, even if the proposed X7 data collection pipeline enables multi radar data acquisition for the first time, the X4-based solution remains the preferred choice in terms of efficiency, robustness, and ease of deployment. Considering both radar specifications and the data collection process, the Novelda X4 is ultimately selected as the radar platform for this thesis, due to its larger bandwidth, higher SNR, and more robust multi radar data acquisition procedure.

5.1.2. Radar Deployment

Based on the radar selection results, the Novelda X4 UWB radar is chosen to validate the proposed 3D grid-based method. This sub-section focuses on the radar deployment, including the configuration parameters of the X4 radar and its ceiling-mounted installation.

Novelda X4 Configuration Parameters

For validation, the configuration parameters of the Novelda X4 radar are summarized in Table 5.2.

Table 5.2: Configuration Parameters of Novelda X4 UWB Radar.

Radar	Radar Number	FPS	Center Frequency (GHz)
X4	3	50 Hz	7.29 GHz

The frame per second (FPS) parameter represents the slow-time sampling frequency, which is set to 50 Hz in this thesis to balance signal quality and the requirements of the tracking experiments. A lower FPS would reduce the temporal resolution of the trajectory and potentially cause missed detections during tracking. Conversely, a higher FPS decreases the time interval between two consecutive frames, thereby limiting the number of pulses that can be coherently integrated within each frame [62]. This reduction in integration directly decreases the processing gain

and consequently degrades SNR [62]. Therefore, an FPS of 50 Hz is adopted as a practical compromise, providing sufficient temporal resolution for human motion while preserving adequate SNR for robust signal detection. In addition, the center frequency of the X4 radar is configured at 7.29 GHz to comply with European Telecommunications Standards Institute (ETSI) regulations [62].

Radar Installation

To mount the Novelda X4 UWB radar on the ceiling, the sensor is fixed onto a rigid cardboard foundation using cable ties, and the entire base is attached to the ceiling with carpet tape. This setup enlarges the contact surface area between the radar module and the ceiling, while also avoiding the use of metal materials that could introduce unwanted reflections. The cardboard foundation is selected because it is lightweight, non-conductive, and easy to modify, making it suitable for repeated installation and adjustment. Figure 5.5 illustrates the ceiling-mounted installation of the Novelda X4 radar.



Figure 5.5: Image of the Novelda X4 radar installed on the ceiling.

5.2. Experimental Setup

From the previous section, the Novelda X4 UWB radar operating at 50 Hz FPS is selected for data collection. In this section, the experimental setup is described in detail, including the characteristics of the room environment, the procedure for ground truth data measurement, and the design of the experimental scenarios. These experiments are conducted to provide the datasets required for evaluating the proposed 3D grid-based method.

5.2.1. Experiment Office Room

As briefly described in Section 1.2.1, the experimental data is collected in a dedicated office within the ‘Workplace Vitality Hub’, which is a data-driven smart office facility [14]. The selected room measures 6m in length, 3.5m in width, and 2.7m in height, and is equipped with furniture. An image of the room is shown in Figure 1.2, while its floor plan is provided in Figure 5.6. In Figure 5.6, the red triangles indicate the locations of the three radars installed on the ceiling. The placement of the radars is determined according to the following criteria:

1. Radars should not be positioned directly above large furniture to avoid extreme clutter.
2. The positions should be sufficiently scattered to improve PDOP.
3. The radars should not be too widely separated, as this would cause large open-angle issues in the center of the room.

Based on these considerations, the radar locations are selected in the red triangles location in Figure 5.6. In detail, the specific location of Radar 1 is (1.195 m, 4.06 m, 2.67 m), Radar 2 is (1.06 m, 1.76 m, 2.67 m), and Radar 3 is

(2.5 m, 3.55 m, 2.67 m)¹. The radar positions are measured at the center of the antenna surface of each UWB radar mounted on the ceiling.

In Figure 5.6, several pieces of furniture, including two tables, a plant, and a chair, are shown to ensure that the test environment resembles a realistic office setting. In addition, several labeled points are marked with blue circles in the room, which are used to measure the ground truth positions. The details of this ground truth measurement procedure will be described in the next section.

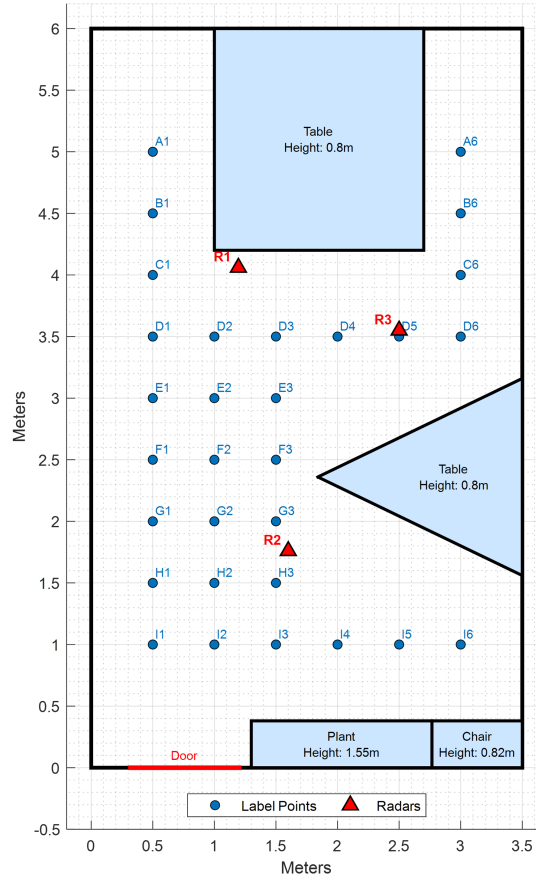


Figure 5.6: The floor plan of the office room for experiment.

5.2.2. Ground Truth Data Measurement

To further evaluate the proposed 3D grid-based method, ground truth data are required to represent the actual walking trajectory and activity types over the entire time sequence. Due to time constraints, this thesis adopts a ticker-timer based ground truth measurement approach for trajectory using floor label points (shown as blue circles in Figure 5.6 and thick black lines sticking on the floor in Figure 1.2), instead of employing motion capture devices.

According to the trajectory ground truth, the general principle of the ticker-timer based ground truth measurement is to record the arrival time when the person passes each floor label point, from which the complete walking trajectory can be reconstructed. The label points are evenly placed on the floor with a horizontal and vertical spacing of 0.5 m. The distance from the rightmost and leftmost label points to the corresponding walls is 0.5 m, while the distance from the topmost and bottommost points to the corresponding walls is 1 m, ensuring a uniform distribution of label points within the room. In addition, each labeled point has its own ID. The nine rows are denoted from

¹It should be mentioned that the radar placement is not derived from mathematical optimization, since determining the optimal ceiling-mounted radar position is beyond the scope of this thesis. The placement is chosen just to support the validation of the proposed tracking and activity recognition algorithms.

‘A’ to ‘I’ and the six columns are numbered from ‘1’ to ‘6’, forming label points from ‘A1’ to ‘I6’. If a label point is covered by or located too close to furniture, it is omitted from the labeling.

Since the spacing between two adjacent label points is fixed at 0.5 m and the walking speed inside the office room is assumed to be stable, the validation experiment only considers straight-line walking between consecutive labels. Under this assumption, the walking speed between two consecutive label points can be regarded as constant. By recording the arrival time at each label point, the intermediate trajectory between two points can be estimated.

In practice, the person under test is instructed to use the timer function on a smartphone to record the arrival time at each label point along the designed trajectory. The timestamp is noted when the person’s entire body covered the corresponding floor label while visually aligning with it. This ground-truth measurement is performed simultaneously with radar data collection to ensure temporal alignment and enable direct comparison. When the radar starts recording, a program-generated beep from the computer is triggered to notify the participant to start the timer on the smartphone. An example of the ground truth measurement procedure for trajectory evaluation is illustrated in Figure 5.7, where the person is walking along the trajectory in the office room while using the timer to record the timestamps.



Figure 5.7: The person records timestamps while walking to provide ground truth for tracking evaluation.

Similar to the trajectory ground truth, the activity ground truth is also recorded using the timer of a smartphone. Except for walking, the other activities such as standing and sitting are only performed at the labeled point locations. The person under test records the start and end times of each activity. For example, the sitting activity is defined as the period between the moment when the person is fully seated and the moment when the person decides to stand up. This indicates that both sitting and standing activities are completely static during their respective time intervals. By applying this procedure, the ground truth activity labels for the entire sequence can be obtained.

This ground truth measurement method, while not as precise as a motion capture system, is still considered acceptable for this thesis. The inaccuracies comes from human factors, such as inconsistent walking speed, a person’s inability to walk in a perfectly straight line, and the human reaction time when recording timestamps. However, these errors are within an acceptable range for the purpose of this thesis: to validate the general performance of the proposed 3D grid-based method and to identify the performed activity rather than the exact timing of the transition. This simple and cost-effective approach give the chance to collect data efficiently, providing a consistent ground truth for evaluating the radar system’s performance under real world conditions.

5.2.3. Experiment Content

From the previous sections, the radar data and ground truth are prepared for collection. To evaluate the performance of the proposed 3D grid-based method for tracking and activity classification, several experiments are designed. These experiments include tests with different walking trajectories as well as activity classification performed at

various locations in the room. Additionally, to ensure redundancy and background reference data, the first five seconds and the last five seconds of each recording are collected while the person remains static. For each test scenario, two to three rounds of experiments are conducted to ensure result reliability, and the results reported in section 5.3.2 and section 5.4.2 correspond to the most representative round.

Experiment Content of Localization and Tracking

To comprehensively evaluate the localization and tracking performance, four walking trajectories are designed in this thesis, covering different areas of the room: the central area, the left side, the right side, and a round path. The selected trajectories are designed to cover most area of the room. Each trajectory is sufficiently long, avoids returning along the same path (to prevent ambiguity during evaluation), and includes both straight segments and changes in direction. All trajectories start from label point 'I1' in Figure 5.6 because it is close to the door, and in a realistic scenario, a person typically enters the office through the door. Based on these considerations, the four selected routes are as follows:

1. **Central area route:** The route goes from 'I1' to 'D1' and then to 'D6'. This trajectory covers the area from the door to the central region with one change of direction. The central area² is defined as the region inside the triangle formed by the three ceiling-mounted Novelda X4 radars. According to the PDOP analysis in Section 3.6.3, this area has high positioning accuracy. Therefore, this trajectory is used to examine the localization and tracking performance in a good positioning region.
2. **Left side route:** The route goes from 'I1' to 'A1'. This trajectory covers the leftmost part of the room and is designed to evaluate the edge tracking problem in an area outside the central region. It tests performance in a low accuracy region as well as along a straight path. In addition, at the upper-left corner ('A1' and 'B1'), a table with a height of 0.8 m partially blocks the lower torso of the participant from Radar 3, introducing a non line of sight (NLOS) condition. Furthermore, the distance from Radar 2 to 'A1' is relatively long, which provides a weak-signal scenario to further evaluate performance.
3. **Right side route:** The route goes from 'I1' to 'D1', then to 'D6', and finally to 'A6'. This trajectory is similar to the central area route but extends from 'D6' to the rightmost part of the room at 'A6'. The purpose is similar to the left side route, aiming to examine performance in a low-accuracy region, under NLOS conditions, and with weak signals.
4. **Round route:** The route goes from 'I1' to 'D1', then to 'D3', to 'I3', and back to 'I1'. This trajectory forms a loop inside the room, including three direction changes. It is used to test the performance of the tracking algorithm in a more complex path scenario.

Table 5.3 summaries the designed walking trajectories for localization and tracking experiments with path ID.

Table 5.3: Designed walking trajectories for localization and tracking experiments.

Path ID	Route	Label Points	Purpose
P1	Central	I1 → D1 → D6	Central region, one turn
P2	Left side	I1 → A1	Left edge, NLOS, weak signal
P3	Right side	I1 → D1 → D6 → A6	Right edge, NLOS, weak signal
P4	Round	I1 → D1 → D3 → I3 → I1	Loop with three turns

Figure 5.8 summarizes the four designed walking trajectory on the floor plan, where the orange arrows representing

²Label points in central area: 'D3', 'D4', 'D5', 'E3', 'F3', 'G3' (from Figure 5.6)

the designed route.

