Efficient and Enhanced Radar Perception for Autonomous Driving Systems

Xiangyu Gao

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Sumit Roy, Chair

Hui Liu

Sajjad Moazeni

Program Authorized to Offer Degree:
Electrical and Computer Engineering

University of Washington

Abstract

Efficient and Enhanced Radar Perception for Autonomous Driving Systems

Xiangyu Gao

Chair of the Supervisory Committee:
Sumit Roy
Department of Electrical and Computer Engineering

Millimeter-wave radars are increasingly integrated into commercial vehicles to support advanced driver-assistance systems, enabling robust object detection, localization, and recognition as a crucial component of environmental perception in autonomous driving systems. This thesis focuses on radar perception algorithm design, incorporating fundamental signal processing, and novel deep learning applications to address open challenges observed in autonomous driving.

To tackle challenging conditions for autonomous driving, where optical sensing may be limited, we propose a novel radar multiple perspectives convolutional neural network (RAMP-CNN). This model extracts object location and class information from range-velocity-angle heatmap sequences. To reduce complexity, we combine lower-dimension network models within our RAMP-CNN, achieving significant performance enhancement. Experimental results demonstrate superior average recall and average precision compared to prior works in all testing scenarios. Notably, the RAMP-CNN model exhibits robust performance during nighttime, showcasing the potential for low-cost radars as substitutes for optical sensing under adverse conditions.

Current vehicular radar imaging suffers from poor azimuth resolution for side-looking operation due to antenna size constraints. To address this limitation, we propose a multiple-input and multiple-output synthetic aperture radar (MIMO-SAR) imaging technique. By applying coherent SAR principles to vehicular MIMO radar, we enhance the side-view

angular resolution. The proposed MIMO-SAR algorithm employs a 2-stage hierarchical workflow, significantly reducing computation load while preserving image resolution. Coherent processing over the synthetic aperture is enabled by integrating a radar odometry algorithm to estimate the trajectory of the ego radar. Validation of the MIMO-SAR algorithm is conducted through simulations and real experiment data collected from a vehicle-mounted radar platform.

Anti-collision assistance (as part of the current push towards increasing vehicular autonomy) critically depends on accurate detection/localization of *moving targets* in vicinity. An effective solution pathway involves removing background or static objects from the scene, so as to enhance the detection/localization of moving targets as a key component for improving overall system performance. We present an efficient algorithm for background removal for automotive scenarios, applicable to commodity frequency-modulated continuous wave (FMCW)-based radars. Our proposed algorithm follows a three-step approach: a) preprocessing of back-scattered received radar signal for 4-dimensional (4D) point clouds generation, b) 3-dimensional (3D) radar ego-motion estimation, and c) notch filter-based background removal in the azimuth-elevation-Doppler domain. The performance of our algorithm is evaluated using both simulated data and experiments with real-world data. By offering a fast and computationally efficient solution, our approach contributes to a potential pathway for challenges posed by non-homogeneous environments and real-time processing requirements.

Overall, this thesis contributes to the advancement of autonomous driving systems by introducing efficient and enhanced radar perception techniques. The proposed algorithms address critical challenges, paving the way for safer and more reliable autonomous vehicles in diverse and complex driving environments.

TABLE OF CONTENTS

Page

LIST OF FIGURES

ACKNOWLEDGMENTS

During my transformative journey over the past five years at the University of Washington, I have been profoundly fortunate to receive unwavering assistance and encouragement from those who have surrounded me. In expressing my heartfelt gratitude, I wish to acknowledge the following individuals whose contributions have shaped my academic and personal growth.

Foremost, my deepest appreciation goes to my esteemed advisor, Professor Sumit Roy. Throughout the entirety of my doctoral pursuit, Professor Roy's sage guidance and mentorship have illuminated my path. His invaluable insights into the rigors of academic scholarship and the art of critical inquiry have indelibly shaped my development as a researcher. Additionally, I extend my gratitude to Professor Hui Liu, my co-advisor, for his instrumental role in initiating my initial research project and for being an unwavering source of advice and support. Professor Liu's guidance was pivotal in honing both my research endeavors and my holistic well-being. I also appreciate his support in guiding my career decisions during a crucial juncture.

I extend my heartfelt thanks to Professors Sajjad Moazeni and Prashanth Rajivan for their dedicated involvement as members of my Ph.D. thesis committee, bearing witness to the milestones of my academic voyage. Their thoughtful feedback and constructive insights have been instrumental in refining my work. A special acknowledgment is reserved for Professors Behcet Acikmese, Payman Arabshahi, and Eli Shlizerman, who served as committee members for my Ph.D. qualifying or general exams. Their rigorous assessments and insightful feedback have greatly enriched the intellectual fabric of my work. I am also indebted to Professors Shunqiao Sun, Akshay Gadre, Chris Rudell, and Jenq-Neng Hwang for their generous help in my academic odyssey.

I extend my heartfelt gratitude to Dr. Guanbin Xing, whose expert guidance and astute insights during the formative years of my Ph.D. paved a strong foundation for my research

endeavors. His expertise in radar and sensing provided an essential starting point for my journey.

My journey would have been incomplete without the camaraderie and shared aspirations of my lab mates: Dr. Sian Jin, Dr. Lyutianyang Zhang, Dr. Hao Yin, Liu Cao, Dr. Collin Brady, and Sachin Nayak. Our collaborative efforts and companionship over the past five years have been a source of continuous inspiration. I am also indebted to scholars such as Dr. Weihua Jiang, Dr. Yizhou Wang, and Zhongyu Jiang, whose collaborations enriched my research pursuits.

I extend my gratitude to all the students who have been integral to my academic voyage, including Leland Wei, Ali Alansari, Alex Luo, Zhichao Lei, Yuchong Liu, Tekhong Meach, Mukund Bharadwaj, Kat Mikhaylyuta, Matthew Sissel, and many others, for their contributions to the research conducted in the FUNLab and in various capstone projects.

To my cherished friends in Seattle and at the University of Washington, including Yangwei Shi, Xiao Lin, Yize Chen, Xuhang Ying, Qiangqiang Guo, Shuo Zhang, Hao Wang, Xinping Yuan, Yaping Sun, Ran Wei, Yang Cheng, Bingqing Peng, Jingyuan Li, Wenqi Cui, Wenzhe Zhang, Xiaofeng Xiang, Baicen Xiao, Zhipeng Liu, Mingfei Chen, Wenqiu Ma, Cheng Chang, Hao Yang, Jiarui Cai, Yue Yang, Qinlin Yu, Haotian Zhang, Yang Zheng, Kun Su, Yonghun Lee, Daniyal Zulfiqar, Yifei Song, Haorui Ji, Jiayi Li, Yuanyuan Shi, Xinyue Sun, Yichen Yuan, Aaron Li, and many others, your camaraderie has brought immeasurable joy to my journey.

I also extend my gratitude to those friends and mentors who have shared moments of camaraderie and support during my internships at Volvo Cars and Intel Labs: Hongyu Cai, Jun Li, Jianhe Yuan, Yuhan Chen, Ekta Samani, Xinkai Zhang, Ying Li, Sihao Ding, Karl Vanas, Harshavardhan Reddy Dasari, Henrik Soderlund, Athma Lakshmi Narayanan, Asma Kuriparambil Thekkumpate, Margaret Huang, Ziwen Song, and Yulin Zeng.

Finally, my heartfelt appreciation goes to my family and my greatest pillar of support, Yan Chen. Their boundless encouragement and unwavering faith have propelled me through the challenges and triumphs of this academic expedition. As I stand at the culmination

of this remarkable journey, I am reminded that my Ph.D. degree is not solely my own achievement; it is a tribute to the collective support and unwavering belief of all those who have been part of this transformative experience.

DEDICATION

to my mom Yunfang Wang, who constantly believed in me.

Chapter 1

INTRODUCTION

1.1 Background

Road accidents, primarily caused by human errors, contribute to approximately 94% of all accidents, as reported by the National Highway Traffic Safety Administration (NHTSA) in a recent technical report [2]. In response to this concerning statistic, autonomous driving systems (ADSs) have emerged as a promising solution with the potential to mitigate accidents, reduce emissions, improve transportation for the mobility-impaired, and alleviate driving-related stress [3]. The widespread deployment of ADSs is projected to yield substantial annual social benefits, estimated to reach nearly $800 billion by the year 2050. These benefits encompass congestion reduction, decreased energy consumption, fewer road casualties, and increased productivity resulting from the reallocation of driving time [4, 5].

The Society of Automotive Engineers (SAE) has defined a comprehensive hierarchy of five driving automation levels [6, 7]. This hierarchy begins with Level-0, indicating no automation, and progresses through Level-1, characterized by primitive driver assistance, such as adaptive cruise control and anti-lock braking systems. Level-2 represents advanced driver-assisted systems (ADAS) with features like emergency braking and collision avoidance. Level-3 involves conditional driving automation with human-in-loop override. Level-4 entails full automation in limited conditions, typically requiring special infrastructure or detailed maps, while Level-5 signifies full automation in all conditions [5, 8]. Presently, most car vendors offer Level-2 automation, exemplified by General Motors' Super Cruise and Tesla's widely recognized Autopilot [9]. Mercedes-Benz and Audi have taken the lead with Level-3 vehicles available to the public. Audi's Traffic Jam Assist enables drivers to disengage from driving under certain conditions, and Waymo operates at Level-4 autonomy in select testing markets, allowing vehicles to navigate alongside other road users without human intervention.

However, achieving Level-4 and 5 driving automation in complex urban road networks poses significant challenges. The dynamic and unpredictable nature of environmental variables, including weather conditions and the behavior of surrounding objects, presents formidable obstacles. In response to these challenges, autonomous vehicles incorporate a diverse array of sensors and vehicle-to-vehicle (V2V) communication [10–14] (or other wireless IoT techniques [15–21]) to empower their perception systems with a comprehensive understanding of the complex and evolving environment. The performance and utilization of these integrated sensors play a crucial role in ensuring the safety and feasibility of automated driving vehicles [22]. Nonetheless, the growing number of sensors also raises concerns regarding hardware costs [23].

In the realm of autonomous vehicles, a multitude of sensors, including cameras, LiDARs, and radars, are strategically mounted on vehicles, earning them the distinction of "smart" sensors due to their ability to provide target tracking, event descriptions, and other relevant information as part of their output [22]. Each sensor type offers unique advantages and contributes to the comprehensive perception system of autonomous vehicles. Cameras, known for their cost-effectiveness, play a pivotal role in detecting both moving and static obstacles within their field of view, offering high-resolution images of the surrounding environment. This capability empowers vehicles to identify critical elements such as road signs, traffic lights, road lane markings, barriers, and various objects, especially relevant in off-road driving scenarios. LiDAR, on the other hand, operates based on emitting pulses of infrared beams or laser light, which then reflect off target objects. Its superiority lies in its high measurement range and accuracy, resilience to environmental changes, and rapid scanning speed. As a result, LiDAR has found widespread application in diverse areas, ranging from autonomous robots, drones, humanoid robots, to autonomous vehicles [22].

Radar, an established technology, employs electromagnetic waves to detect and measure the range, velocity, and angle of objects in the vehicle's environment. Distinguished by its weather resistance and independence from illumination conditions, radar stands as a reliable and robust sensing modality for autonomous vehicles. Moreover, mmWave radar provides fine Doppler resolution and precise estimation for moving targets. Despite its relatively coarse resolution compared to cameras, radar sensors compensate for this limitation by

synergistically integrating data from other sensors, such as cameras and LiDAR, thereby enhancing object recognition capabilities [23].

To enable a reliable and accurate understanding of the environment, this thesis proposal focuses on using radar perception to address three open challenges observed in autonomous driving systems.