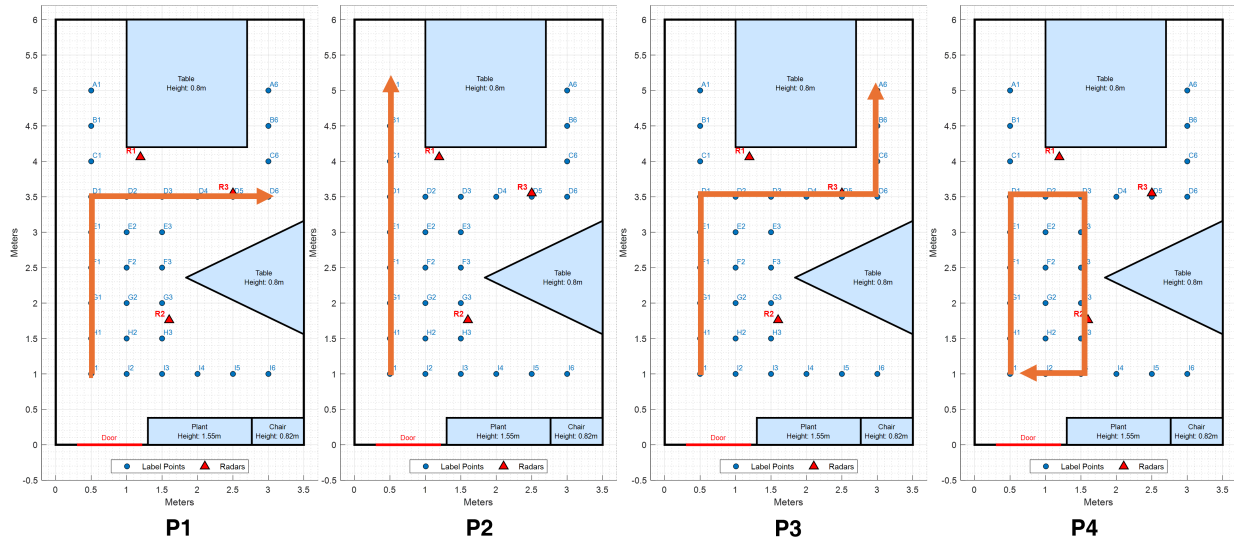


Figure 5.8: Route of four designed walking trajectory on the floor plan.

To fully test the robustness of the proposed method against variations in human height, two participants with different heights are asked to conduct the experiments. One participant is 1.73 m tall (denoted as “H1”), and the other is 1.87 m tall (denoted as “H2”). Due to time limitations, not all designed trajectories are tested by both participants. All trajectories are performed by H1, while only the central route is tested by both H1 and H2.

Based on this setup, the test data ID is labeled as “Test Location”_“Test Module”_“Activity Type”_“Path ID”_“Human ID”. The explanation of the components is provided in Table 5.4. For example, *VH_T_W_P1_H2* indicates a tracking test module with walking activity on Path 1 in the Workplace Vitality Hub conducted by participant two. Similarly, *VH_C_WS_P1_H1* indicates a classification test module with walking and standing activities on Path 1 in the Workplace Vitality Hub conducted by participant one.

Table 5.4: Explanation of the Test Data ID Format.

Label Component	Code	Description	Example	Notes
Test location	VH	VH = Vitality Hub	VH	Test environments
Module type	T / C	T = Tracking C = Classification	T	Defines experiment module
Activity type	W / WS / WSSi	W = walk S = stand Si = sit	WS	Single or combined activities
Path ID	P1 / P2 / .../ Pn	ID of walking path	P3	Different test trajectories
Human ID	H1 / H2 / .../ Hn	Subjects with different height	H2	Participant identifier

In summary, five experiments are conducted to evaluate the localization and tracking performance, labeled as follows:

- *VH_T_W_P1_H1*
- *VH_T_W_P2_H1*

- *VH_T_W_P3_H1*
- *VH_T_W_P4_H1*
- *VH_T_W_P1_H2*

Together, they cover most areas of the room and include tests with different human heights to examine robustness.

Experiment Content of Activity Recognition

In this thesis, three activity types are considered during the trajectory: walking, standing, and sitting. Since the trajectories are generated by human walking, only two additional activities occur at specific locations: standing and sitting. To examine the performance of these two activities at different positions, the experiments focus on two designated points: one in the central area and one in the non-central area. Label point ‘D3’ is selected to represent the central area, and label point ‘F1’ is selected to represent the non-central area.

The central area route is chosen for this experiment because it is a representative path that crosses the central region with a direction change. At both ‘D3’ and ‘F1’, activity transitions between walking and standing, or between walking and sitting, are tested. For the walk-to-stand test, the participant walks to the specified point, stands for approximately five seconds, and then continues the trajectory. For the walk-to-sit test, the participant walks to the specified point, sits on a chair for approximately five seconds, and then stands up to continue the trajectory.

Due to time limitations, experiments that include multiple activity types within one trajectory (for example, walk–stand–sit–stand in sequence) are not conducted. However, separate tests for walk-to-stand and walk-to-sit transitions are sufficient to evaluate the performance of the proposed activity recognition method. To standardize the evaluation in the activity classification experiment module, path ID “P1” denotes activity transitions only at ‘F1’, while path ID “P2” denotes transitions only at ‘D3’. In addition, due to time limitations, only one participant with human height of 1.73 m (“H1”) is tested.

In summary, four experiments are conducted to evaluate the activity recognition performance, covering both standing and sitting activities at the two designated locations ‘D3’ and ‘F1’:

- *VH_C_WS_P1_H1*
- *VH_C_WS_P2_H1*
- *VH_C_WSi_P1_H1*
- *VH_C_WSi_P2_H1*

5.3. Evaluation of Localization and Tracking

Based on the previous test plan, the evaluation of localization and tracking is conducted in this section. The content is organized into several topics including evaluation metrics, experimental results, and related analysis.

5.3.1. Evaluation metric of Localization and Tracking

To evaluate the performance of the estimated trajectory, two main evaluation metrics are employed to quantify the tracking accuracy: the Root Mean Square Error (RMSE) and the Relative Pose Error (RPE).

Root Mean Square Error

The Root Mean Square Error (RMSE) is a metric to evaluate the accuracy of estimated trajectories. It quantifies the average magnitude of the error between the estimated positions and the ground truth positions by taking the

square root of the mean squared differences. Formally, the RMSE is defined in Equation 5.3, where $\hat{\mathbf{p}}_t = [\hat{x}_t, \hat{y}_t]^\top$ denotes the estimated position at time index t , $\mathbf{p}_t = [x_t, y_t]^\top$ is the corresponding ground truth position, and T is the total number of trajectory points. A lower RMSE indicates that the estimated trajectory is closer to the ground truth, thus reflecting better localization and tracking performance.

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T \|\hat{\mathbf{p}}_t - \mathbf{p}_t\|^2}, \quad (5.3)$$

Relative Pose Error

The Relative Pose Error (RPE) is employed to evaluate the local consistency of the estimated trajectory [76]. Instead of comparing absolute positions at each time step, RPE quantifies the accuracy of the relative motion between two trajectory points separated by a fixed spatial distance. For each valid pair of start and end points, the relative displacement vectors from the estimated and ground truth trajectories are computed, and their Euclidean distance is taken as the error. The final RPE is obtained as the root mean square of these segment-wise errors.

Equation 5.4 describes the computation of RPE, where $\hat{\mathbf{p}}_{\tau_s^{(i)}}$ and $\hat{\mathbf{p}}_{\tau_e^{(i)}}$ denote the estimated positions at the start and end indices of the i -th segment, $\mathbf{p}_{\tau_s^{(i)}}$ and $\mathbf{p}_{\tau_e^{(i)}}$ are the corresponding ground truth positions, and I is the number of valid trajectory segments defined by the segment length threshold:

$$\text{RPE} = \sqrt{\frac{1}{I} \sum_{i=1}^I \left\| (\hat{\mathbf{p}}_{\tau_e^{(i)}} - \hat{\mathbf{p}}_{\tau_s^{(i)}}) - (\mathbf{p}_{\tau_e^{(i)}} - \mathbf{p}_{\tau_s^{(i)}}) \right\|^2}. \quad (5.4)$$

In this thesis, the segment length between the start and end indices is set to 10 cm, as too big segment length can not reflect relative motion consistency of the whole trajectory in detail while too small segment length may influence by small fluctuation in the trajectory. The selected segment length meaning that the RPE evaluates the relative motion consistency every 10 cm. A lower RPE indicates that the estimated trajectory better preserves local motion characteristics, ensuring that the movement direction and relative displacement trends are consistent with the ground truth.

5.3.2. Experimental Results of Localization and Tracking

Based on the test plan and the defined evaluation metrics, the performance of the proposed 3D grid-based method is compared with the baseline method across the five experiments described in Section 5.2.3. It should be emphasized that identical parameter settings are applied to both the proposed method and the baseline method (presented in section 3.8) in order to ensure a fair comparison. The plots of different experiment results are simply listed in the following sub-sections, while the value of RMSE, median error, 90-th percentile error and RPE for the different tests are summarized in Table 5.5. A discussion on the results then follows in section 5.3.3.

1. Central Area Route (Test ID: VH_T_W_P1_H1)

Figure 5.9 shows the comparison between the tracking trajectories obtained by the proposed 3D grid-based method and the baseline method for the central area route. The gray line represents the ground truth trajectory. The color variation along the estimated trajectory indicates the distance error relative to the ground truth. In addition, gray dashed lines connect each estimated position with its corresponding ground truth point.

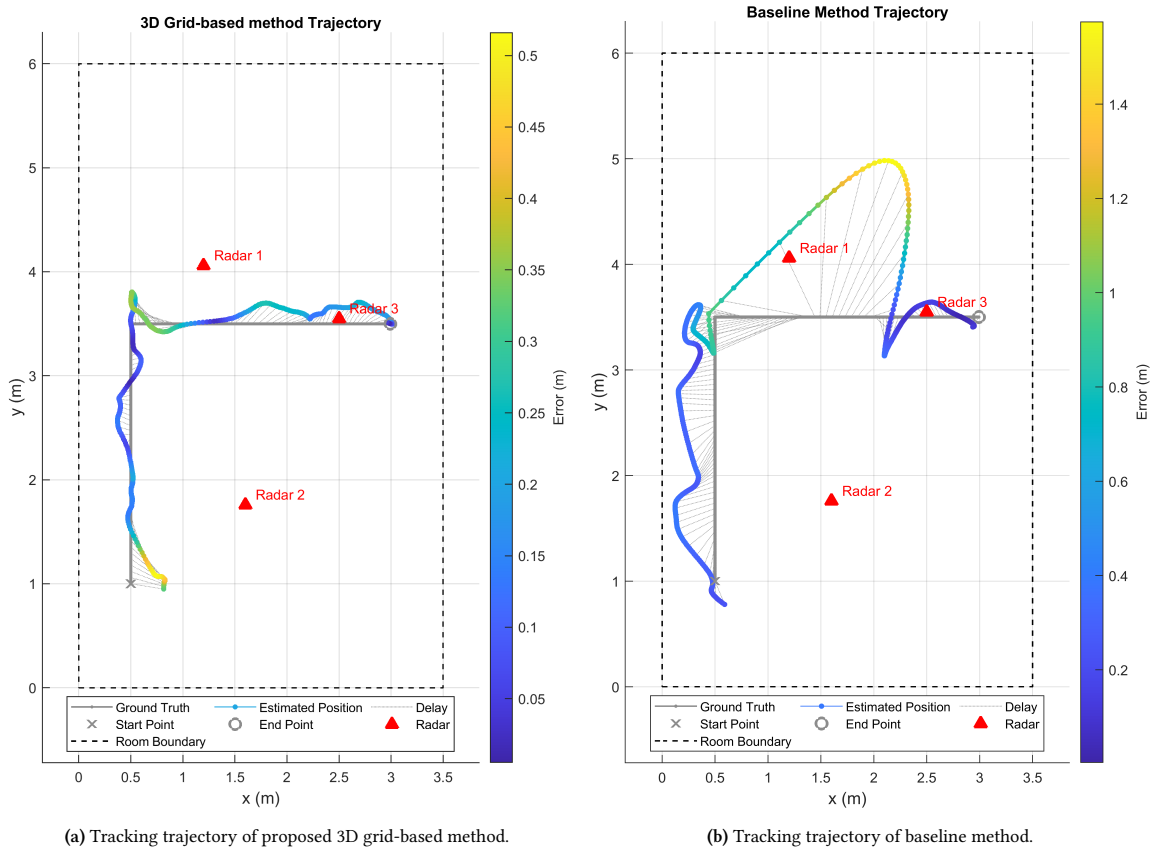


Figure 5.9: Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID *VH_T_W_P1_H1*.

In Figure 5.10, the distance error and cumulative distribution function (CDF) of the proposed method and the baseline method for the central area route are compared. In Figure 5.10, the orange line represents the proposed method and the blue line represents the baseline method.