The first challenge revolves around object detection under various weather conditions. Inclement weather, such as rain, snow, fog, and harsh sunlight, can significantly impact the performance of traditional sensors like cameras and LiDAR, which are central to object detection in autonomous vehicles. These weather-related obstacles can lead to compromised perception accuracy and decision-making, making it difficult for autonomous vehicles to reliably navigate through diverse environmental conditions. Addressing this challenge is crucial for ensuring the safety and effectiveness of autonomous driving systems, necessitating the development of robust sensor fusion techniques and advanced algorithms that can adapt and perform optimally in adverse weather conditions.

The second challenge is the limitation of having a low-resolution image of the surrounding environment. Low-resolution images may lack the necessary level of detail required for accurate object detection and localization, potentially compromising the safety of passengers and other road users. Moreover, in complex urban environments with dense traffic and intricate structures, low-resolution images may not provide sufficient information for precise path planning and decision-making, leading to suboptimal navigation and increased chances of collisions or traffic violations. The lack of high-resolution images can severely hamper the perception capabilities of autonomous vehicles, hindering their ability to operate effectively and safely in real-world driving scenarios, especially in adverse weather conditions or low-light situations.

The third challenge lies in complicated stationary backgrounds, particularly in urban environments, which can create cluttered and visually complex scenes. This clutter can interfere with the accurate detection and recognition of dynamic objects, making it challenging for autonomous vehicles to distinguish between moving and stationary elements. Additionally, complicated stationary backgrounds may lead to false positives in object detection, potentially causing the vehicle to react inappropriately. Advanced sensor fusion

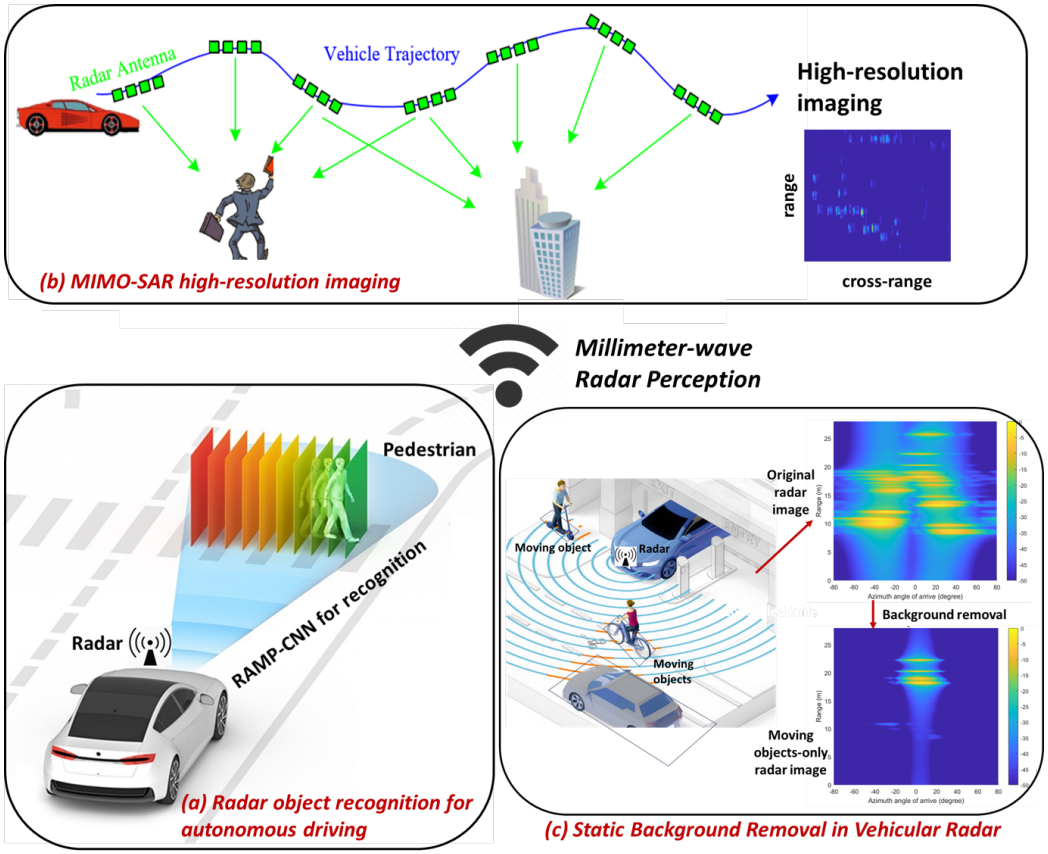

Figure 1.1: Outline of our research contributions: (a) Radar object recognition for autonomous driving, (b) MIMO-SAR high-resolution imaging, (c) Static background removal in vehicular radar.

techniques and sophisticated algorithms are required to effectively differentiate between static and dynamic elements in urban environments, ensuring accurate perception and safe navigation in dense urban areas.

By addressing these challenges through the utilization of mmWave radar perception and developing innovative algorithms, this thesis aims to enhance the perception capabilities of autonomous driving systems, making them more robust, reliable, and safe in various environmental conditions and complex scenarios.

1.2 Summary of Research Contributions

In the deployment of fully autonomous vehicles, numerous challenges arise across perception, decision-making, and control domains. In this thesis, we specifically delve into the role of mmWave radar within the autonomous system and explore its potential in addressing the open challenges faced by autonomous driving systems, with a particular focus on perception. To ensure a reliable and accurate understanding of the environment, this thesis centers on addressing three key perception challenges through the utilization of mmWave radar, as shown in Fig. 1.1.

Enhanced Perception in Adverse Conditions: Adverse environmental conditions, such as nighttime, glaring sunlight, snow, rain, or fog, can significantly impact optical sensing capabilities. To maintain robust perception under such challenging circumstances, we propose a novel radar multiple perspectives convolutional neural network (RAMP-CNN). This network extracts crucial information, including the location and class of objects, by further processing the range-velocity-angle heatmap sequences from mmWave radar. By employing a combination of lower-dimensional network models, we achieve remarkable performance enhancement while managing complexity. The proposed RAMP-CNN aims to deliver semantic environment perception comparable to optical sensors, even in adverse conditions.

Improved Side-View Angular Resolution: Present-day vehicular radar imaging often suffers from poor azimuth resolution during side-looking operation due to limitations in antenna size and placement. To overcome this limitation and enhance side-view angular resolution, we introduce a novel multiple-input and multiple-output synthetic aperture radar (MIMO-SAR) imaging technique. By applying coherent SAR principles to vehicular MIMO radar, we improve imaging capabilities and address the challenge of poor azimuth resolution. The proposed MIMO-SAR algorithm employs a 2-stage hierarchical workflow, significantly reducing computation load while preserving image resolution. We also integrate a radar odometry algorithm to enable coherent processing over the synthetic aperture, ultimately providing improved side-view imaging for enhanced perception.

Efficient Background Removal in Dense Traffic Scenarios: Image understanding within

complex and dense traffic environments presents significant challenges for autonomous driving systems. In these scenarios, removing background or static objects becomes crucial to enhance the detection of moving targets and improve overall system performance. To address this challenge, we present an efficient algorithm for background removal in automotive radar applications, specifically utilizing frequency-modulated continuous wave (FMCW) radar. Our algorithm follows a three-step approach, involving radar signal preprocessing, robust 3D ego-motion estimation, and notch filter-based background removal in the azimuth-elevation-Doppler domain. By efficiently removing background clutter, our approach enhances perception within complex traffic scenarios and offers a fast and computationally efficient solution suitable for real-time processing requirements [24].

In this thesis, we explore the capabilities of mmWave radar in addressing these perception challenges, aiming to contribute to the advancement of autonomous driving technology. By leveraging the strengths of mmWave radar and innovative algorithm design, we seek to improve the safety, reliability, and efficiency of autonomous vehicles, thereby bringing us closer to the realization of fully autonomous driving in diverse and challenging environments.

In addition to the research work related to the thesis, several non-thesis research projects that contribute to the field of radar perception and autonomous driving systems have been undertaken. One such project involved presenting basic radar signal processing algorithms and setting up an experimental testbed [25, 26]. This work laid the foundation for understanding radar signal processing techniques and their application in real-world scenarios. Another significant contribution is the development of an updated testbed featuring Texas Instrument's mmWave cascaded-chip 2D-array radar, along with corresponding signal processing algorithms [27]. Furthermore, I have demonstrated expertise in the development of an open carry and concealed object detection system based on the cascaded-chip radar platform [28]. This research endeavor highlights the practical application of radar perception in security-related scenarios, showcasing the versatility of radar-based systems beyond autonomous driving. Besides, I have shown interest in autonomous driving occupancy representation and automotive radar cognition [29–31]. These interests indicate a holistic approach to advancing the field of autonomous driving, encompassing both perception and decision-making aspects. Moreover, my efforts in sharing open-source radar datasets (raw

I-Q data) [32, 33] collected using 77 GHz FMCW radar showcase a commitment to fostering collaboration and facilitating progress in the field. These datasets serve as valuable resources for researchers to validate and refine various radar algorithms in real-world driving and indoor scenarios, further contributing to the advancement of autonomous driving systems. Outside the realm of autonomous driving, I possess a keen interest in the domain of edge computing. Specifically, I am drawn to the intricate dynamics of mobile edge computing networks (MEC), where my focus lies in the intricate interplay of caching, pushing, and computing, seeking avenues for joint optimization [34, 35].

1.3 Thesis Outline

The remainder of this thesis is organized as follows. In Chapter 2 of this thesis, we proposed an enhanced automotive radar object detection system. In Chapter 3 of this thesis, we propose a hierarchical high-resolution imaging algorithm for side-looking MIMO-SAR radar. In Chapter 4 of this thesis, we propose a static background removal algorithm in the complicated driving environment for vehicular radar. In Chapter 5, we conclude this thesis and point out the future directions of interest. This thesis includes the material in the author's previous papers published on IEEE [25, 27, 36, 37] [1] and arXiv [24].

1.4 Conventions and Notations

Throughout this thesis, we use $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ to represent inverse, transpose, conjugate, and conjugate transpose. x, $\mathbf{x}$, and $\mathbf{X}$ denotes the scalar, vector, and matrix respectively. $\mathrm{Tr}(\bullet)$ denotes the trace. $\lceil \bullet \rceil$ and $\lfloor \bullet \rfloor$ denote integer round up and down.

---

[1] In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University of Washington's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

Chapter 2

ENHANCED AUTOMOTIVE RADAR OBJECT DETECTION

2.1 Motivation

In challenging conditions - nighttime, glaring sunlight, snow, rain, or fog - the utility of pure optical sensing (camera and lidar) is diminished [38]; hence the primary objective of this work is to enable low-cost mmWave radar as a potential substitute. To achieve this, radars should deliver semantic environment perception close to what optical sensors provide. The analog-to-digital converted (ADC) raw data of FMCW radar has 3 dimensions: samples (fast time), chirps (slow time), and receivers), which can be mapped to the 3D radar cube with 3 new dimensions: range, Doppler velocity, and angle. In this work, we adopt the 3-DFFT [25] to obtain the 3D radar cube that is named the range-velocity-angle (**RVA**) heatmap[1].

The small form factor of TI 77 GHz boards - while a desirable feature - limits the number of antennas that can be integrated, resulting in poor angular resolution. Specifically, two targets at the same distance and radial velocity are not resolved in angle if separated by less than resolution beamwidth; even if resolvable, the spatial dimension is not well-defined. Hence to achieve reliable object recognition using such hardware, [25, 39] have sought to exploit the unique movement patterns over time for different classes of objects, i.e. rely on temporal patterns over multiple frames rather than spatial discrimination from single-frame data. While several prior works [25, 39–42] explore radar object recognition with various input data formats using neural network (NN), none has ever combined the spatial and temporal domain information, i.e., by jointly processing the 3D radar cube *sequences* (from multiple frames). Our fundamental contribution is a *deep learning* network design with 3D radar cube sequences as input that approaches performance upper bound by exploiting all available information. However, it is impractical to implement 4D (3D from radar cube plus

[1]In general, by the heatmap we refer to the complex image resulting from the FFT operations. When for visualization purposes, we take the amplitude value of the complex heatmap. In this work, the angle represents the azimuth angle if not specified.

time sequences) convolution processing as the resulting computation complexity exceeds any known CNN models. Therefore, we propose to combine several lower-dimension (3D) models, which nonetheless exceed the performance of prior methods with acceptable computation complexity. Basically, each 3D radar cube (RVA heatmap) is sliced into 2D images from 3 perspectives, that is, range-angle (RA) heatmap, range-velocity (RV) heatmap, and velocity-angle (VA) heatmap. The RA, RV, and VA heatmap sequences are then processed by three parallel DL models to generate different feature bases, which are fused to make the object recognition decision. We name the above radar network architecture RAMP-CNN.

Supervised learning methods need a huge amount of training data and corresponding ground truth labels, a challenge for every deep learning (DL)-based approach. In particular, human labeling of radar data is unreliable even in good conditions, in contrast to labeling camera images. To solve this problem, we propose to set up a vehicular radar-camera system to collect synchronized radar data and camera images for building the UWCR dataset. Cameras are used for teaching radars the locations (range and angle) and classes of objects under good light and weather conditions, which is achieved by implementing the object detection and depth estimation algorithm on captured images. To further improve the performance of the RAMP-CNN model, we propose the following two modifications to avoid overfitting in the training stage, which have been validated in the ablation study.

Radar data augmentation algorithms: Data augmentation encompasses a suite of techniques that enhance the size and quality of training datasets such that better DL models can be built. Traditional image augmentation algorithms include geometric transformations, color space augmentations, mixing images, etc. We propose 4 basic data augmentation operations - flipping, translating, interpolating, and mixing - that work for radar data by accounting for radar imaging physics: energy loss with range and nonuniform angular resolution[12].

New loss function design: The feature basis from the RA, RV, and VA input - named RA, RV, and VA features respectively - are fused within the feature fusion module. When fusing, RA features remain unchanged while RV and VA features are mapped to the range-angle domain. The resulting issue is that NN may well give more weight to the straightforward and accessible RA features than other velocity-based features, leading to

overfitting. Therefore, to push NN to effectively utilize RV and VA features, we add a new term - that only takes RV and VA inputs - to the original loss function.

In summary, the main contributions of this chapter are four-fold:

- Design a novel temporal-spatial RAMP-CNN model that jointly processes the 3D radar cube sequences to achieve superior performance than all prior works, as validated by extensive testing with our UWCR dataset. The proposed RAMP-CNN model is validated to work robustly in the nighttime, where cameras (and other passive optical sensors) are largely ineffective.

- For training the RAMP-CNN model, we propose and establish a vehicular radar-camera system that uses cameras to teach radars the locations and classes of objects under good light and weather conditions.

- To avoid overfitting in the training stage, we propose the modified data augmentation algorithms suitable for radar data and design a new loss function that pushes the RAMP-CNN model to utilize more velocity-related features.

2.2 FMCW Radar Signal Model and Processing

2.2.1 Frequency-Modulated Continuous Wave (FMCW) Radar

FMCW radar transmits a periodic wideband linear frequency-modulated (LFM, also called *chirps*) signal as shown in Fig. 2.1. The transmit (TX) signal is reflected from targets and received at the radar receiver. FMCW radars can detect targets' range and velocity from the receive (RX) signal using the stretch or de-chirping [25] processing structure in Fig. 2.2. The mixer at the receiver multiplies the RX signal with the TX signal to produce an intermediate frequency (IF) signal. At the receiver end, the IF signal is passed into an anti-aliasing low-pass filter (LPF) and an ADC for the following digital signal processing.

2.2.2 Signal Model

The source transmits a sequence of N_{c} chirps (with chirp duration T_{c}) that constitute a frame of duration T_{f}, as shown in Fig. 2.1. With carrier frequency f_{c}, sweep slope S_{w}, amplitude

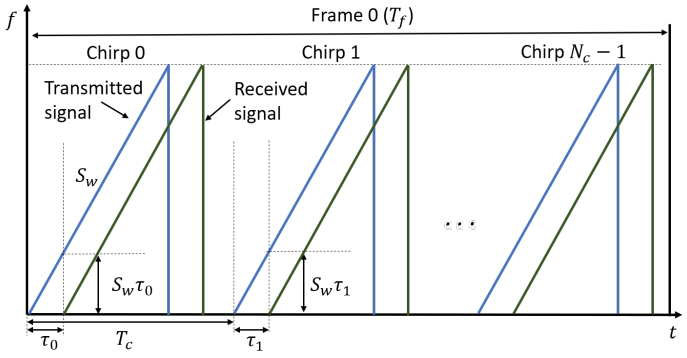

Figure 2.1: The transmitted FMCW chirps and corresponding return signal.

A_{T}, and initial phase ϕ_0, the FMCW transmit signal for chirp 0 is given by [43]:

$$s_{\mathrm{T}}(t) = A_{\mathrm{T}} \cos\left(2\pi\left(f_{\mathrm{c}}t + \frac{1}{2}S_{\mathrm{w}}t^2\right) + \phi_0\right) \tag{2.1}$$

For a target at distance r from radar, the delayed return of the transmitted signal corresponding to the round-trip delay $\tau = \frac{2r}{c_0}$ is given by $s_{\mathrm{R}}(t) = s_{\mathrm{T}}(t-\tau)$, where c_0 is the speed of light. In the receiver chain shown in Fig. 2.2, the received signal $s_{\mathrm{R}}(t)$ is mixed with the transmitted signal $s_{\mathrm{T}}(t)$ to obtain the de-chirped signal which is applied to a low-pass filter to filter out the high-frequency component and retain the difference signal [44]. The resulting intermediate frequency (IF) signal [45] corresponding to chirp 0 is given by:

$$s_{\mathrm{IF}}(t) = \frac{A_{\mathrm{T}}A_{\mathrm{R}}}{2}\left\{\cos\left[2\pi\left(S_{\mathrm{w}}\tau t + f_{\mathrm{c}}\tau - \frac{1}{2}S_{\mathrm{w}}\tau^2\right)\right]\right\} \tag{2.2}$$

Consider the scenario shown in Fig. 2.3, where the vehicle-mounted radar is assumed to have a constant velocity $\boldsymbol{v}_{\mathrm{s}} = (v_{\mathrm{x}}, v_{\mathrm{y}})$ relative to the stationary target within a frame[2], and v_{x}, v_{y} are the velocity components along x- and y-direction. The radar is assumed located at the origin at $t = 0$ without loss of generality; at time t, the distance between radar location $(x_{\mathrm{s}}(t), y_{\mathrm{s}}(t))$ and that of stationary target $(x_{\mathrm{tg}}, y_{\mathrm{tg}})$ is given by $r(t) = \sqrt{(v_{\mathrm{x}}t - x_{\mathrm{tg}})^2 + (v_{\mathrm{y}}t - y_{\mathrm{tg}})^2}$.

[2]We assume the vehicle-mounted radar motion is linear over the (short) frame T_{f} [46]. For different frames, the velocity can be different, as when the trajectory is non-linear.

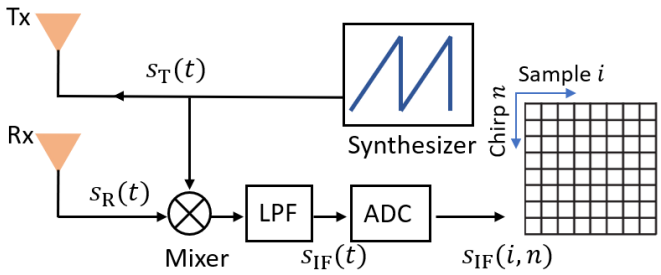

Figure 2.2: The receiver chain for FMCW radar.

For very short chirp duration T_c (a few microseconds), the round-trip delay is assumed fixed for each chirp and therefore can be represented as $\tau_n = \frac{2r(nT_c)}{c_0}$ for chirp n. Across chirps within a frame, τ_n can be expressed as a linear approximation with the relative radial velocity v_r (see Fig. 2.3, where v_r is velocity component projected onto the radial direction θ) [47], i.e.,

$$\tau_n = \tau_0 - \frac{2v_r n T_c}{c_0} \tag{2.3}$$

where $\tau_0 = \frac{2\sqrt{x_{tg}^2+y_{tg}^2}}{c_0}$ is the round-trip delay for chirp 0.

For the above scenario, the quadrature ADC samples of the IF signal of chirp n is given by $s_{IF}(i,n)$, where i is the index of ADC samples within a chirp (see Fig. 2.2). The samples $s_{IF}(i,n)$ are obtained by replacing t and τ in Eq. (2.2) with i/f_s and τ_n, where f_s is the ADC sampling frequency [47].

$$s_{IF}(i,n) = \frac{A_T A_R}{2} \exp\left(j2\pi\left(S_w \tau_n \frac{i}{f_s} + f_c \tau_n - \frac{1}{2} S_w \tau_n^2\right)\right) \tag{2.4}$$

For the MIMO radar with N_{Tx} transmitters and N_{Rx} receivers, we can create a $N_{Tx}N_{Rx}$-element virtual array via transmitting orthogonal waveforms. Therefore, the ADC-sampled IF signal Eq. (2.4) is extended for virtual MIMO operation as follows: for Rx q of the virtual array, we add a phase term $2\pi q h \sin\theta/\lambda$ corresponding to the far-field target's angle of arrival θ [48, 49] into Eq. (2.4).

$$s_{IF}(i,n,q) = \frac{A_T A_R}{2} \exp\left(j2\pi\left(S_w \tau_n \frac{i}{f_s} + f_c \tau_n - \frac{1}{2} S_w \tau_n^2 + \frac{qh\sin\theta}{\lambda}\right)\right) \tag{2.5}$$

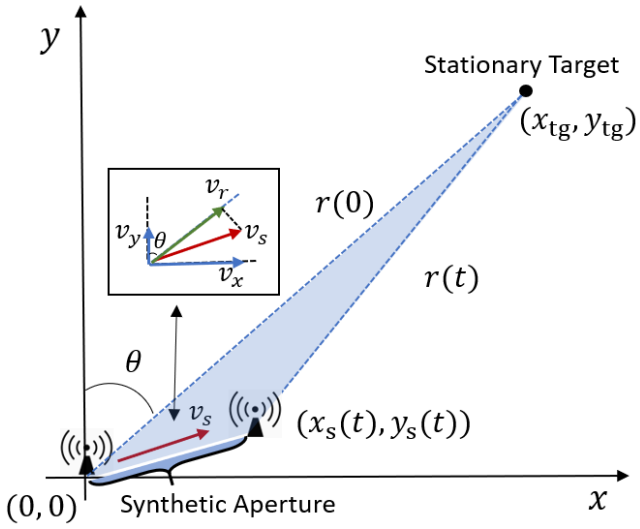

Figure 2.3: The illustration of the scenario for a moving radar within one frame.

where h is inter-Rx distance of virtual array ($h = \lambda/2$).

2.2.3 Range, Velocity and Angle Estimation

We operate the 3-dimensional fast Fourier transform (**3D-FFT**) - Range FFT, Velocity FFT, and Angle FFT - on the digitized IF signal Eq. (2.5) to estimate the range, Doppler velocity, and azimuth angle spectrums. The obtained 3D spectrum is named Range-Velocity-Angle *RVA* cube.