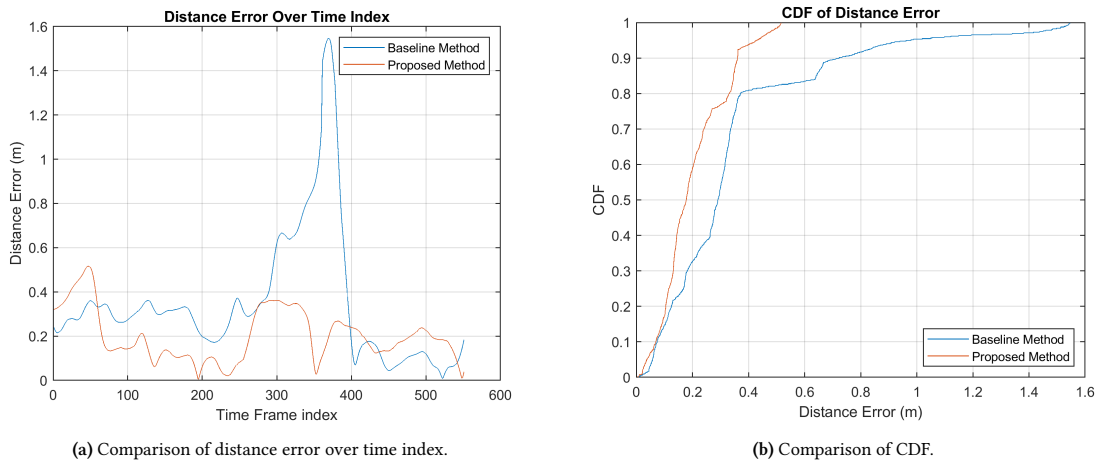


Figure 5.10: Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID *VH_T_W_P1_H1*.

2. Left Side Route (Test ID: *VH_T_W_P2_H1*)

Figure 5.11 shows the comparison between the tracking trajectories obtained by the proposed 3D grid-based method and the baseline method for the left side route.

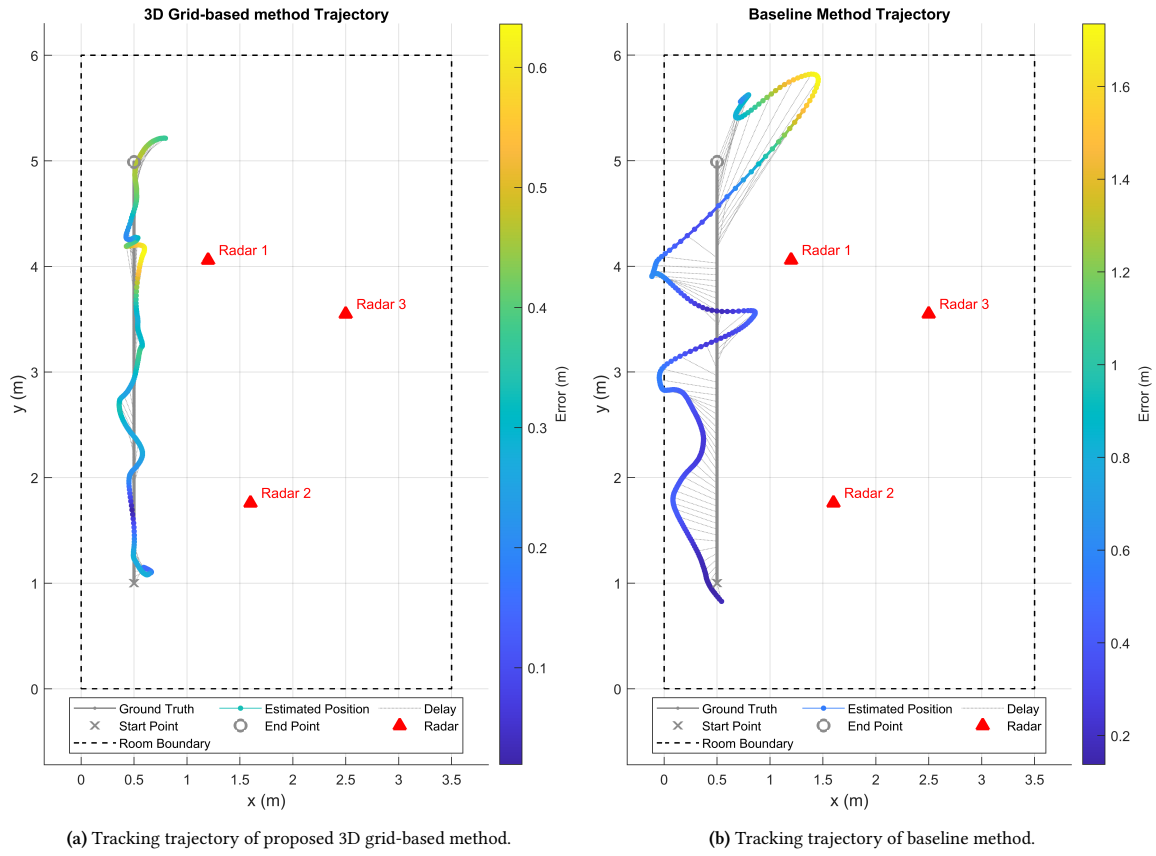


Figure 5.11: Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID $VH_T_W_P2_H1$.

In Figure 5.12, the distance error and CDF of the proposed method and the baseline method for the left side route are compared.

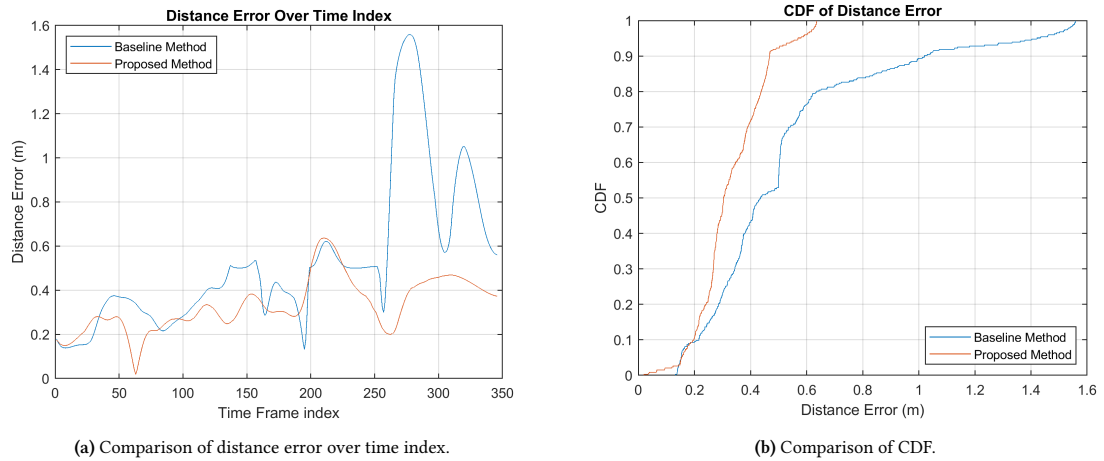


Figure 5.12: Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID $VH_T_W_P2_H1$.

3. Right Side Route (Test ID: $VH_T_W_P3_H1$)

Figure 5.13 shows the comparison between the tracking trajectories obtained by the proposed 3D grid-based method and the baseline method for the right side route.

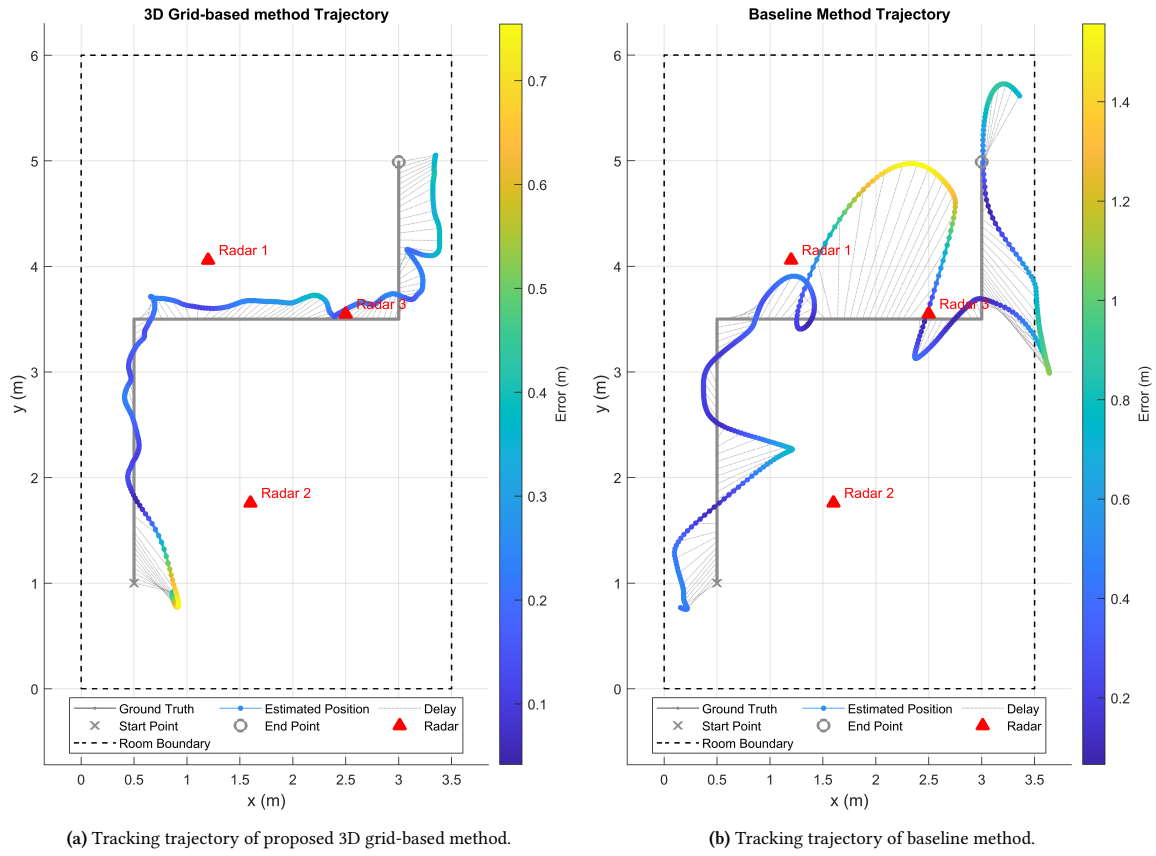


Figure 5.13: Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID *VH_T_W_P3_H1*.

In Figure 5.14, the distance error and CDF of the proposed method and the baseline method for the right side route are compared.

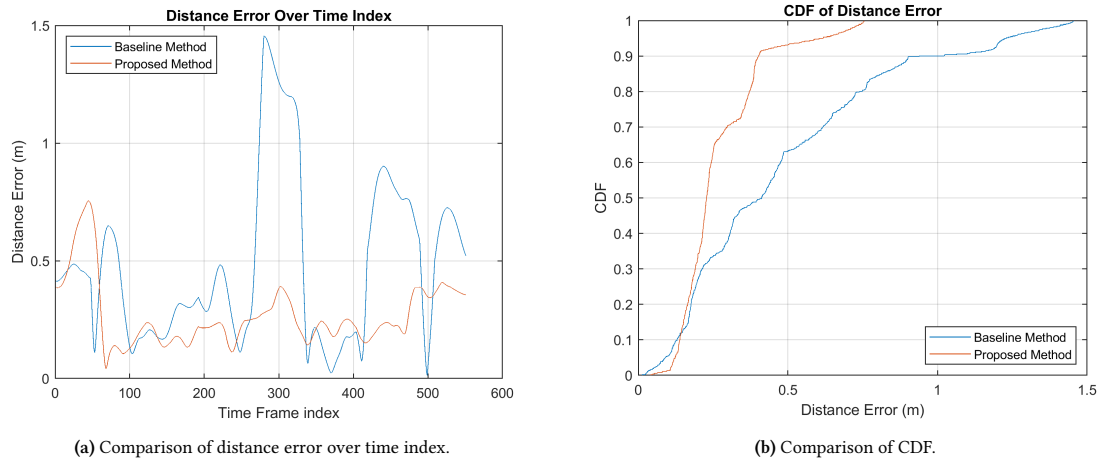


Figure 5.14: Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID *VH_T_W_P3_H1*.

4. Round Route (Test ID: *VH_T_W_P4_H1*)

Figure 5.15 shows the comparison between the tracking trajectories obtained by the proposed 3D grid-based method and the baseline method for the round route.