Range Estimation

The IF signal has a beat frequency $f_{\mathrm{b}} = S_{\mathrm{w}}\tau_n$, where τ_n is related to the distance to the target. To estimate the beat frequency, a fast Fourier transform (*Range FFT*) is used to convert the time domain IF signal into the frequency domain [25]; the peaks in the resulting spectrum (or range profile) can be transformed to the distance of target. The Range FFT implemented on chirp n and receiver q is given by $S_{\mathrm{R}}(m_{\mathrm{r}}, n, q) = \mathcal{F}\{s_{\mathrm{IF}}(i, n, q)\}$, where m_{r} is the range bin. The range resolution is determined by the swept RF bandwidth B with the well-known equation $R_{\mathrm{res}} = \frac{c_0}{2B}$.

Velocity Estimation

According to Eq. (2.3) and Eq. (2.4), the relative radial velocity v_{r} will cause a Doppler phase shift $\Delta\phi_{\mathrm{v}} = \frac{4\pi v_{\mathrm{r}} T_{\mathrm{c}}}{\lambda}$ in IF signal between consecutive chirps. Hence, a fast Fourier transform (*Velocity FFT*) is executed across chirps to estimate phase shift and then transform it to velocity [25]. The Velocity FFT performed on the range profile is expressed as $S_{\mathrm{RV}}(m_{\mathrm{r}}, m_{\mathrm{v}}, q) = \mathcal{F}\{S_{\mathrm{R}}(m_{\mathrm{r}}, n, q)\}$, where m_{v} is velocity bin. The velocity resolution of this method is given by $V_{\mathrm{res}} = \frac{\lambda}{2N_{\mathrm{c}}T_{\mathrm{c}}}$ [50], where N_{c} is number of chirps in a frame.

Angle Estimation

The return from a target located at angle θ (far field) results in a steering vector with fixed phase shift $\Delta\phi_{\theta} = \frac{2\pi h \sin\theta}{\lambda}$ for a uniform linear array [49]. Then the angle estimation can be conducted by a fast Fourier transform (*Angle FFT*) across the signal over the Rx elements [25, 51]. The Angle FFT is represented as $S_{\mathrm{RVA}}(m_{\mathrm{r}}, m_{\mathrm{v}}, m_{\theta}) = \mathcal{F}\{S_{\mathrm{RV}}(m_{\mathrm{r}}, m_{\mathrm{v}}, q)\}$, where m_{θ} is azimuth angle bin. The angle resolution for MIMO radar is $\theta_{\mathrm{res}} = \frac{\lambda}{N_{\mathrm{Tx}} N_{\mathrm{Rx}} h \cos\theta}$ [50]. For non-stationary targets, the motion-induced phase errors should be compensated before Angle FFT on the virtual Rx elements corresponding to the second Tx in the case of TDM-MIMO [51]. According to [52], these are corrected via phase compensation of $\frac{\Delta\phi_{\mathrm{v}}}{2}$ (half the estimated Doppler phase shift), which can be obtained from the Velocity FFT results.

2.2.4 MIMO and Virtual Array

The MIMO radar is efficient in improving angular resolution by forming a virtual array and increasing valid antenna aperture. This is achieved by sending orthogonal signals on multiple TX antennas, which enables the contribution of each TX signal to be extracted at each RX antenna. Hence, a physical TX array with N_{T} elements and RX array with N_{R} elements will result in a virtual array with up to $N_{\mathrm{T}} N_{\mathrm{R}}$ unique (non-overlapped) virtual elements [53]. The virtual array is located at the spatial convolution of TX antennas and RX antennas, i.e., convolution produces a set of virtual element locations that are the summed locations of each TX and RX pair. To reduce array cost (fewer physical antenna elements), non-uniform

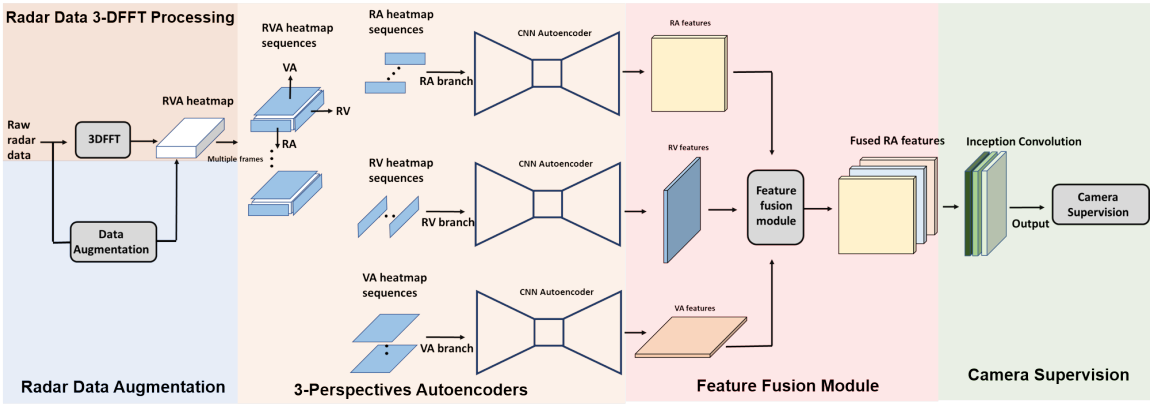

Figure 2.4: The architecture of RAMP-CNN model.

arrays spanning large apertures have been proposed, e.g., the minimum redundancy array (MRA) [27, 54].

When performing angular estimation on a MIMO virtual array, the motion-induced phase errors (i.e., for non-stationary targets) should be compensated on virtual elements before performing Angle FFT. The motion-included phase difference has to be considered under the time-division multiplexing (TDM) scheme because of the switching time between the transmitters. The correction of phase error can be done via compensation of half the estimated Doppler phase shift ($\Delta\phi_{\mathrm{v}}/2$) obtained from the Velocity FFT results [52].

2.3 RAMP-CNN model: A Convolutional Neural Network for Radar Data

2.3.1 3-Perspectives Autoencoders Design

As shown in Fig. 2.4, the main body of the RAMP-CNN architecture is composed of 3 convolutional autoencoders. These autoencoders extract features from the heatmap sequences of different perspectives: RA, RV, and VA.

Why Convolutional Autoencoder: CNNs are known for excellent feature extractors in some tasks, e.g., object detection, and segmentation. The convolutional autoencoders (CAE) - consisting of an encoder and a decoder - render a compact feature representation of the input, by learning the optimal filters that minimize the reconstruction error [55, 56]. The

output feature representation/basis is the same size grid as the input and the cell value of the grid is the feature strength. Each (input) image-plane pixel location maps to multiple feature grid indices. Thus, operations such as weighted-sum followed by a suitable non-linearity on the feature grid cells can be used to determine the presence of particular-type objects at one location. The parameters for such operations (i.e., weights and bias) can be trained by suitable NN iteration.

The Physical Significance of Network Design: The first CAE processes the complex-valued RA heatmap sequences with 3D *conv* layers and *transposed conv* layers. Similar to [42], we pick one RA heatmap from each frame to form the heatmap sequences, and the singled out RA heatmap is obtained by computing Range FFT and Angle FFT at an arbitrarily selected chirp[3]. For the RA heatmap sequences input, those 3D convolution operations take advantage of not only the object's spatial patterns in a single frame but also the temporal information behind the change of spatial patterns across frames. Some aspects of spatial patterns - like the distribution of reflection intensity - directly contribute to object recognition, e.g., larger objects (vehicles) contain stronger-reflectors than small objects (humans). Across multiple frames, the convolution in time would utilize the temporal information behind the change of spatial patterns. As the RA heatmap input is with the complex-valued format, the temporal information can be expressed as the change in both amplitude and phase. The phase change of mmWave signal along time is sensitive to the object movement, e.g. 1mm position movement results in phase shift $\Delta\phi = \pi$ for 77 GHz radar. While the sampling rate of RA heatmap input is a bit lower (30 FPS), we believe the embedded phase shift would still be beneficial.

The second and third CAE process the absolute-value RV and VA heatmap sequences respectively. The RV and VA heatmaps are calculated from the original RVA heatmap by summing the power over the omitted dimension. What two CAEs have in common is: in a single RV or VA heatmap, they extract features from the distribution of range-velocity or velocity-angle cells; while across multiple heatmaps, they extract an object's movement patterns from the change of radial velocity with time. These two CAEs essentially utilize

[3]For the range bin where there exist CFAR detections, we compensate the Doppler phase shift with its maximum-intensity velocity (obtained from the range-velocity spectrum).

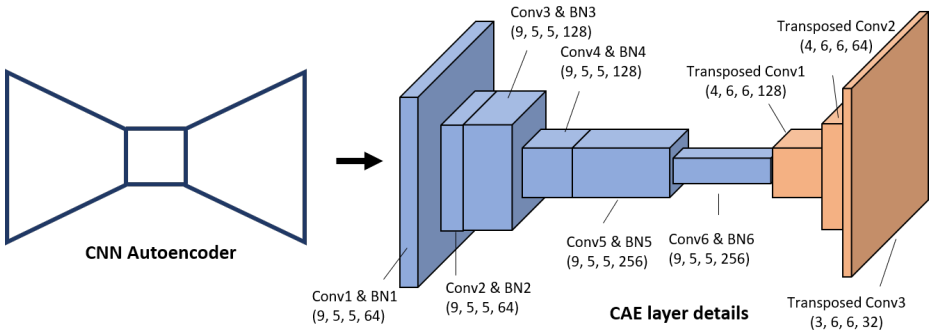

Figure 2.5: Details of Convolutional AutoEncoder (CAE), consist of six 3D *conv* layers and three 3D *transposed conv* layers.

the abundant velocity-based information behind the phase change across chirps within each frame, which is the biggest difference from [42].

Network Details: We adopt the 3D Convolutional-De-Convolutional [55] (shown in Fig. 2.5) model as our CAE, which is effective in summarizing spatiotemporal patterns from raw data into high-level semantics. Each CAE includes six 3D *conv* layers and three 3D *transposed conv* layers. All 3D *conv* layers are followed by a *bn* (batch-normalization) layer and the *relu* activation function. The first two 3D *transposed conv* layers are followed by the *prelu* activation function. The layer details (including parameter selection) of CAE are presented in Fig. 2.5. For illustration, the first blue cuboid part in Fig. 2.5 represents the 3D *conv* layer and a *bn* layer. The kernel size of the 3D *conv* layer is (9, 5, 5) and the number of output feature channels is 64. To preserve the phase information in RA heatmap input, we represent complex-valued heatmap by two real-valued channels that store the real and imaginary parts following [57]. While for RV and VA heatmap inputs, it suffices to only keep the absolute value and use the one-channel representation.

2.3.2 Feature Fusion Module

The feature basis extracted from RA, RV, and VA inputs all support the final classification decision. Our final output is expressed as an image in the RA domain, which means the RA feature can be directly inputted into the network to obtain the corresponding RA-format

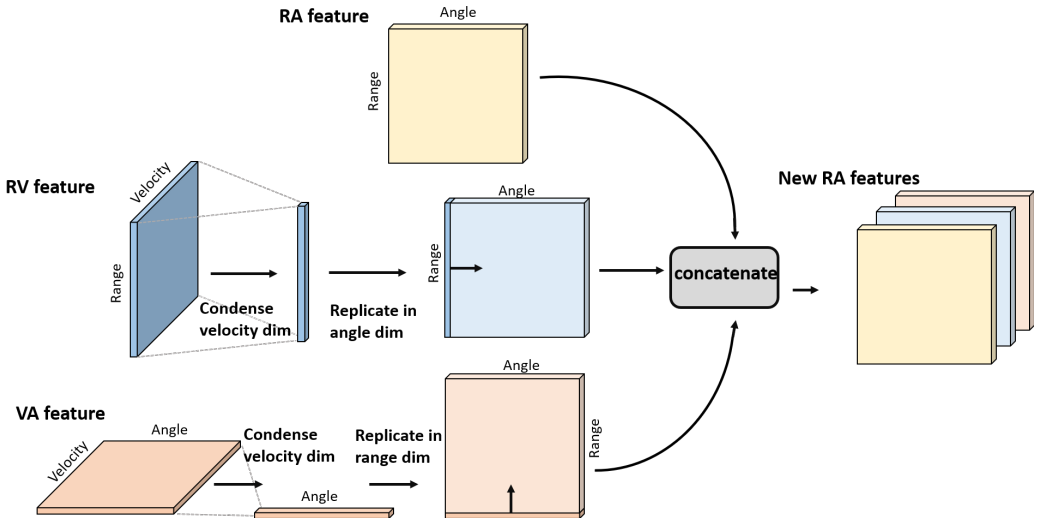

Figure 2.6: The structure of feature fusion module.

output. The key issue is how to further use RV and VA features to support an improved final classification. This is similar to initial human perception using the visual sensor (eyes) supported by supplementary sense organs (ears, nose) for final determination. A person with impaired eyesight will rely more on other sensors, e.g. acquire initial angle information/feature via the ear[4]. The above analogy applied to radar processing suggests how to use the VA feature - which provides good azimuth angle information and no range information. As shown in Fig. 2.6, the VA feature is condensed along the velocity dimension by summing, then the condensed vector is replicated in the missing dimension - range. Similarly, we replicate the RV feature along the angle dimension. Thereafter, we concatenate all features along the channel dimension and input them into the deep network for classification decisions.

Convolution Layers after Feature Fusion Module: There are two *conv* layers that take as input the fused features and make recognition decisions: one 3D *inception* layer, and one ordinary 3D *conv* layers with kernel size (3, 3, 3). Note that the ordinary *conv* layer operates on time, range, and angle dimension, while the *inception* layer operates on channel, range, and angle dimension of fused features. To avoid collapsing the time dimension on the

[4]The ear does not provide good range localization, and hence suggests an equal probability prior to the range.

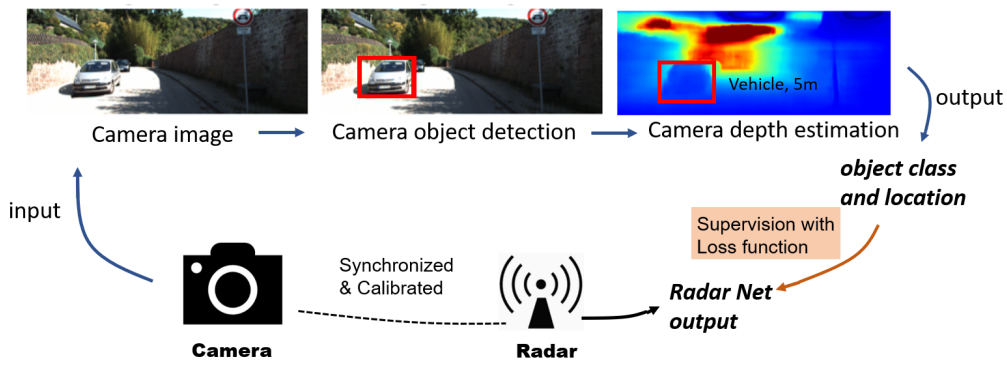

Figure 2.7: The framework of camera teaching radar for the training stage.

inception layer, we repeat the operation on each timestamp and concatenate all inception results along the time dimension. The 3D *inception* layer includes 3 convolution kernels: (3, 5, 5), (3, 3, 3), (3, 1, 21). The first two kernels allow NN to take advantage of multi-level feature extraction, i.e. it extracts both general 5×5 size feature and local 3×3 size features. The last kernel with dilation 6 is used to push the NN to observe a larger area in angle - hence to solve the false alarm problem on the side-lobes[5]. We make it dilated convolution - with angular kernel size 21 and dilation 6 - to cover almost all angle cells, as well as to reduce complexity.

2.3.3 Camera Supervision

The established radar camera system generates synchronized camera images and raw radar data. As shown in Fig. 2.7, we apply the object detection [58] and depth estimation [59] algorithm on captured camera images to obtain the locations (i.e., range and azimuth angle) and classes of all objects of interest in the scene. Under good light and weather conditions, the obtained information from cameras is treated as ground truth for supervising the output of RAMP-CNN. Note that camera-assisted radar learning only happens in the training stage,

[5]The side-lobe in radar heatmap is easy to be recognized as objects of a certain class. This is because the convolution-kernel operators of CAEs do not force each feature to be global (i.e., to span the entire visual field) [56].

while in testing, radar acts independently. To ease the training burden, we use the center key point to represent the existence of objects following [60]. For each ground truth center point $\mathbf{p}$ with location (p_r, p_θ), class id p_c and frame id p_t, we compute its Gaussian representation:

$$Y_{t,r,\theta,c} = \begin{cases} \exp(-\frac{(r-p_r)^2+(\theta-p_\theta)^2}{2\sigma_p^2}) & \text{if } c = p_c \text{ and } t = p_t \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

where σ_p is an object size-adaptive standard deviation. We then splat all ground truth center points onto $Y \in [0,1]^{D\times W\times H\times C}$, and take the element-wise maximum if two Gaussians of the same class and same frame overlap[6]. Y is used as the ground truth in the loss function.

2.3.4 Loss Function for All-perspectives Learning

Let $X_{\text{RA}}, X_{\text{RV}}, X_{\text{VA}}$ be the input RA, RV, and VA heatmap sequences, the aim of RAMP-CNN model is to predict center-point heatmap sequences $\hat{Y} \in [0,1]^{D\times W\times H\times C}$ in RA domain, where $\hat{Y}_{t,r,\theta,c} = 1$ corresponds to a detected center point at range r, azimuth angle θ, frame t and class c, while $\hat{Y}_{t,r,\theta,c} = 0$ represents background. The prediction $\hat{Y}$ includes a map for every frame time. The center point types of each map include $C = 3$ classes of objects: pedestrian, cyclist, and car. For the prediction $\hat{Y}$ and ground truth Y, the training objective is a modified penalty-reduced pixel-wise logistic regression with focal loss [61, 62]:

$$\begin{aligned} L_{\hat{Y}Y} =& \frac{-1}{N_{\text{obj}}} \sum_t \sum_r \sum_\theta \sum_c l(\hat{Y}_{t,r,\theta,c}, Y_{t,r,\theta,c}), \\ \text{where } l =& \begin{cases} \kappa(1-\hat{Y}_{t,r,\theta,c})^\alpha \log(\hat{Y}_{t,r,\theta,c}) & \text{if } Y_{t,r,\theta,c} = 1 \\ \kappa(1-Y_{t,r,\theta,c})^\beta (\hat{Y}_{t,r,\theta,c})^\alpha \log(1-\hat{Y}_{t,r,\theta,c}) & \text{if } Y_{t,r,\theta,c} = 0 \\ & \text{and } Y_{t,r,\theta,\bar{c}} > 0 \\ (1-Y_{t,r,\theta,c})^\beta (\hat{Y}_{t,r,\theta,c})^\alpha \log(1-\hat{Y}_{t,r,\theta,c}) & \text{otherwise} \end{cases} \end{aligned} \tag{2.7}$$

where α and β are hyper-parameters of focal loss [62], and N_{obj} is the number of objects in ground truth. The normalization by N_{obj} is chosen to normalize all positive focal loss instances to 1. Compared to [62], we add a new scalar hyper-parameter κ, which put more

[6]The symbols D, W, H and C here represent the size of Y on time, range, azimuth angle, and class dimension respectively.

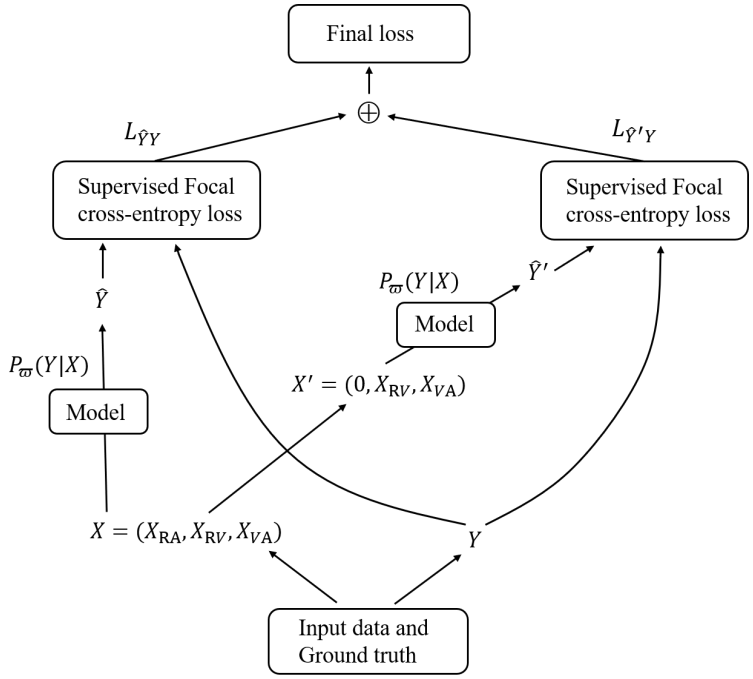

Figure 2.8: The loss function for all-perspectives supervision.

loss/focus at the region where objects exist to shorten the training time. In this work, we choose $\kappa = 4$, and following [61], we use $\alpha = 2$ and $\beta = 4$ in all our experiments.

When designing the loss function, we also take into account the fact that NN may well give more weight to the straightforward and accessible RA features than other velocity-based features, leading to overfitting. This point can be illustrated with the above human perception example again. A person with unimpaired eyesight will not rely much on the other sensors (ears, nose), thus resulting in weaker supplementary function compared to a person with impaired eyesight[7]. The above analogy applied to loss function design suggests how to make NN fully utilize all three perspectives (RA, RV, and VA) and particularly enhance the supplementary function provided by RV and VA perspectives. We add a new loss constraint $L_{\hat{Y}'Y}$ besides the original loss term $L_{\hat{Y}Y}$ mentioned above. To obtain $L_{\hat{Y}'Y}$, we set $X_{\mathrm{RA}} = 0$ in the new loss term, i.e. we input $X = (0, X_{\mathrm{RV}}, X_{\mathrm{VA}})$ to the NN $P_w(Y|X)$

[7]To train this supplementary function for a non-disabled person, it is better to create a situation where eyes are not working, e.g., blindfolding

again such that getting the new prediction $\hat{Y}'$. Then $\hat{Y}'$ is also supervised by ground truth Y with Eq. (2.7) to obtain $L_{\hat{Y}'Y}$.

The final loss is the weighted sum of two terms:

$$L_{loss} = L_{\hat{Y}Y} + \gamma L_{\hat{Y}'Y} \tag{2.8}$$

where γ is the hyper-parameter to balance two terms, chosen to be $\gamma = 0.5$ in this work.

2.4 Radar Data Augmentation Algorithms

Many image data augmentation algorithms have been proposed to increase the amount of *relevant* data and prevent the NN from overfitting, thus essentially boosting overall performance. However, most of the existing data augmentation algorithms cannot be applied to the radar data because of a few key differences from traditional RGB images - complex inputs, energy loss with range, and nonuniform resolution in the angular domain. In this subsection, we focus on 4 basic data augmentation operations and explain how to apply them to radar data: flipping, translating, interpolating, and mixing.