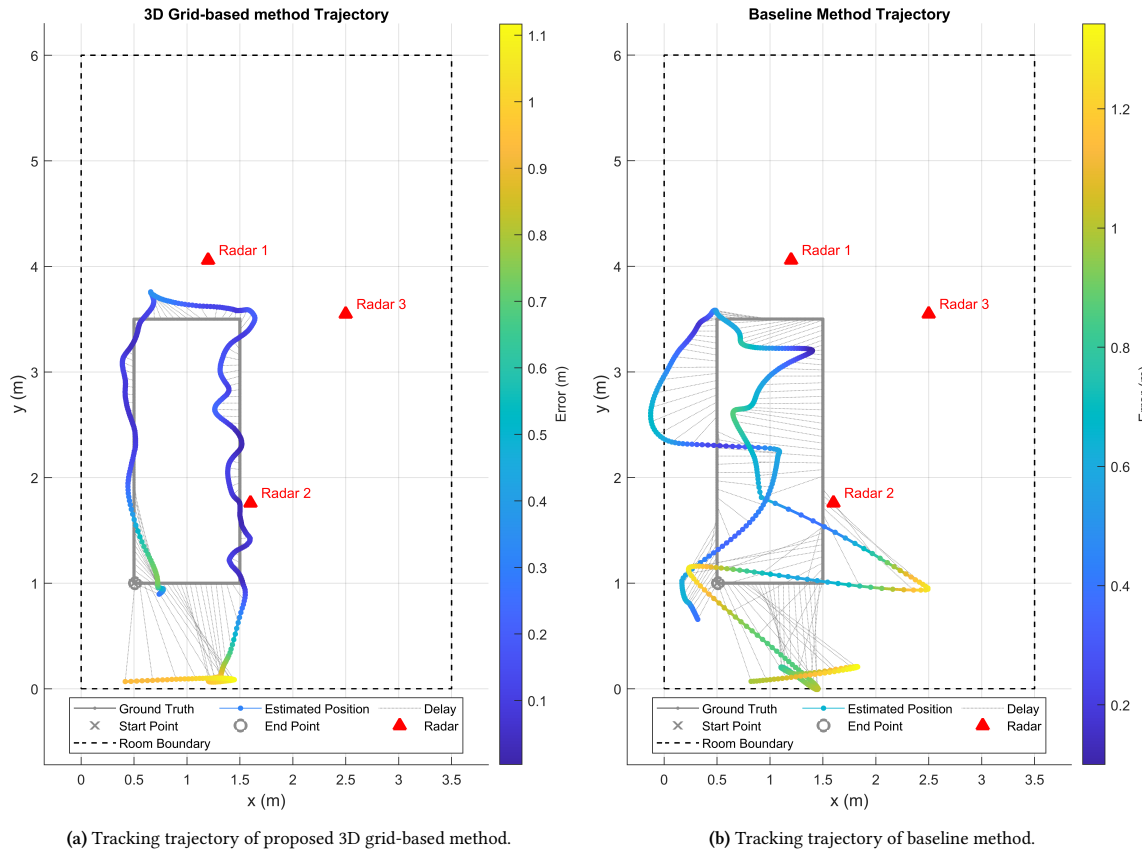


Figure 5.15: Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID *VH_T_W_P4_H1*.

In Figure 5.16, the distance error and CDF of the proposed method and the baseline method for the round route are compared.

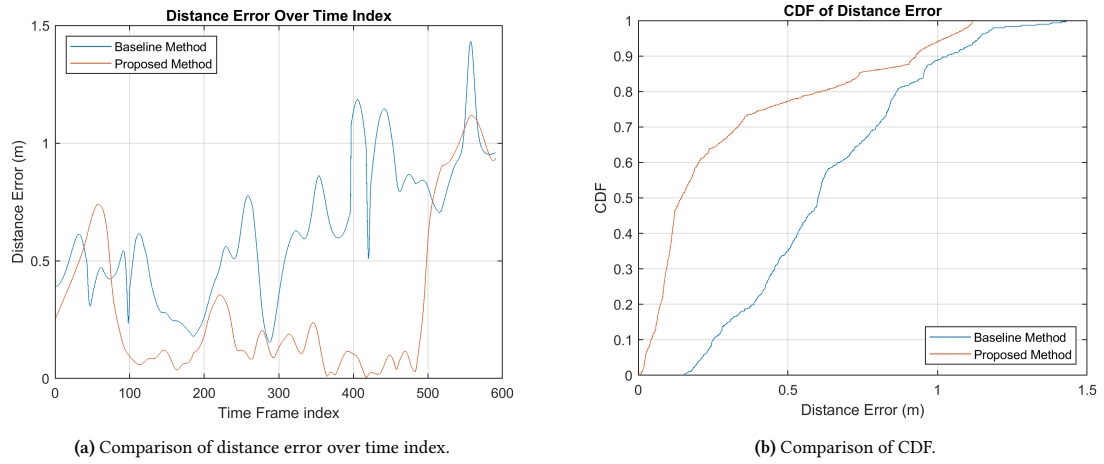


Figure 5.16: Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID *VH_T_W_P4_H1*.

5. Central Area Route with taller participant (Test ID: *VH_T_W_P1_H2*)

Figure 5.17 shows the comparison between the tracking trajectories obtained by the proposed 3D grid-based method and the baseline method for the central area route with the taller participant.

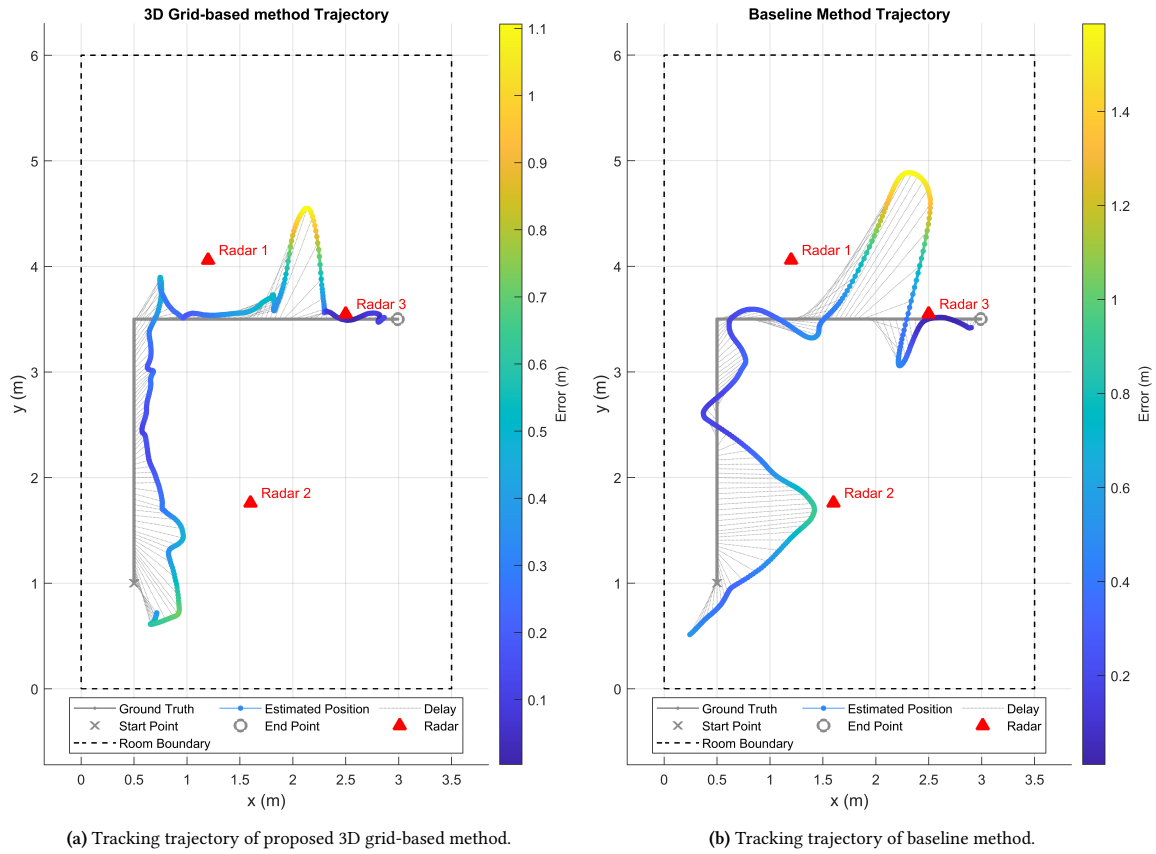


Figure 5.17: Comparison between the tracking trajectory of proposed 3D grid-based method and baseline method for test ID $VH_T_W_P1_H2$.

In Figure 5.18, the distance error and CDF of the proposed method and the baseline method for the central area route with the taller participant are compared.

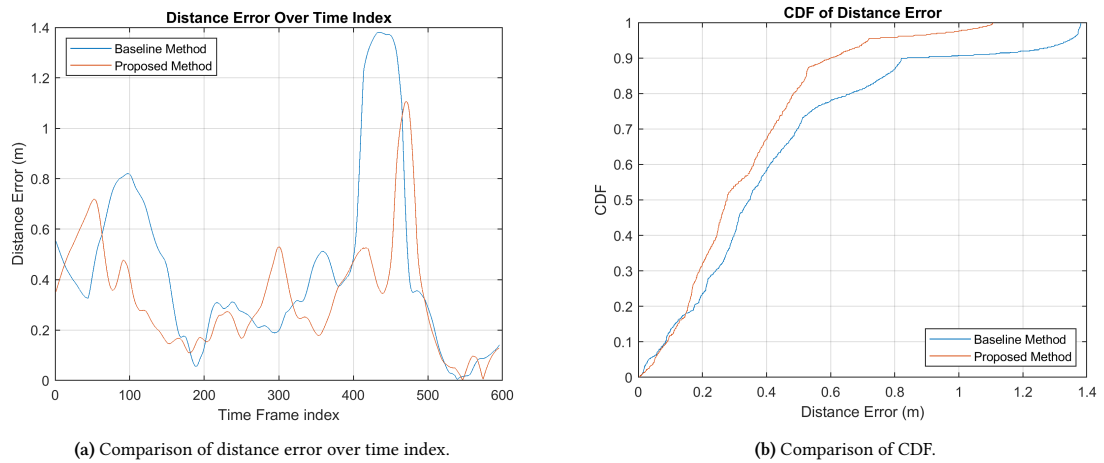


Figure 5.18: Comparison of distance error and CDF between the proposed 3D grid-based method and the baseline method for test ID $VH_T_W_P1_H2$.

Table 5.5: Tracking performance comparison of proposed 3D grid-based method and baseline.

Route & Test ID	Method	RMSE (m)		Median Error (m)		90th Percentile Error (m)		RPE (m)	
		Value	Δ^3	Value	Δ	Value	Δ	Value	Δ
Central area <i>VH_T_W_P1_H1</i>	Proposed	0.2327	+52.6%	0.1775	+38.9%	0.3606	+50.1%	0.0984	+63.3%
	Baseline	0.4913		0.2905		0.7231		0.2678	
Left side <i>VH_T_W_P2_H1</i>	Proposed	0.3522	+46.6%	0.3034	+30.4%	0.4683	+54.1%	0.1055	+63.5%
	Baseline	0.6593		0.4358		1.0201		0.2887	
Right side <i>VH_T_W_P3_H1</i>	Proposed	0.3024	+53.9%	0.2277	+44.8%	0.4017	+58.7%	0.1040	+61.5%
	Baseline	0.6554		0.4124		0.9725		0.2703	
Round <i>VH_T_W_P4_H1</i>	Proposed	0.4478	+37.4%	0.1432	+76.1%	0.9272	+10.2%	0.1356	+53.5%
	Baseline	0.7159		0.6003		1.0320		0.2913	
Central (H2) <i>VH_T_W_P1_H2</i>	Proposed	0.3979	+34.4%	0.2765	+20.4%	0.6003	+28.3%	0.1772	+25.9%
	Baseline	0.6061		0.3472		0.8370		0.2391	

5.3.3. Analysis of Localization and Tracking Results

In the previous sub-section, the detailed results of the five tracking experiments were presented. This sub-section provides a comprehensive analysis of the performance differences between the proposed 3D grid-based method and the baseline approach.

From the numerical results in Table 5.5, it is evident that the proposed method consistently outperforms the baseline across all evaluation metrics, namely RMSE, median error, 90th percentile error, and RPE. Furthermore, the trajectory plots in section 5.3.2 demonstrate that the proposed method produces estimates that are closer to the ground truth with smoother trajectories. The distance error plots and CDF plots in section 5.3.2 also confirm that the proposed method achieves lower distance errors for most time indices. These improvements are substantial rather than marginal, reaching up to 53.9% reduction in RMSE for the right side route and 63.5% reduction in RPE for the left side route, thereby demonstrating both the accuracy and robustness of the proposed approach.

For the central area route (*VH_T_W_P1_H1*), the RMSE decreases from 0.4913 m to 0.2327 m, corresponding to an improvement of 52.6%. Similarly, the median error and 90th percentile error improve by 38.9% and 50.1%, respectively. Notably, the 90th percentile error indicates that in 90% of cases, the distance error of the proposed method remains below 0.36 m. The RPE shows the improvement (63.3%), highlighting that the proposed method not only enhances overall accuracy but also ensures smoother and temporally consistent trajectories. This is further illustrated in Figure 5.9, where the proposed method aligns closely with the ground truth, whereas the baseline suffers from a large outlier (approximately 1.5 m error at around the 350th time index). Despite this, the baseline is still able to capture the general walking trend.

For the left side route (*VH_T_W_P2_H1*), the RMSE and RPE improvements are 46.6% and 63.5%, respectively, confirming the stability of the proposed method even near room edges. As shown in Figure 5.11 and Figure 5.12a, the proposed trajectory initially matches the ground truth with very low error. However, in the end of the trajectory, delay and error increase appear due to NLOS conditions caused by table occlusion of radar 3. Besides, the end of

³ $\Delta = (\text{Baseline} - \text{Proposed}) / \text{Baseline} \times 100\%$

the trajectory is also located in the corner of the room, which corresponds to a limited opening-angle area and introduces a viewpoint variation problem. Even in this adverse scenario, the particle filter effectively follows the motion trend, whereas the baseline trajectory fluctuates more severely.

The right side route (*VH_T_W_P3_H1*) exhibits similar results. The RMSE is reduced by 53.9% and the RPE by 61.5%, while the median and 90th percentile errors improve by 44.8% and 58.7%, respectively. As seen in Figure 5.13, the proposed trajectory generally follows the ground truth except for minor deviations at the start and end caused by NLOS effects and limited opening-angle effect. The baseline, in contrast, produces a large error in the central area, confirming the higher robustness of the proposed approach.

For the round route (*VH_T_W_P4_H1*), improvements are slightly smaller in terms of the 90th percentile error (10.2%), but remain significant for the other metrics: RMSE improves by 37.4%, median error by 76.1%, and RPE by 53.5%. The trajectory plot in Figure 5.15 shows that while the baseline method fails to capture the complex route, the proposed method achieves a reasonable trajectory with an RMSE of 0.4478 m, except for a large error segment between positions 'I3' and 'I1', where radar 1 and 2 are too far from the target.

Finally, for the central route with a taller subject height (*VH_T_W_P1_H2*), the proposed method maintains consistent improvements, though with smaller margins compared to the 1.73 m subject. The RMSE and RPE improve by 34.4% and 25.9%, respectively. Only a single small outlier point is observed in the proposed trajectory, whereas the baseline suffers from larger deviations, demonstrating the adaptability of the method to variations in subject height.

The poorer performance of the baseline can be attributed to the limited number of radars (three) and the reliance on Taylor localization in 3D space. Since the head position is unknown, any small range estimation error at a single radar can make the localization matrix nearly singular, leading to severe inaccuracies. In contrast, the proposed method uses the extended target characteristics of walking humans by estimating the joint distribution of reflections along vertical columns, effectively increasing redundancy and mitigating the influence of small errors.

Across all tests, the best RMSE for the proposed method is 0.2327 m (central area, subject height 1.73 m), while the worst case is 0.4478 m in the round route with multiple turns. In most scenarios, the RMSE remains around 0.3-0.4 m, the median error is below 0.3 m, the 90th percentile error is below 0.6 m, and the RPE is approximately 0.1-0.15 m. These results confirm that the proposed method achieves both high accuracy and robustness under diverse test routes and subject conditions.

Compared with existing studies on UWB radar tracking, most prior works rely on wall-mounted UWB radars based on Decawave or Novelda chipsets, with reported average accuracies generally ranging from 0.25 to 0.6 m [7]. In contrast, the proposed method employs a ceiling-mounted multistatic configuration, which is less explored in the literature yet more suitable for office-like environments. Methodologically, a 3D grid-based method is introduced, following a TBD framework. As a result, the system achieves an RMSE of 0.23–0.45 m, which is competitive with state-of-the-art methods while using a limited number of radars.

Overall, the proposed 3D grid-based method provides consistent improvements over the baseline in both average-case errors (RMSE, median) and extreme-case errors (90th percentile), as well as in temporal consistency (RPE). This demonstrates its robustness and reliability for practical indoor human tracking applications by using the ceiling-mounted UWB radars and the proposed method.

5.4. Evaluation of Activity Recognition

Similar to the evaluation of localization and tracking, the content of this section is also organized into several topics including evaluation metrics, experimental results, and related analysis.

5.4.1. Evaluation Metrics

Since activity transitions occur along the trajectory, the trajectory is also estimated during activity recognition. Therefore, the previously defined tracking metrics (RMSE, median error, 90th percentile error, and RPE) are applied to quantify localization accuracy.

For activity classification, performance is then evaluated using the confusion matrix, where the counts of correct and incorrect decisions are summarized. In this context, “+” is used to denote the positive class (e.g., *walking*) and “−” is used to denote the negative class (e.g., *static*). The matrix is defined as follow:

- **TP** (True Positive): an actual “+” is predicted as “+”.
- **TN** (True Negative): an actual “−” is predicted as “−”.
- **FP** (False Positive): an actual “−” is predicted as “+”.
- **FN** (False Negative): an actual “+” is predicted as “−”.

Table 5.6: Binary confusion matrix (rows: actual, columns: predicted).

	Pred: +	Pred: −
Actual: +	TP	FN
Actual: −	FP	TN

From this matrix, the accuracy used in the evaluation are computed in Equation 5.5.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.5)$$

By combining these classification metrics with the tracking metrics, a complete evaluation is obtained that reflects both the spatial accuracy of the estimated trajectory and the correctness of the recognized activities along that trajectory.

5.4.2. Experimental Results of Activity Recognition

Based on the test plan and the defined evaluation metrics, the activity recognition performance of the proposed 3D grid-based method is evaluated across the four experiments described in Section 5.2.3. A detailed summary of the classification results is presented in Table 5.7, with a more detailed discussion in the later section 5.4.3. The key metrics are discussed in the previous sub-section, while the transition distance is used to describe the difference between the estimated activity transition point and the ground truth transition point.

1. Walk-Stand test at ‘F1’ (Test ID: VH_C_WS_P1_H1)

Figure 5.19 shows the trajectory of the walk–stand test at label point ‘F1’. This plot is similar to the previous trajectory plots, with the only difference being the addition of markers that indicate the standing position. The blue asterisk denotes the ground truth location where the standing activity occurs, while the magenta asterisk denotes the estimated position of the standing activity occurs.

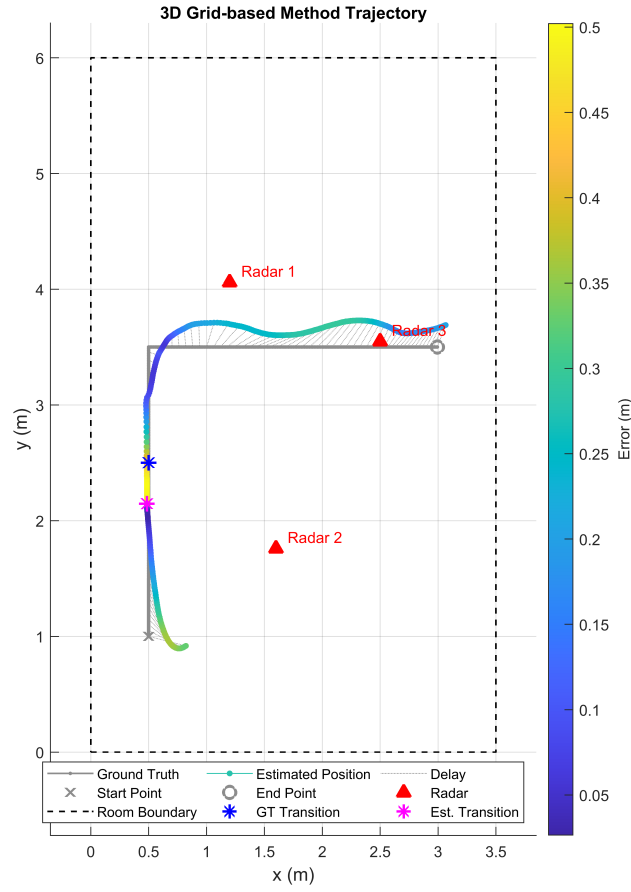


Figure 5.19: Tracking trajectory of Walk-Stand test at 'F1'.

In Figure 5.20, the distance error and CDF plot of the proposed method for Walk-Stand test at 'F1' are illustrated.

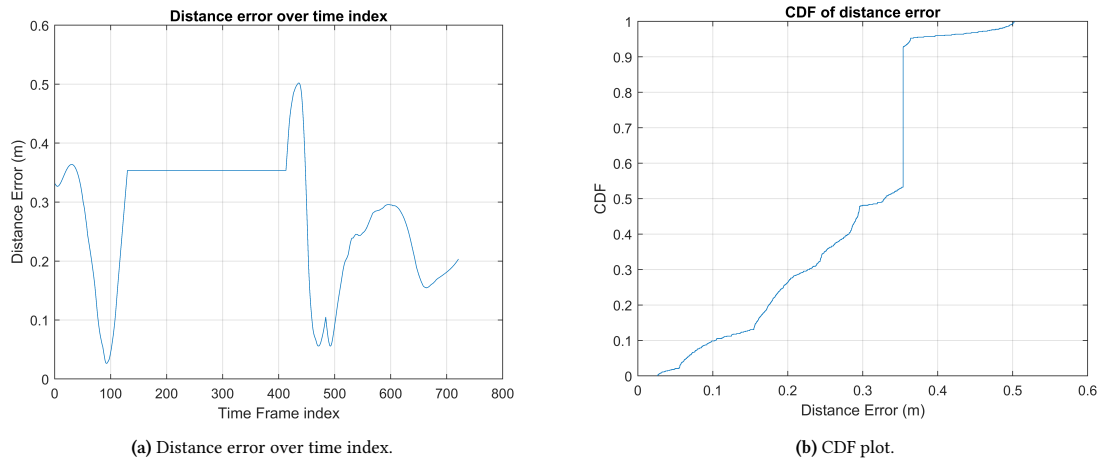


Figure 5.20: Distance error and CDF of the proposed 3D grid-based method for Walk-Stand test at 'F1'.

Figure 5.21 shows the classification results of this experiment. The classification result over the time index illustrates the predicted activity at each timestamp, while the confusion matrix presents the overall classification accuracy.

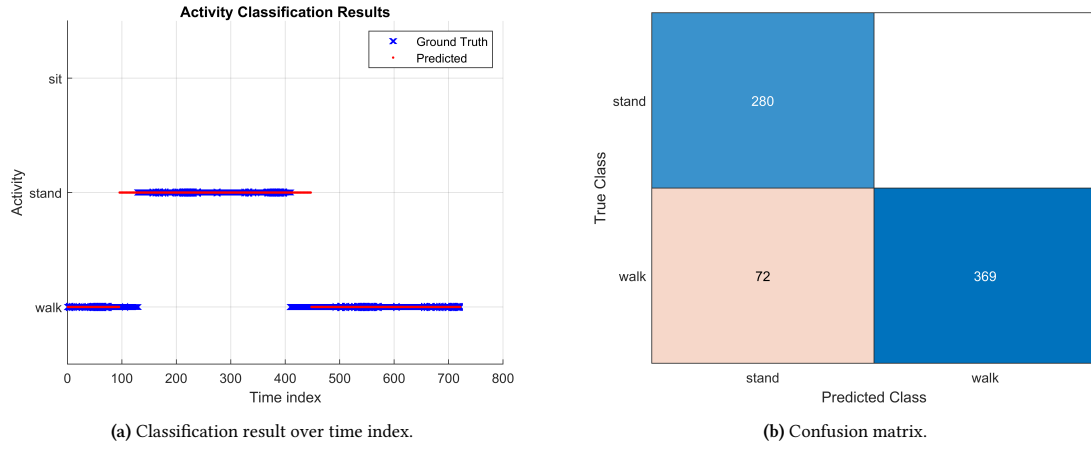


Figure 5.21: Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Stand test at 'F1'.

2. Walk-Stand test at 'D3' (Test ID: VH_C_WS_P2_H1)

Figure 5.22 shows the trajectory of the walk-stand test at label point 'D3'.

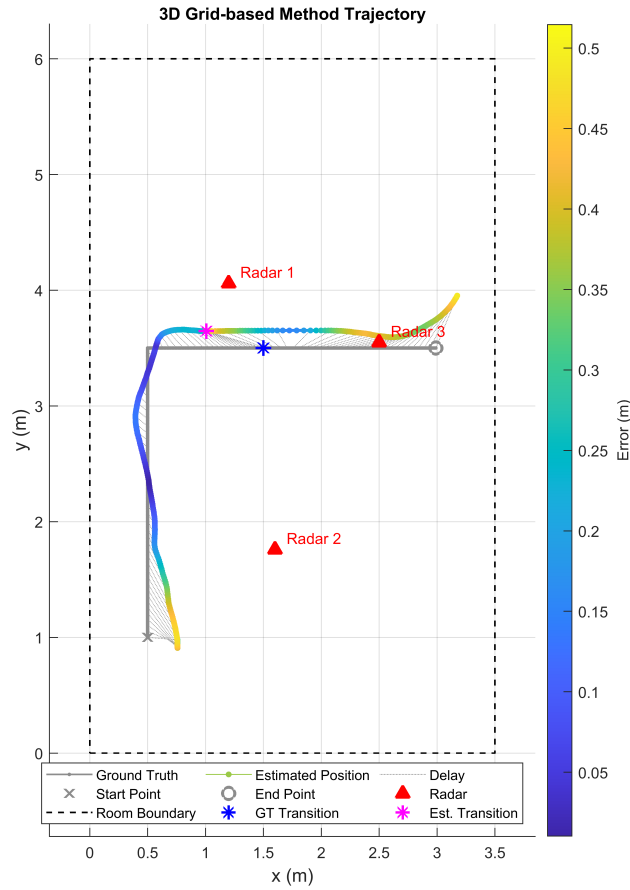


Figure 5.22: Tracking trajectory of Walk-Stand test at 'D3'.