Horizontal Flipping: The horizontal flipping operation will swap the left and right parts of the processed 3D radar cube along the azimuth angle dimension. This operation can be applied to radar data directly as to images because radar has symmetric properties (resolution and antenna gain) in the angular domain.

Translating in Range: With the translating in range operation, we will do a pre-defined range shift Δr for all objects in the RA domain. As shown in Fig. 2.9, for the first step, we transform the polar-coordinates[8] (r, θ) radar data into the uniform Cartesian-coordinates (x, y) radar data with the well-known projection: $x = r\sin\theta,\ y = r\cos\theta$. This relation is nonlinear, therefore the new cell in Cartesian coordinates is not rectangular, interpolation and down-sampling operations are needed to sample the Cartesian plane uniformly. Assume a target at location (r_1, θ_1) with corresponding Cartesian coordinates (x_1, y_1), and the distance from target to radar boresight is fixed even with range shift. Then, the range shift Δr for this target is equivalent to the shift y-axis while keeping the x-axis fixed in the

[8]The processed 3D radar cube is represented as the polar-coordinates format in the RA domain, which is non-uniform for the representation of objects.

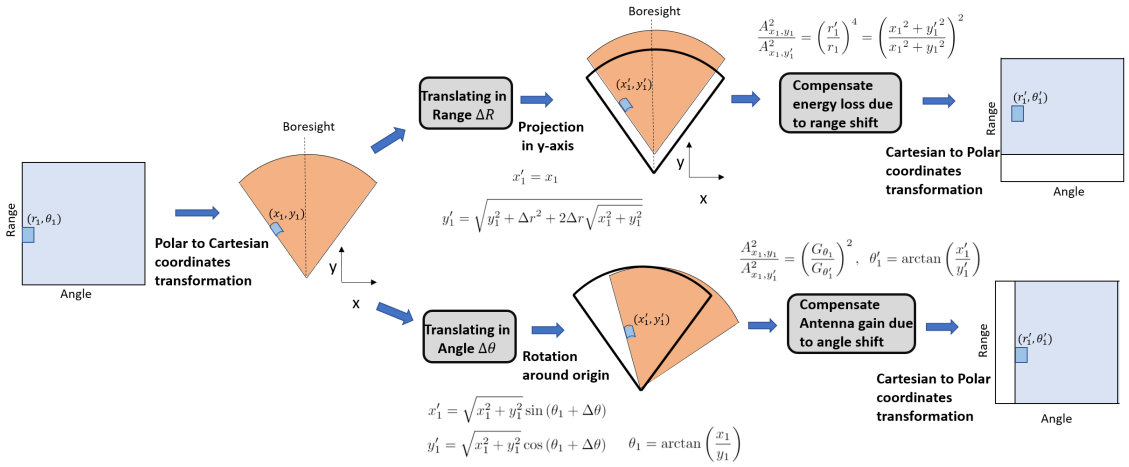

Figure 2.9: The design of translating in range and angle.

Cartesian plane. This maps the previous cell (x_1, y_1) to the new cell (x_1', y_1') with relation: $x_1' = x_1,\ y_1' = \sqrt{y_1^2 + \Delta r^2 + 2\Delta r r_1}$, where $r_1 = \sqrt{x_1^2 + y_1^2}$, and $r_1' = \sqrt{{x_1'}^2 + {y_1'}^2} = r_1 + \Delta r$.

It is easy to prove that the above projection operations can be approached by changing the phase across antennas:

$$\frac{\phi_{r_1',q}}{\phi_{r_1,q}} = \frac{qd\sin\theta'}{qd\sin\theta} = \frac{r_1}{r_1 + \Delta r} \tag{2.9}$$

where $\phi_{r_1,q}$ is the original phase for range r_1 and q^{th} Rx, $\phi_{r_1',q}$ is the phase for projected range r_1' and q^{th} Rx.

Meanwhile, the energy loss due to range shift needs to be compensated according to the radar range equation [63]:

$$\frac{A_{x_1,y_1}}{A_{x_1',y_1'}} = \left(\frac{r_1'}{r_1}\right)^2 = \left(\frac{r_1 + \Delta r}{r_1}\right)^2 \tag{2.10}$$

where A is the signal amplitude.

The details of deriving the relationship for translating the range dimension are presented in Appendix A.1.

Translating in Angle: With the translating in angle operation, we will do a pre-defined angle shift $\Delta\theta$ for all objects in the RA domain. The angle shift in the polar plane is equivalent to the rotation around the origin in the Cartesian plane. Therefore, as shown in Fig. 2.9,

after transforming to the Cartesian-plane data, we use the following relation to map the target in cell (x_1, y_1) to the new cell (x_1', y_1'): $x_1' = r_1 \sin(\theta_1 + \Delta\theta)$, $y_1' = r_1 \cos(\theta_1 + \Delta\theta)$, where $\theta_1 = \arctan(\frac{x_1}{y_1})$, and $\theta_1' = \arctan(\frac{x_1'}{y_1'}) = \theta_1 + \Delta\theta$. If there is no more than one target in a range bin, we can approximate the above projections with shifting $\lfloor \frac{N_\theta d(\sin\theta_1 - \sin\theta_1')}{\lambda} \rceil$ cells in the polar-coordinates angular spectrum, where N_θ is the number of points for Angle FFT. Based on radar range equation [63], we also need to compensate for the antenna gain (G) loss due to angle shift:

$$\frac{A_{x_1,y_1}}{A_{x_1',y_1'}} = \frac{G_{\theta_1}}{G_{\theta_1'}} \tag{2.11}$$

The details of deriving the relationship for translating the angle dimension are presented in Appendix A.2.

Interpolating: The interpolating operation is to fill in the blanks (the white stripes in the last two images of Fig. 2.9) left by the translating operation. We utilize the environment noise for interpolating in order to imitate the situation where there is no object in the blank area. The environment noise samples are obtained by sorting all data of the 3D radar cube with amplitude and then taking the bottom (smallest) 5% of it.

Mixing: The mixing operation is to add up two 3D radar cubes that could have been done with other augmentation techniques - like flipping and translating.

2.5 Experiment

2.5.1 UW Camera-Radar (UWCR) Dataset

A large camera image and radar raw data (I-Q samples post-demodulated at the receiver) dataset for various objects has been collected for multiple scenarios - parking lot, curbside, campus road, city road, freeway, etc. - by a vehicle-mounted platform that is driven. In particular, significant effort was placed into collecting data for situations where cameras are largely ineffective, i.e. under challenging light conditions. We show the camera images and radar range-angle heatmaps of several scenario examples in our UWCR dataset in Fig. 2.11, and present detailed dataset distribution in Table. 2.1.

The data collection platform shown in Fig. 2.10 consists of 2 FLIR cameras (left and right) and two TI AWR1843 EVM radars [64] with configurations in Table. 2.2. Two radar EVM

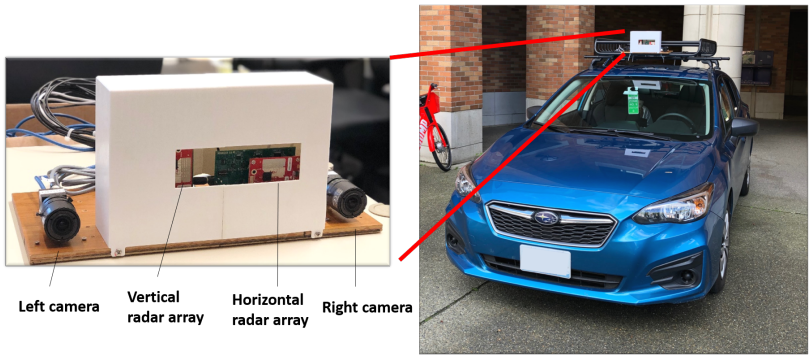

Figure 2.10: Radar-Camera data capture platform: (a) This platform consists of 2 FLIR cameras and two perpendicular radars from TI - the right radar is with the 1D horizontal antenna array, and the left one is with the 1D vertical antenna array. (b) Data capture platform mounted on a vehicle with front view.

boards are placed to form a '2D' antenna array system[9] that can provide more abundant object information. We place one radar array horizontally and the other one vertically to collect the data from both the range-azimuth angle domain and the range-elevation angle domain. We only use the radar data from the horizontal array so far, and we will incorporate the vertical array data into future work. The binocular cameras are synchronized with radars, and they can provide the location and class of semantic objects after we implement the Mask R-CNN detection model [58] and unsupervised depth estimation model [59] on the captured camera images. The semantic object detection results and depth estimation results generated from cameras are manually calibrated and then saved as the requisite ground truth for the following training and evaluation.

[9]Here, '2D' is equivalent to two perpendicular 1-D arrays.

[10]The training set here doesn't count the augmented data.

[11]The nighttime data are not labeled, so we don't count the number of different classes of objects here.

Table 2.1: Dataset distribution for training and test

	Augmented data	Training set[10]	Testing set
Frames	26462	67198	25098
Included ped. cyc. car	53275, 18840, 18731	55203, 24742, 46446	14541, 7607, 7471

	Parking Lot	Curbside	On-road	nighttime
Frames	9900	7200	4398	3600
Included ped. cyc. car	5750, 4501, 2700	3581, 1172, 2039	5210, 1934, 2732	-[11]

Table 2.2: Parameters and configurations of TI's AWR1843 FMCW radar [25]

Parameter	Calculation Equation	Configuration	Value
Range resolution (R_{res})	$R_{\text{res}} = \frac{c_0}{2B} = 0.23$ m	Frequency (f_{c})	77 GHz
Velocity resolution (V_{res})	$V_{\text{res}} = \frac{\lambda}{2N_{\text{c}}T_{\text{c}}} = 0.065$ m/s	Sweep Bandwidth (B)	670 MHz
Angle resolution (θ_{res})	$\theta_{\text{res}} = \frac{\lambda}{N_{\text{Rx}}d\cos\theta} \approx 15^\circ$	Sweep slope (S)	21 MHz/μs
Max. operating range (R_{max})	$R_{\text{max}} = \frac{f_{\text{s}}c_0}{2S} = 28.5$ m	Sampling frequency (f_{s})	4000 Ksps
Max. operating velocity (V_{max})	$V_{\text{max}} = \frac{\lambda}{4T_{\text{c}}} = 8.3$ m/s	Num. of chirps in one frame (N_{c})	255
Max. operating angle (θ_{max})	$\theta_{\text{max}} = \sin^{-1}\left(\frac{\lambda}{2d}\right) = 90^\circ$	Num. of samples of one chirp (N_{s})	128
		Num. of transmitters, receivers	2, 4
		Frame rate (f_{F})	30 FPS

Figure 2.11: 8 scenario examples in the collected UWCR dataset: rows 1, 3 are the camera images; rows 2, 4 are the corresponding radar Range-Azimuth angle heatmaps.

2.5.2 Data Processing

3-DFFT Preprocessing: We implement the 3-DFFT [25] algorithm on the raw I-Q radar data samples to obtain the RVA heatmap sequences. The FFT on range, angle, and velocity dimensions are all 128 points. We choose the input frame number $M = 16$. Therefore, the size of the preprocessed input data is $128 \times 128 \times 128 \times 16$ that corresponds to [range bins $\times$ angle bins $\times$ velocity bins $\times$ frame number].

Data Augmentation: We implement the proposed data augmentation algorithms on the processed RVA heatmap sequences, which include flipping, range translating, angle translating, and mixing the input data. The augmented data is saved locally and **only** is used as part of the training data to avoid overfitting.