In Figure 5.23, the distance error and CDF plot of the proposed method for Walk-Stand test at 'D3' are illustrated.

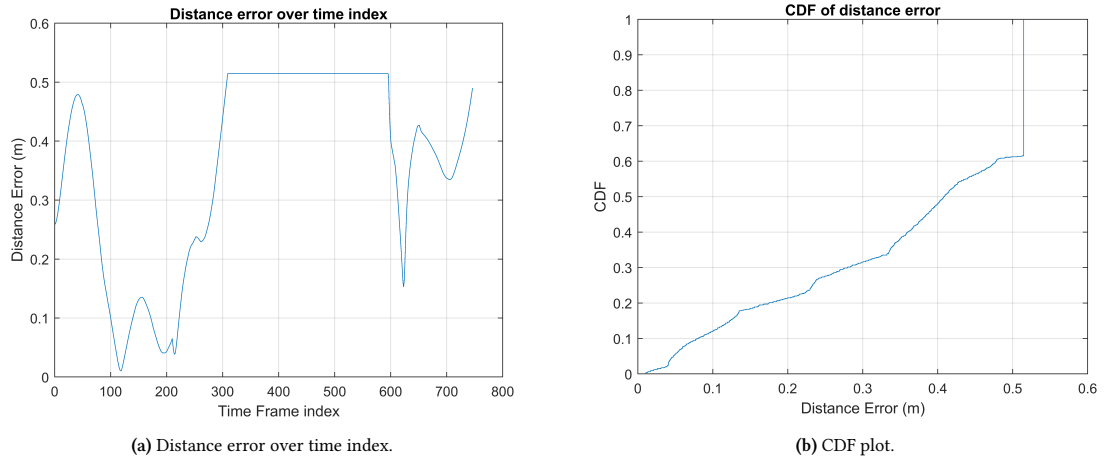


Figure 5.23: Distance error and CDF of the proposed 3D grid-based method for Walk-Stand test at 'D3'.

Figure 5.24 shows the classification results of this experiment. The classification result over the time index illustrates the predicted activity at each timestamp, while the confusion matrix presents the overall classification accuracy.

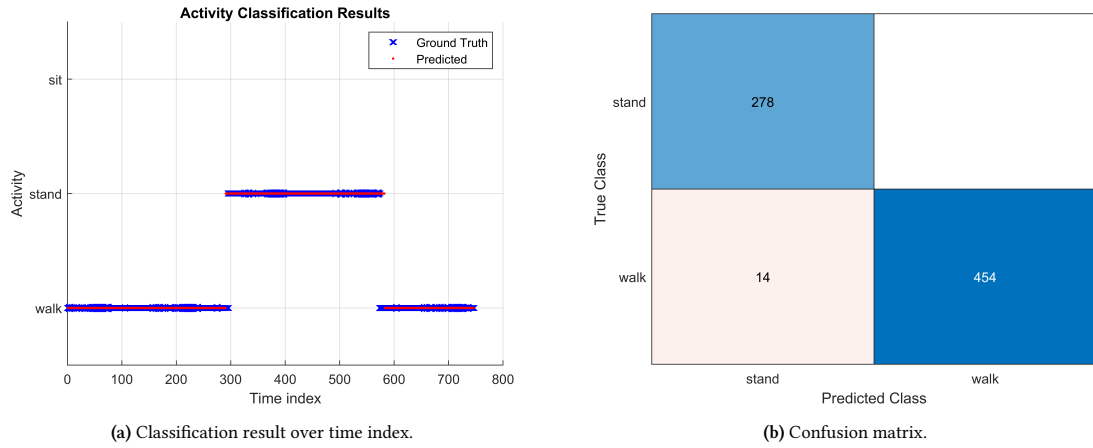


Figure 5.24: Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Stand test at 'D3'.

3. Walk-Sit test at 'F1' (Test ID: VH_C_WSi_P1_H1)

In Figure 5.25, the distance error and CDF plot of the proposed method for Walk-Sit test at 'F1' are illustrated.

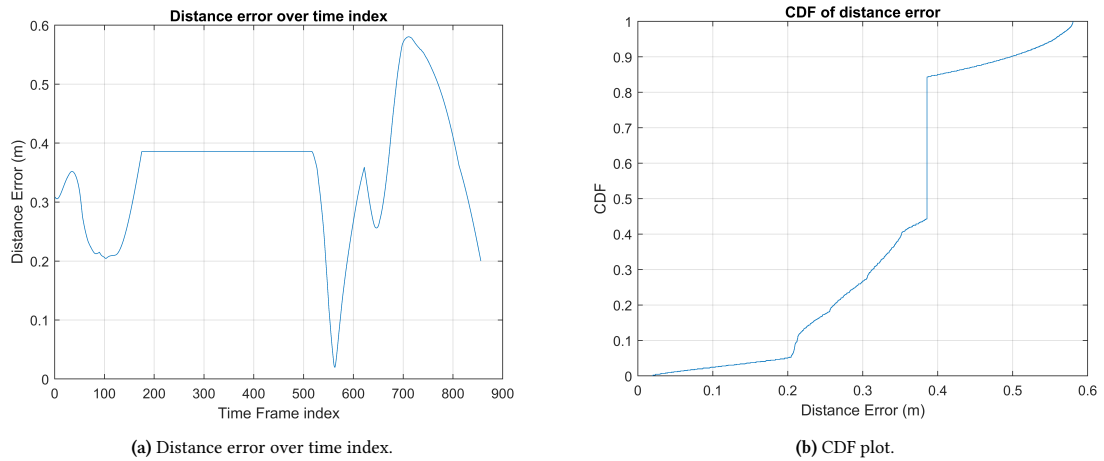


Figure 5.25: Distance error and CDF of the proposed 3D grid-based method for Walk-Sit test at 'F1'.

Figure 5.26 shows the trajectory of the walk-sit test at label point ‘F1’.

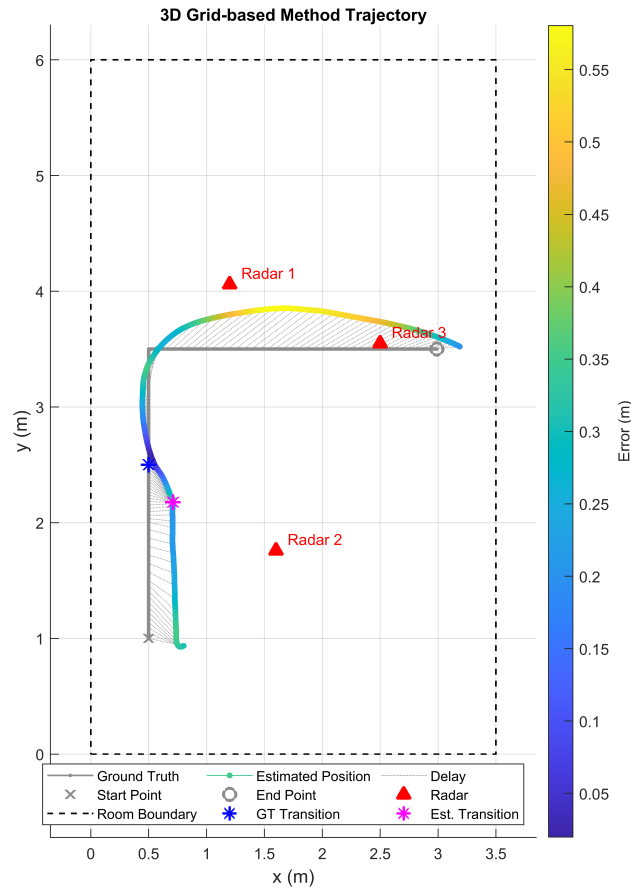


Figure 5.26: Tracking trajectory of Walk-Sit test at ‘F1’.

Figure 5.27 shows the classification results of this experiment. The classification result over the time index illustrates the predicted activity at each timestamp, while the confusion matrix presents the overall classification accuracy.

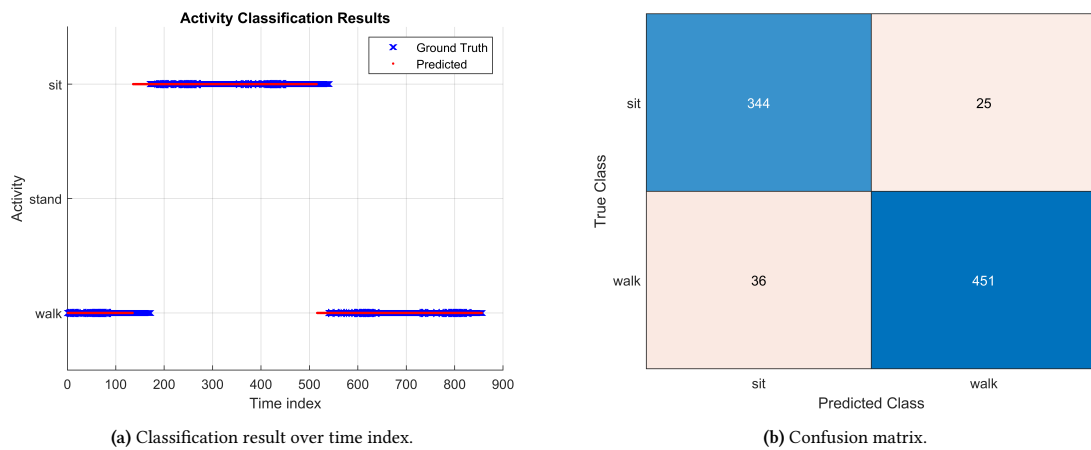


Figure 5.27: Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Sit test at ‘F1’.

4. Walk-Sit test at ‘D3’ (Test ID: VH_C_WSi_P2_H1)

Figure 5.28 shows the trajectory of the walk-sit test at label point ‘D3’.

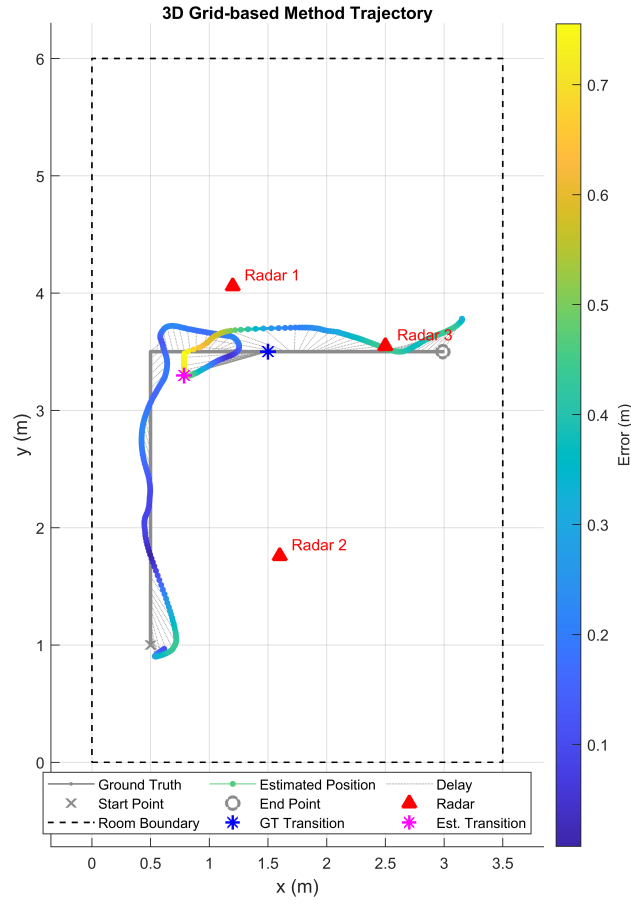


Figure 5.28: Tracking trajectory of Walk-Sit test at 'D3'.

In Figure 5.29, the distance error and CDF plot of the proposed method for Walk-Sit test at 'D3' are illustrated.

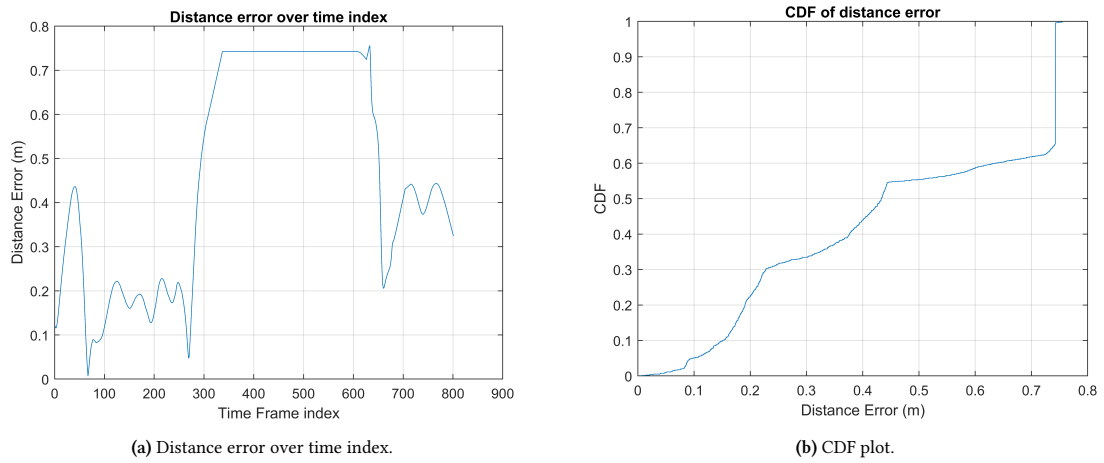


Figure 5.29: Distance error and CDF of the proposed 3D grid-based method for Walk-Sit test at 'D3'.

Figure 5.30 shows the classification results of this experiment. The classification result over the time index illustrates the predicted activity at each timestamp, while the confusion matrix presents the overall classification accuracy.

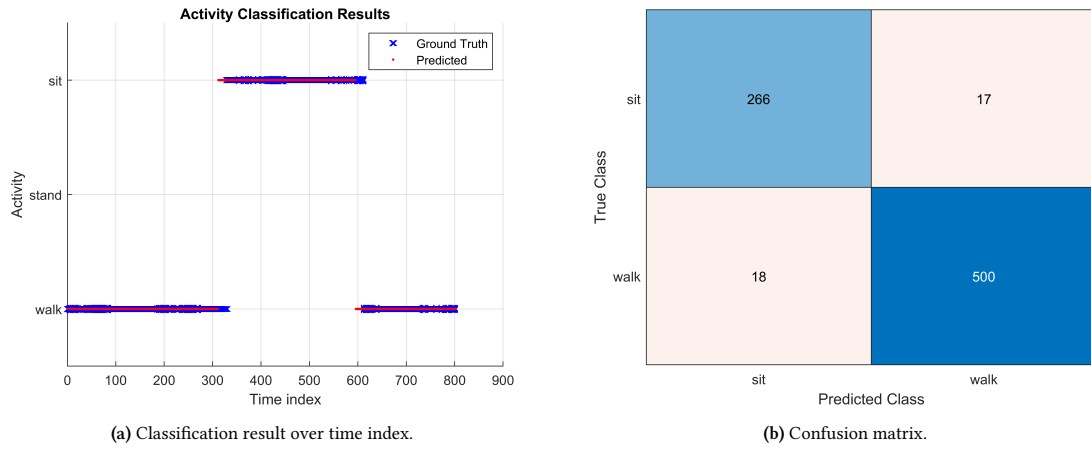


Figure 5.30: Classification result over time index and confusion matrix of the proposed 3D grid-based method for Walk-Sit test at 'D3'.

Table 5.7: Summary of activity recognition results.

Test & Test ID	Accuracy	RMSE (m)	RPE (m)	Median (m)	90th Percent. (m)	Transition Dist. (m)
Walk-Stand test at 'F1' <i>VH_C_WS_P1_H1</i>	0.90	0.2974	0.0808	0.3292	0.3538	0.3538
Walk-Stand test at 'D3' <i>VH_C_WS_P2_H1</i>	0.98	0.3976	0.1473	0.4084	0.5147	0.5147
Walk-Sit test at 'F1' <i>VH_C_WSi_P1_H1</i>	0.93	0.3673	0.2807	0.3857	0.4972	0.3857
Walk-Sit test at 'D3' <i>VH_C_WSi_P2_H1</i>	0.96	0.5285	0.2484	0.4328	0.7425	0.7425

5.4.3. Analysis of Activity Recognition Results

From the numerical activity recognition results in Table 5.7, it is observed that the classification accuracy is consistently above 90%, while most of the related trajectory RMSE values remain below 0.4, m. When the activity transition distance is larger, the corresponding test RMSE also increases. This is because the RMSE is computed along the entire trajectory, including the static periods of standing and sitting, where the accumulated transition distance error contributes to a higher RMSE. This also explains why identical error values are observed for a certain duration in the distance error plots in Section 5.4.2. Moreover, the trajectory plots in Section 5.4.2 demonstrate that smooth trajectories are still obtained when the target remains static.

According to the activity results presented in Table 5.7, as well as the classification plots over time and the confusion matrices in Section 5.4.2, the proposed method achieves reliable classification of walk-stand and walk-sit activities, with only a few misclassifications occurring during activity transitions. These errors are mainly caused by inaccurate assumptions of the transition time and the inaccuracy in ground truth that is measured by the subjects themselves. Besides, although the activity classification in both the central area ('D3') and the non-central area ('F1') achieves an accuracy above 90%, the performance in the central area ('D3') is still slightly better than that in the non-central area. However, since the proposed activity recognition method is threshold-based and relies on human height, its robustness for participants with different body heights still needs to be validated.

5.5. Summary

In this chapter, the radar used for testing and the experimental environment are first introduced. Then, a detailed test plan is designed to perform various scenario based experiments for tracking and activity recognition, in order to evaluate the robustness and accuracy of the proposed method. Finally, the results of tracking and activity recognition are presented, achieving a best RMSE of 0.23, m and a best classification accuracy of 98%.

Conclusion and Future Work

This chapter summarizes the main conclusion of this thesis and discuss the potential improvement of this thesis

6.1. Conclusion

This thesis successfully presented a novel indoor human tracking and activity recognition system utilizing multiple ceiling-mounted ultra-wideband (UWB) radars. By mounting the UWB radar on the ceiling, the likelihood of occlusion by people or furniture is reduced. The core contribution of this work is the development of a proposed 3D grid-based joint tracking and classification method for a single person. A key innovation of this method is its ability to fully leverage the extended target characteristics of a walking human, thereby significantly improving localization performance. The method works by mapping the Channel Impulse Response (CIR) to a 3D grid, and a subsequent compensation procedure helps to generate an accurate 2D likelihood map that represents the presence probability of a walking person. A particle filter is then used for trajectory tracking, while a parallel estimation of the person's height aids in the classification of activities.

The proposed methodology was validated in a real-world office environment using three Novelda X4 UWB radars, demonstrating significant performance improvements in the limited radar number scenario. Experimental results show the superior performance of this method in both localization and tracking when compared to the baseline approach. Notably, in the central area of the room, the root mean square error (RMSE) was reduced from 0.4913 meters to 0.2327 meters, a 52.6% improvement. The relative position error (RPE) also saw a substantial improvement of 63.3%, which indicates that the generated trajectories are smoother and more temporally consistent. Furthermore, the stability and robustness of the method were confirmed even in challenging non-line-of-sight (NLOS) conditions and areas with weak signals and limited opening angle issues with viewpoint variation problem, particularly at the room's edges. In terms of activity recognition, the method achieved a high classification accuracy of approximately 90% for walking, standing, and sitting activities, and this performance was maintained in both central and non-central areas along the walking trajectory.

In summary, this thesis validates the feasibility and effectiveness of using limited number of ceiling-mounted UWB radars for robust human tracking and activity recognition in smart office environments. By proposing a novel method that uses extended target characteristics and provides a reliable, privacy-preserving solution, this work lays a solid foundation for future research and practical applications in the field of indoor monitoring and analysis.

6.2. Future Work

Although the proposed method demonstrates accuracy and robustness, several directions remain open for further research and improvement. These directions can be broadly categorized into validation and robustness, computational efficiency, and methodological extensions.

- **Validation and Robustness:** To strengthen the generalizability and reliability of the proposed approach, additional experiments are necessary:
 - **Test with multi-activity sequences:** Future datasets should include sequences with multiple activity transitions (e.g., walk $\rightarrow$ stand $\rightarrow$ sit $\rightarrow$ stand). Such scenarios are closer to real-world human behavior and will allow for a more comprehensive evaluation of the activity classification module.
 - **Test with more participant diversity:** The current experiments are limited in terms of body height variation. Including participants with a broader range of body heights will enable more rigorous validation of robustness and improve applicability to diverse populations.
 - **Test with more radar configurations:** This thesis employs only three ceiling-mounted radars. Experiments with additional radars, or alternative deployment strategies (e.g., hybrid wall/ceiling placements), are needed to determine whether the proposed method consistently outperforms baseline approaches under varied setups.
 - **Test with better ground truth measurement:** The current ground truth relies on timer-based measurement. Replacing this with motion capture systems or wearable sensors would provide higher temporal and spatial accuracy, thereby enabling more reliable evaluation of tracking and classification performance.
- **Computational Efficiency:** Profiling results indicate that the dominant bottleneck lies in the construction of the 3D CIR grid. To move towards real-time deployment, several optimization strategies are possible:
 - **Construct CIR grid based on radial line value:** Instead of constructing a dense voxel grid, only the values along the radial line in each layer can be generated, capturing the essential information while avoiding irrelevant voxel computation.
 - **Joint grid construction with particle filter:** Grid construction can be directly coupled with the particle filter: rather than computing the entire grid, only the vertical voxel columns at hypothesized particle locations are constructed.
- **Future Extensions:** Beyond single-person scenarios, extending the method to multi-person tracking represents a key research direction:
 - **Multi-person tracking:** Leveraging the extended-target modeling introduced in this thesis may improve data association across multiple reflections, thus enhancing trajectory estimation in crowded or multi-target environments.

References

- [1] Ryan Peppin. *What Is Smart Office Technology? Tools, Features, & Benefits*. May 2025. URL: <https://spaces.cisco.com/what-is-smart-office-technology/>.
- [2] Irfanullah Khan et al. "Occupancy prediction in iot-enabled smart buildings: Technologies, methods, and future directions". In: *Sensors* 24.11 (2024), p. 3276.
- [3] Jonathan Synnott, Joseph Rafferty, and Chris D Nugent. "Detection of workplace sedentary behavior using thermal sensors". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, pp. 5413–5416.
- [4] Huthaifa Obeidat et al. "A review of indoor localization techniques and wireless technologies". In: *Wireless Personal Communications* 119.1 (2021), pp. 289–327.
- [5] Faheem Zafari, Athanasios Gkelias, and Kin K Leung. "A survey of indoor localization systems and technologies". In: *IEEE Communications Surveys & Tutorials* 21.3 (2019), pp. 2568–2599.
- [6] Utku Kumbul et al. "Indoor Localization with Passive RFID System Based on Phase Difference of Displacement". In: *2025 19th European Conference on Antennas and Propagation (EuCAP)*. 2025, pp. 1–5. DOI: 10.23919/EuCAP63536.2025.10999484.
- [7] Mohammad Cheraghinia et al. "A comprehensive overview on UWB radar: Applications, standards, signal processing techniques, datasets, radio chips, trends and future research directions". In: *IEEE Communications Surveys & Tutorials* (2024).
- [8] Y. Kilic et al. "Device-free person detection and ranging in UWB networks". Undefined. In: *IEEE journal of selected topics in signal processing* 8.1 (Feb. 2014). eemcs-eprint-24025, pp. 43–54. ISSN: 1932-4553. DOI: 10.1109/JSTSP.2013.2281780.
- [9] Anton Ledergerber and Raffaello D'Andrea. "A Multi-Static Radar Network with Ultra-Wideband Radio-Equipped Devices". In: *Sensors* 20.6 (2020). ISSN: 1424-8220. DOI: 10.3390/s20061599. URL: <https://www.mdpi.com/1424-8220/20/6/1599>.
- [10] B. Yan, A. Giorgetti, and E. Paolini. *A Track-Before-Detect Algorithm for UWB Radar Sensor Networks*. 2021. arXiv: 2108.00501 [eess.SP]. URL: <https://arxiv.org/abs/2108.00501>.
- [11] Yuan He, Timofey Savelyev, and Alexander Yarovoy. "Two-stage algorithm for extended target tracking by multistatic UWB radar". In: *Proceedings of 2011 IEEE CIE International Conference on Radar*. Vol. 1. IEEE. 2011, pp. 795–799.
- [12] Wei Lu et al. "An unobtrusive fall detection system using ceiling-mounted ultra-wideband radar". In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2023, pp. 1–5.
- [13] *Home | imec the Netherlands*. URL: <https://www.imec-int.com/en/the-netherlands>.
- [14] *Data driven office environment*. URL: <https://workplacevitalityhub.nl/innovatie/data-driven/>.
- [15] Jana Fortes et al. "Positioning and Tracking of Multiple Humans Moving in Small Rooms Based on a One-Transmitter-Two-Receiver UWB Radar Configuration". In: *Sensors* 22.14 (2022). ISSN: 1424-8220. DOI: 10.3390/s22145228. URL: <https://www.mdpi.com/1424-8220/22/14/5228>.