Training and Test sets: We partition our UWCR dataset into the training set and test set. Any nighttime scenario data cannot be used in the training set as the corresponding low-light camera images cannot provide the ground truth labels. So all nighttime data is placed into the test set for qualitative performance evaluation only, i.e., the performance of nighttime data isn't evaluated with numerical metrics. The data distribution for the training, augmented, and test set are shown in Table. 2.1 (rows 1-3). Note that the training set in the table doesn't count augmented data, and the whole training data are the collection of the training set and augmented data. The test set is divided into 4 scenarios: parking lot, curbside, on-road, and nighttime. Table. 2.1 also shows their data distribution.

2.5.3 Experiments

We compare the RAMP-CNN model with RODNet-CDC [42], RODNet-HG [42], the state-of-art radar object classification models, as well as the CDMC [25], a model that fully exploits the micro-Doppler signatures of moving objects. We train the RAMP-CNN model and retrain the RODNet-CDC, RODNet-HG, and CDMC following the below details.

Proposed RAMP-CNN model: We train the RAMP-CNN model on a complete training set (includes augmented data) with Cyclic learning rate (minimum 5×10^{-6}, maximum 5×10^{-5}, and cycle duration 860 iterations) [65], batch size 5 for the first 10 epochs. Then we continue to train the RAMP-CNN model with a Step learning rate (starts from 5×10^{-6}, and decays 0.2 every 5 epochs), batch size 4 for the next 24 epochs. We use the Adam gradient descent optimizer [66] and 1 TITAN RTX GPU for the training of all experiments. To verify the capability of the proposed radar data augmentation algorithms (see Section 2.6.2), we also train a new RAMP-CNN model following the same procedures as above, but without augmented data.

RODNet-CDC and RODNet-HG: We train the RODNet-CDC and RODNet-HG model with Cyclic learning rate (same as above), batch size 4 for 10 epochs, and then train them with Step learning rate (same as above), batch size 4 for the following 22 epochs. The gradient descent optimizer is Adam [66]. The loss function for RODNet-CDC is the Minimum Square Error (MSE) provided by PyTorch, and the loss function for RODNet-HG is Cross Entropy. Note that the training set for RODNet-CDC, RODNet-HG, and CDMC [25] model doesn't include the augmented data.

CDMC: Following [25], we generated about 1.2×10^5 concatenated STFT heatmaps in total from the training set. By feeding the training STFT heatmaps to the VGG16 classifier, we trained the model from scratch with the batch size 5, learning rate 1×10^{-4} for the first 10 epochs, and learning rate 1×10^{-5} for the next 10 epochs. The gradient descent optimizer is also Adam [66] and the loss function is the Cross Entropy provided by TensorFlow.

Table 2.3: Performance comparison between different models

Model	Overall		Parking Lot Scenario		Curbside Scenario		On-road Scenario	
	AP	AR	AP	AR	AP	AR	AP	AR
CDMC [25]	30.55%	54.79%	65.74%	76.56%	28.68%	53.93%	4.88%	25.76%
RODNet-HG [42]	71.84%	76.03%	93.87%	95.36%	61.65%	70.09%	41.97%	53.04%
RODNet-CDC [42]	71.46%	78.15%	92.72%	95.07%	64.01%	71.97%	46.52%	58.61%
Prop. RAMP-CNN	**81.23%**	**84.25%**	**97.38%**	**98.37%**	**79.25%**	**84.21%**	**57.07%**	**64.85%**

2.5.4 Evaluation Metrics

We use the average precision (AP) and average recall (AR) to evaluate performance - using true positive, false positive, and false negative rates in Eq. (2.12). Here, true positive (tp) represents correctly located and classified instances, false positive (fp) represents the false alarm, false negative (fn) represents the missed detection and/or incorrectly classified instance.

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}, \quad \text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{2.12}$$

We adopt the CFAR [63] and threshold 0.2 to filter out the target center points from prediction $\hat{Y}$. Whether the targets are correctly located is determined by an object size-adaptive distance threshold, i.e., if the distance between the prediction and ground truth is smaller than the threshold, we assume the prediction is correctly located.

2.5.5 Evaluation Results

We test the RAMP-CNN under 4 different scenarios: parking lot, curbside, on-road, and nighttime (See Fig. 2.12 and Fig. 2.13 for testing examples). The parking lot scenario is manually controlled to have moving and/or static objects in a clear parking lot. For the curbside scenario, we set up the data collection platform on the curbside and then recorded multiple moving objects on a clear road. The on-road scenario is more like an autonomous driving scenario where we drive around and record all objects on the city road. For the

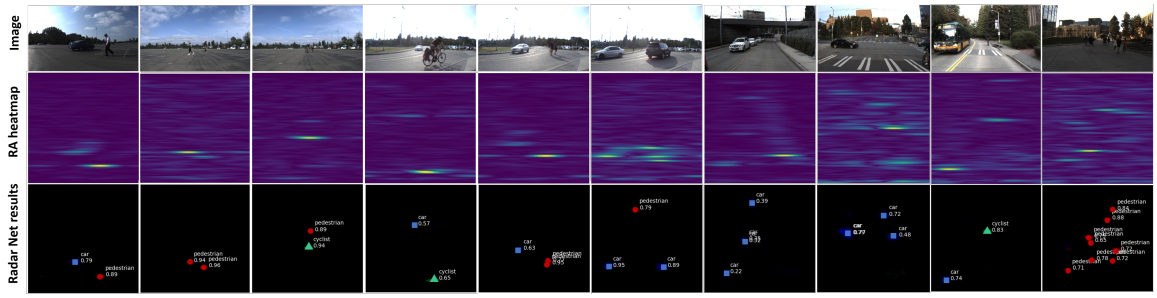

Figure 2.12: 10 test examples from the parking lot scenario, curbside scenario, and on-road scenario: Columns 1-3 are the parking lot scenario; Columns 4-6 are the curbside scenario; Columns 7-10 are the on-road scenario. For each column, the top-row image is the synchronized camera image for visualization, the second-row image is the corresponding radar RA heatmap, and the bottom-row image is the visualization of the RAMP-CNN model results.

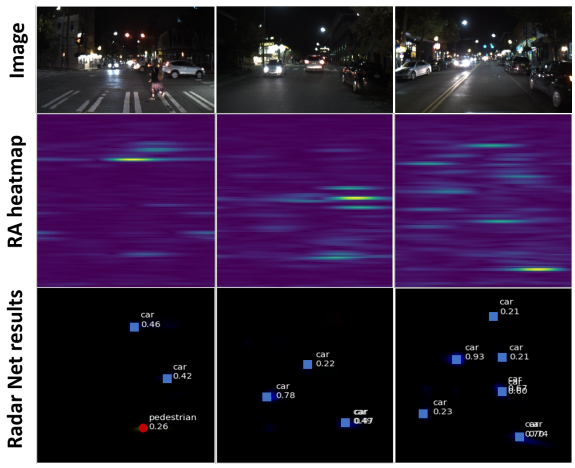

Figure 2.13: 3 test examples from the nighttime scenario. The arrangement of this figure is the same as that in Fig. 2.12.

nighttime scenario, we collect the data under challenging light conditions where cameras are largely ineffective. The testing results of RAMP-CNN and other baselines are shown in Table. 2.3.

The Parking Lot Scenario The parking lot test set has the data of 9900 frames which contain 5750 pedestrians, 4501 cyclists, and 2700 cars. The parking lot scenario is relatively easy for the object recognition task as the background is clean and the objects are few. The RAMP-CNN model achieves **nearly perfect** performance (97.38% AP, 98.37% AR), and beat all prior works. We show 3 test examples in Fig. 2.12.

The Curbside Scenario: The curbside test set has the data of 7200 frames which contain 3581 pedestrians, 1172 cyclists, and 2039 cars. This scenario allows multiple objects to appear at the same time and some of them to be close so that it is harder than the parking lot scenario. The AP (79.25%) and AR (84.21%) of the RAMP-CNN model have **around 15% improvement** over the best results of prior work - RODNet-CDC model (64.01% AP, 71.97% AR). We show 3 test examples in Fig. 2.12.

The On-road Scenario: The on-road test set has the data of 4398 frames which contain 5210 pedestrians, 1934 cyclists, and 2732 cars. This test set is collected from the city-road driving experiments which include several challenging situations, e.g., the reflections from the environment, a large number of cars in the field of view, crowded pedestrians, etc. The RAMP-CNN model obtains **10% improvement in AP** and **6% improvement in AR** over the RODNet-CDC baseline. We show 3 test examples in Fig. 2.12.

The Nighttime Scenario: We test the RAMP-CNN under nighttime to support the objective of this work - advance the cause of radar as a low-cost substitute for optical sensors that fail under such severe conditions. As shown in Fig. 2.13, the RAMP-CNN model performs as well as under the daytime scenario, i.e. radar is impervious/robust to sunlight change. As it is hard to implement ground truth labeling on the nighttime set, we don't numerically evaluate the performance here.

Table 2.4: The performance of RAMP-CNN model for static object scenario and moving object scenario

Model	Static Object Scenario		Moving Object Scenario	
	AP	AR	AP	AR
CDMC [25]	32.34%	50.73%	28.68%	53.93%
RODNet-HG [42]	56.26%	61.36%	61.68%	70.11%
RODNet-CDC [42]	65.02%	69.91%	64.01%	71.97%
RAMP-CNN	**67.58%**	**70.27%**	**79.25%**	**84.21%**

2.6 Analysis and Ablation study

2.6.1 Impact of adding Temporal information

Compared to prior works, the proposed RAMP-CNN model fully exploits the temporal information behind the chirps within one frame, as well as the change of spatial information (range-angle info.) across frames; hence we expect it to essentially achieve performance improvements for moving objects. To verify this, we choose a part of the data from the parking lot and curbside scenario, and redivide them into the static object set and moving object set. We evaluate the performance of RAMP-CNN, CDMC, RODNet-HG, and RODNet-CDC models on the static object set and moving object set respectively. From the evaluation results (Table. 2.4), we know that for static object scenario, the performance of RAMP-CNN (AP around 67%, AR around 70%) in the same level with RODNet-CDC, but much better than RODNet-HG model; for moving object scenario, there is a performance gap (about 15% AP and 13% AR) between RAMP-CNN model (AP 80%, AR 84%) and other baselines (AP around 65%, AR around 71%). The results above verify that the added temporal information in the RAMP-CNN model is helpful for the recognition of moving objects.

Table 2.5: Ablation Study

Model	Data Augmentation	New training loss	Overall		Parking lot scenario		Curbside scenario		On-road scenario	
			AP	AR	AP	AR	AP	AR	AP	AR
RAMP-CNN		✓	76.78%	81.39%	95.91%	97.21%	66.47%	75.83%	52.79%	62.66%
RAMP-CNN	✓		76.93%	81.41%	93.90%	95.46%	75.52%	81.81%	54.10%	61.86%
RAMP-CNN	✓	✓	**81.23%**	**84.25%**	**97.38%**	**98.37%**	**79.25%**	**84.21%**	**57.07%**	**64.85%**

2.6.2 Ablation Study

An ablation study refers to removing some "part" or "module" of the model or algorithm, and seeing how that affects performance. In this subsection, we will study the contribution of two parts of the RAMP-CNN model - data augmentation and new training loss - to the performance. Following the training procedure mentioned in Section 2.5.3, we train two 'incomplete' RAMP-CNN models - one removes the data augmentation part, and the other one replaces the proposed training loss with the ordinary focal loss [61, 62]. The trained models are all evaluated on 4 test sets mentioned above for comparison with the performance of the 'complete' RAMP-CNN model (presented in Table. 2.3). The experiment results of the ablation study are shown in Table. 2.5.