- [16] Mária Švecová et al. "Through-the-floor localization of a static person by a multistatic UWB radar". In: *Microwave and Optical Technology Letters* 61.3 (2019), pp. 825–831.
- [17] Amirashkan Farsaei et al. "Gesture Detection Using IEEE 802.15.4z-Compliant Impulse Radio UWB Radar System". In: *2024 IEEE 4th International Symposium on Joint Communications & Sensing (JC&S)*. 2024, pp. 1–4. doi: 10.1109/JCS61227.2024.10646221.
- [18] Meiqiu Jiang et al. "A robust target tracking method for crowded indoor environments using mmWave radar". In: *Remote Sensing* 15.9 (2023), p. 2425.
- [19] M.I. Skolnik. *Introduction to Radar Systems*. Electrical engineering series. McGraw-Hill, 2001. ISBN: 9780071181891. URL: <https://books.google.nl/books?id=Y6-APwAACAAJ>.
- [20] Novelda AS. *X4A02 Datasheet: X4 Antenna Board*. B Preliminary - October 31, 2017. Oct. 2017.
- [21] Novelda AS. *X7F202 Datasheet: Multichannel Ultra Wideband (UWB) Impulse Radar Sensor*. Version Rev. 3.1 - Advance Information. Published: 2024-10-21. Novelda AS. Oct. 2024.
- [22] Berk Gulmezoglu, Mehmet Burak Guldogan, and Sinan Gezici. "Multiperson tracking with a network of ultra-wideband radar sensors based on Gaussian mixture PHD filters". In: *IEEE Sensors Journal* 15.4 (2014), pp. 2227–2237.
- [23] Van-Han Nguyen and Jae-Young Pyun. "Location Detection and Tracking of Moving Targets by a 2D IR-UWB Radar System". In: *Sensors* 15.3 (2015), pp. 6740–6762. ISSN: 1424-8220. DOI: 10.3390/s150306740. URL: <https://www.mdpi.com/1424-8220/15/3/6740>.
- [24] Sungwon Yoo et al. "A Multiple Target Positioning and Tracking System Behind Brick-Concrete Walls Using Multiple Monostatic IR-UWB Radars". In: *Sensors* 19.18 (2019). ISSN: 1424-8220. DOI: 10.3390/s19184033. URL: <https://www.mdpi.com/1424-8220/19/18/4033>.
- [25] Federal Communications Commission et al. "Revision of part 15 of the commission's rules regarding ultra-wideband transmission systems". In: *FIRST REPORT AND ORDER FCC 02-48* (2002).
- [26] Marko Malajner, Danijel Šipoš, and Dušan Gleich. "Design of a low-cost ultra-wide-band radar platform". In: *Sensors* 20.10 (2020), p. 2867.
- [27] Jonas Ninnemann, Paul Schwarzbach, and Oliver Michler. "Toward UWB Impulse Radio Sensing: Fundamentals, Potentials, and Challenges". In: *UWB Technology*. Ed. by Rafael Vargas-Bernal. Rijeka: IntechOpen, 2023. Chap. 5. doi: 10.5772/intechopen.110040. URL: <https://doi.org/10.5772/intechopen.110040>.
- [28] Yakup Kilic. *Device-free detection and localization of people using uwb networks*. University of Twente, 2016.
- [29] Anton Ledergerber and Raffaello D'Andrea. "A multi-static radar network with ultra-wideband radio-equipped devices". In: *Sensors* 20.6 (2020), p. 1599.
- [30] Josef Kulmer et al. "Using DecaWave UWB transceivers for high-accuracy multipath-assisted indoor positioning". In: *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE. 2017, pp. 1239–1245.
- [31] Ben Van Herbruggen et al. "Impact of CIR processing for UWB radar distance estimation with the DW1000 transceiver". In: *2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE. 2023, pp. 1–7.
- [32] "IEEE Standard Radar Definitions". In: *IEEE Std 686-2008 (Revision of IEEE Std 686-1997)* (2008), pp. 1–44. doi: 10.1109/IEEESTD.2008.4530766.
- [33] Leander Humbert, Tim Fountain, and Rohde & Schwarz. *AN INTRODUCTION TO PASSIVE RADAR SYSTEMS*. 2021. URL: https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/premiumdownloads/premium_dl_pdm_downloads/3684_0182_52/An-introduction-to-passive-radar-systems_wp_en_3684-0182-52_v0100.pdf?

- [34] Daniel W. O'Hagan, Shaun R. Doughty, and Michael R. Inggs. "Chapter 5 - Multistatic radar systems". In: *Academic Press Library in Signal Processing, Volume 7*. Ed. by Rama Chellappa and Sergios Theodoridis. Academic Press, 2018, pp. 253–275. ISBN: 978-0-12-811887-0. DOI: <https://doi.org/10.1016/B978-0-12-811887-0.00005-5>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128118870000055>.
- [35] Fan Zhang and Chang Liu. "Auxiliary Particle Flow Track-Before-Detect Algorithm for Marine Neighboring Weak Targets". In: *Remote Sensing* 17 (Apr. 2025), p. 1547. DOI: 10.3390/rs17091547.
- [36] Xinghui Wu et al. "Dynamic Programming-Based Track-before-Detect Algorithm for Weak Maneuvering Targets in Range–Doppler Plane". In: *Remote Sensing* 16.14 (2024). ISSN: 2072-4292. DOI: 10.3390/rs16142639. URL: <https://www.mdpi.com/2072-4292/16/14/2639>.
- [37] Xiangyu Peng et al. "An Hybrid Integration Method-Based Track-before-Detect for High-Speed and High-Maneuvering Targets in Ubiquitous Radar". In: *Remote Sensing* 15.14 (2023). ISSN: 2072-4292. DOI: 10.3390/rs15143507. URL: <https://www.mdpi.com/2072-4292/15/14/3507>.
- [38] Chenglong Li et al. "Device-free pedestrian tracking using low-cost ultrawideband devices". In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–4.
- [39] Duo Zhang et al. "mmrotation: Unlocking versatility of a single mmwave radar via horizontal mobility and azimuthal rotation". In: *IEEE Transactions on Mobile Computing* (2025).
- [40] Ingrid Ullmann et al. "A survey on radar-based continuous human activity recognition". In: *IEEE Journal of Microwaves* 3.3 (2023), pp. 938–950.
- [41] Ingrid Ullmann et al. "Classification Strategies for Radar-Based Continuous Human Activity Recognition with Multiple Inputs and Multi-Label Output". In: *IEEE Sensors Journal* (2024).
- [42] Nicolas C Kruse et al. "Reconstruction of Extended Target Intensity Maps and Velocity Distribution for Human Activity Classification". In: *IEEE Transactions on Radar Systems* (2024).
- [43] Ronny Gerhard Guendel et al. *Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes*. Nov. 2021. DOI: 10.4121/16691500. URL: https://data.4tu.nl/articles/dataset/Dataset_of_continuous_human_activities_performed_in_arbitrary_directions_collected_with_a_distributed_radar_network_of_five_nodes/16691500.
- [44] Nir Regev and Dov Wulich. "Radar-Based, Simultaneous Human Presence Detection and Breathing Rate Estimation". In: *Sensors* 21.10 (2021). ISSN: 1424-8220. DOI: 10.3390/s21103529. URL: <https://www.mdpi.com/1424-8220/21/10/3529>.
- [45] Naoki Honma et al. "Human Monitoring Using MIMO Radar". In: *2018 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM)*. 2018, pp. 1–2. DOI: 10.1109/iWEM.2018.8536630.
- [46] Jiacheng Wu, Han Cui, and Naim Dahnoun. "A Novel High Performance Human detection, Tracking and Alarm System Based on millimeter-wave Radar". In: *2021 10th Mediterranean Conference on Embedded Computing (MECO)*. 2021, pp. 1–4. DOI: 10.1109/MECO52532.2021.9460150.
- [47] Dominik Jan Schott et al. "Comparison of direct intersection and sonogram methods for acoustic indoor localization of persons". In: *Sensors* 21.13 (2021), p. 4465.
- [48] Sungwon Yoo et al. "Adaptive clutter suppression algorithm for detection and positioning using IR-UWB radar". In: *2018 9th International Conference on Ultrawideband and Ultrashort Impulse Signals (UWBUSIS)*. IEEE. 2018, pp. 40–43.
- [49] Daehyeon Yim et al. "Quantified activity measurement for medical use in movement disorders through IR-UWB radar sensor". In: *Sensors* 19.3 (2019), p. 688.

- [50] *findpeaks* - Find local maxima - MATLAB. URL: <https://nl.mathworks.com/help/signal/ref/findpeaks.html>.
- [51] R Yarlagadda et al. "Gps gdop metric". In: *IEE Proceedings-radar, sonar and navigation* 147.5 (2000), pp. 259–264.
- [52] Zhixin Zhao et al. "Receiver placement in passive radar through GDOP coverage ratio with TDOA-AOA hybrid localization". In: *IET Conference Proceedings CP779*. Vol. 2020. 9. IET. 2020, pp. 476–480.
- [53] Huazheng Du et al. "A navigation satellites selection method based on ACO with polarized feedback". In: *IEEE Access* 8 (2020), pp. 168246–168261.
- [54] Dennis Milbert. "Dilution of precision revisited". In: *Navigation* 55.1 (2008), pp. 67–81.
- [55] Tiancheng Li, Tariq Pervez Sattar, and Shudong Sun. "Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters". In: *Signal Processing* 92.7 (2012), pp. 1637–1645.
- [56] Adam Johansen. "A tutorial on particle filtering and smoothing: Fifteen years later". In: (2009).
- [57] T. Schon, F. Gustafsson, and P.-J. Nordlund. "Marginalized particle filters for mixed linear/nonlinear state-space models". In: *IEEE Transactions on Signal Processing* 53.7 (2005), pp. 2279–2289. doi: 10.1109/TSP.2005.849151.
- [58] Neil J Gordon, David J Salmond, and Adrian FM Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE proceedings F (radar and signal processing)*. Vol. 140. 2. IET. 1993, pp. 107–113.
- [59] M.A. Richards. *Fundamentals Of Radar Signal Processing*. McGraw-Hill Education (India) Pvt Limited, 2005. ISBN: 9780070607378. URL: <https://books.google.nl/books?id=qizdSv8MEngC>.
- [60] Greg Welch, Gary Bishop, et al. "An introduction to the Kalman filter". In: (1995).
- [61] Dušan Kocur, Mária Švecová, and Jana Rovňáková. "Through-the-wall localization of a moving target by two independent ultra wideband (UWB) radar systems". In: *Sensors* 13.9 (2013), pp. 11969–11997.
- [62] Novelda AS. *X4 - Datasheet: Impulse Radar Transceiver SoC*. Version Rev. E-Preliminary. Nov. 2017.
- [63] Florentin Thullier et al. "A systematic evaluation of the XeThru X4 ultra-wideband radar behavior". In: *Procedia Computer Science* 198 (2022), pp. 148–155.
- [64] Novelda AS. *X4M03 Datasheet: X4 Radar Development Kit*. Version C Preliminary. Oct. 2017.
- [65] Novelda AS. *X4M300 Datasheet: Presence Sensor*. Version Rev. G - Preliminary. Oct. 2018.
- [66] Novelda ASA. *X7F202 Development Kit – Human Presence / UWB Radar Sensor Evaluation Board*. <https://www.digikey.at/en/products/detail/novelda/X7F202-Development-Kit/25558545>. Digi-Key online catalogue, accessed on 16 August 2025. Digi-Key, n.d.
- [67] Nikolaj Andersen et al. "A 118-mW 23.3-GS/s dual-band 7.3-GHz and 8.7-GHz impulse-based direct RF sampling radar SoC in 55-nm CMOS". In: *2017 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE. 2017, pp. 138–139.
- [68] Nikolaj Andersen et al. "4.1 A 79.7μW Two-Transceiver Direct-RF 7.875GHz UWB Radar SoC in 40nm CMOS". In: *2024 IEEE International Solid-State Circuits Conference (ISSCC)*. Vol. 67. 2024, pp. 76–78. doi: 10.1109/ISSCC49657.2024.10454396.
- [69] Christoph Reising et al. "Noise Figure Calculation For Space Surveillance Radar Systems". In: *8th European Conference on Space Debris*. Vol. 8. 2021.
- [70] novelda. *Legacy-SW*. GitHub repository. Software which is currently not under active development. Mar. 2020. URL: <https://github.com/novelda/Legacy-SW>.

- [71] Dingyang Wang, Sungwon Yoo, and Sung Ho Cho. “Experimental Comparison of IR-UWB Radar and FMCW Radar for Vital Signs”. In: *Sensors* 20.22 (2020). ISSN: 1424-8220. DOI: 10.3390/s20226695. URL: <https://www.mdpi.com/1424-8220/20/22/6695>.
- [72] Francesco Blangiardi et al. “Exploring Data Fusion for AI-Based People Counting in a Static Public Transport Vehicle Using Multiple IR-UWB Transceivers”. In: *2024 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE. 2024, pp. 1–8.
- [73] Renming Lu et al. “Through-Wall Human Activity Recognition with Compact MIMO Ultra-Wideband Radar”. In: *IEEE Internet of Things Journal* (2025).
- [74] Oracle VirtualBox. URL: <https://www.virtualbox.org/>.
- [75] Windows 11 System Requirements - Microsoft Support. URL: <https://support.microsoft.com/en-us/windows/windows-11-system-requirements-86c11283-ea52-4782-9efd-7674389a7ba3>.
- [76] Zichao Zhang and Davide Scaramuzza. “A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry”. In: *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE. 2018, pp. 7244–7251.