The experiments between the RAMP-CNN model with and without data augmentation confirm that the proposed augmentation algorithms help to boost model performance by avoiding overfitting. For illustration, the data augmentation algorithms make RAMP-CNN get 12% AP improvement and 8% AR improvement in the curbside scenario and get 4% AP improvement and 2% AR improvement in the on-road scenario. The experiments between the RAMP-CNN model with and without the proposed training loss function verify that the proposed training loss helps improve performance by pushing the RAMP-CNN model to learn more Doppler-related features. Specifically, RAMP-CNN obtains around 4% AP improvement and 3% AR improvement in both the parking lot scenario and curbside scenario as well as the on-road scenario.

2.6.3 Complexity Analysis

Time complexity: Time complexity is the amount of time it takes to run the algorithm. We count the number of floating-point operations (**FLOPs**) required by the algorithm, to measure time complexity. The time complexity of the overall CNN is the sum of the time complexity of all *conv* layers[12] [67]:

$$\textbf{Time} \sim \mathcal{O}\left(\sum_{l=1}^{N_{\text{conv}}} I_l^n \cdot K_l^n \cdot C_{l-1} \cdot C_l\right) \tag{2.13}$$

where N_{conv} is the number of *conv* layers, n is the dimension of convolution kernels (1-dim, 2-dim or 3-dim convolution), I is the size of the feature map, K is the size of convolution kernel, C_l is the number of output channels of the l^{th} *conv* layer, that is, number of convolution kernels of this layer. Another indicator to measure a model's time complexity is the training or prediction time. If the time complexity is too high, it will lead to a large amount of time for model training and prediction. Therefore, we also measure the frame-level prediction/testing time for different models to evaluate the time complexity. We show the results in Table. 2.6.

Space complexity: Space complexity quantifies the amount of memory needed by an algorithm to run as a function. This consists of two parts: the total number of parameters (first term of Eq. (2.14)), and the occupied memory of the feature map output at all layers (second term of Eq. (2.14)).

$$\textbf{Space} \sim \mathcal{O}\left(\sum_{l=1}^{N_{\text{conv}}} K_l^n \cdot C_{l-1} \cdot C_l + \sum_{l=1}^{N_{\text{conv}}} I_l^n \cdot C_l\right) \tag{2.14}$$

We show the space complexity results, as well as the number of layers for different models in Table. 2.7. From Table. 2.6 and Table. 2.7, we know that compared to the 4D-CDC model, RAMP-CNN needs almost **100** times fewer FLOPs, around **half** amount of parameters, and **35** times smaller feature map size. For practical application, this means RAMP-CNN would not only run 100 times faster than 4D-CDC for both training and prediction but also take 35 times less memory. That confirms the claimed statement - RAMP-CNN has much less computation complexity than the 4D model. Also, compared to the RODNet-CDC model,

[12]The time cost of fully connected layers and pooling layers is not involved in this formulation. These layers typically take 5-10% computational time.

Table 2.6: Time Complexity Analysis

Model	FLOPs	Prediction time (per frame)
RODNet-CDC	4.75×10^{11}	11.2 ms
4D-CDC[13]	1.64×10^{14}	-[14]
RAMP-CNN	1.41×10^{12}	31.1 ms

Table 2.7: Space Complexity Analysis

Model	Parameters amount	Feature map size	Layers number[15]
RODNet-CDC	3.47×10^{7}	6.31×10^{7}	6, 3
4D-CDC	1.79×10^{8}	6.58×10^{9}	6, 3
RAMP-CNN	1.04×10^{8}	1.89×10^{8}	20, 9

the time and space complexity of RAMP-CNN is around 3 times higher. That, however, means the performance improvement of the RAMP-CNN model comes at the expense of increased complexity.

2.7 Summary

The proposed RAMP-CNN model achieves significant performance improvement over prior works on object recognition under the parking lot, curbside, and on-road scenario, which establishes a new state-of-art baseline. In some hard cases, the radar object recognition func-

[13]To compare the complexity between one 4D model and the RAMP-CNN model, we replace the 3D convolution kernels in the RODNet-CDC model with the 4D convolution kernels and call the new model 4D-CDC.

[14]Note that we didn't implement the 4D-CDC model, so the prediction time and layer numbers are ignored here.

[15]The number of *conv* layers and *transposed conv* layers in models. For RODNet-CDC and RAMP-CNN model, the layers are all 3D; while for 4D-CDC model, all layers are 4D.

tionality of RAMP-CNN might still be poor for supporting autonomous driving presently[16]. However, it can be further improved in the future by incorporating more preprocessing to increase spatial resolution or adopting an advanced radar platform with more antennas. RAMP-CNN is also verified to work in nighttime scenarios, where cameras are largely ineffective due to the low light. Further, prior works [38, 69, 70] have been showing that mmWave radars are with excellent environmental resistance and robustness because the millimeter-wave is less attenuated by fog, rain, or snow. Therefore, we have reason to believe that RAMP-CNN can be applied to these adverse conditions as a good substitute for optical sensors. However, due to the difficulty of capturing data in such circumstances locally, this must be left for future work.

There are several other advantages of the RAMP-CNN model - it has excellent range localization ability because of the centimeter-level range resolution (∼3.75 cm with 4 GHz sweep bandwidth). As shown in Fig. 2.12 (column 10), RAMP-CNN can resolve multiple close pedestrians with range and localize them separately. Besides, the RAMP-CNN model has great generalization for the input data with a higher dimension. For example, if we add the elevation dimension (from the vertical radar array) to the current RAMP-CNN input, then the formed 5D data can still be sliced and processed by several lower-dimension (3D) models that nonetheless achieve better performance with acceptable computation complexity. RAMP-CNN model fully exploits the temporal information behind the chirps in one frame, as well as the change of spatial information (range-angle info.) across frames. Thus, the performance of RAMP-CNN particularly for moving objects, shows significant improvements relative to state-of-art. The ablation study shows that both the proposed data augmentation algorithms and training loss are helpful for boosting the performance of RAMP-CNN. It is worth noting that major performance improvement comes from the main body of the RAMP-CNN model (3-Perspectives model); the cumulative impact of all elements in the RAMP-CNN architecture results in promising performance improvement, at the expense of increased complexity.

[16]Based on the performance of camera object detection in [68], we infer that the acceptable/desired average precision and recall for radar would be 0.8 with different classes of objects.

Chapter 3

HIGH-RESOLUTION SIDE-LOOKING RADAR IMAGING

3.1 Motivation

Side-looking radars are widely used to support lane change or keeping assist, blind spot detection, and rear cross-traffic alert. For these applications, a high-resolution radar image is key to the effective separation of close objects, detection of the spatial extension of traffic participants, and enhanced object recognition [36]. 77 GHz FMCW radars can already achieve centimeter-level range resolution (~ 3.75 cm with 4 GHz sweep bandwidth). The remaining challenge is to improve the azimuth angle (cross-range) resolution. There exist two main classes of approaches for side-view automotive radars to improve azimuth resolution: MIMO processing [25, 27, 51, 71], and synthetic aperture radar (SAR) [72–74].

MIMO radar systems transmit mutually orthogonal signals from multiple transmit antennas, followed by *joint* processing of the signals at all receive antennas [51] for extraction of target information. For example, if a MIMO radar has N_{Tx} Tx antennas and N_{Rx} Rx antennas with appropriate arrangement, a $N_{\mathrm{Tx}}N_{\mathrm{Rx}}$-element virtual antenna array [51] can be created by conducting beamforming on the received signals, thereby obtaining a finer azimuth resolution compared with its phased array counterpart. The state-of-art Texas Instrument (TI) 77 GHz FMCW radar chip AWR1843 is capable of synthesizing 8 virtual receivers with 2 Tx and 4 Rx [64] using TDM [51, 75], resulting in 15° azimuth resolution [25] which is insufficient for high-resolution target discrimination.

SAR is a well-known imaging technique that overcomes the limits of a small physical aperture on the angular resolution by coherently processing the returns from a series of transmitted pulses to create a large *synthetic* aperture[48]. SAR is applicable for imaging stationary targets and background with an active source mounted on a moving platform (e.g. vehicle-mounted radar), while the moving targets can be observed via inverse SAR (ISAR) concepts with a stationary radar setup [76, 77]. Besides, automotive (vehicle-based)

SAR algorithms need to cope with specific difficulties - notably estimation of vehicle motion and (importantly) achieving real-time operation. Accordingly, prior research on automotive SAR algorithms has focused on speeding up post-processing via techniques to reduce the algorithmic run-time complexity [72, 78–80].

In this section, we propose a new MIMO-SAR algorithm that exploits key features of FMCW MIMO radar to achieve computationally efficient SAR imaging. Specifically, MIMO processing is used for initial low-cost target detection and localization to narrow down the *region of interest* (ROI) for subsequent finer-resolution SAR processing. We adopt the time-domain Back-projection SAR algorithm [73] - that lends itself naturally to parallel processing with graphics processing units (GPU) [81, 82] - to progressively operate on ROI as new snapshots are coherently added and processed. To reduce SAR processing frequency (i.e. PRF) and consequent complexity, the returns from a train of FMCW chirps are stored in a range-velocity-angle (RVA) data cube that is processed via the 3D-FFT algorithm; thereafter we select the max-intensity velocity component for subsequent SAR processing. The above 2-stage hierarchical workflow drastically reduces the computation load while preserving high-resolution imaging.

Coherent SAR processing requires phase compensation of source motion trajectory to achieve coherent in-phase processing. In turn, this requires accurate vehicle ego-motion estimation. The traditional solution is to use onboard inertial measurement units (IMU) [83] that combine measurements from the wheel speed sensor, gyroscopes, and accelerometers. However, high-precision IMUs are cost-prohibitive for automotive applications, inspiring the need for self-contained alternatives such as radar odometry, to determine the velocity and direction of motion of the vehicular radar [1, 46, 84]. Our approach is based on analyzing the distribution of the radial velocities of the received reflections (targets) over their azimuth angles, which can be provided by the MIMO processing stage described above.

In summary, the main novel contributions of this chapter are four-fold:

- A new hierarchical MIMO-SAR algorithm that reduces computational complexity while preserving high-resolution imaging.

- A radar odometry algorithm to estimate the trajectory of ego-radar and enable MIMO-

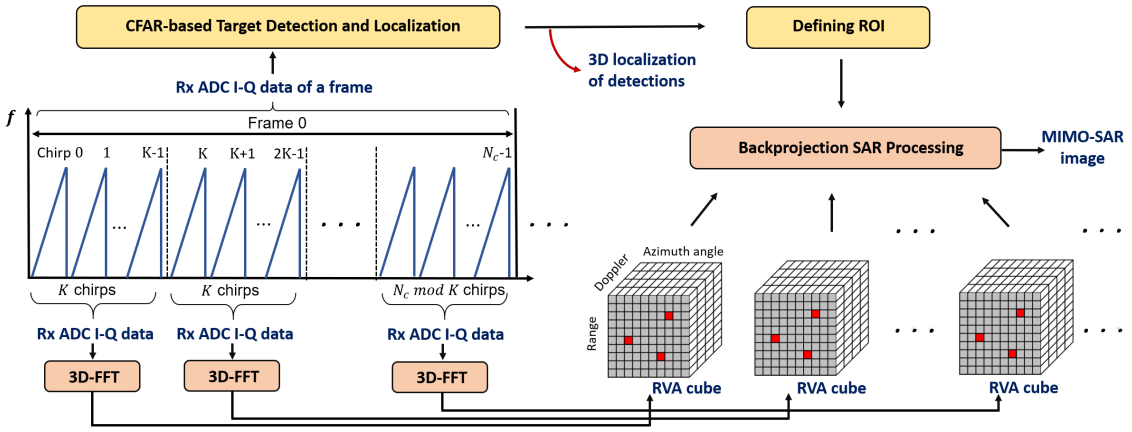

Figure 3.1: MIMO-SAR algorithm illustration: target detection and localization (MIMO processing) are operated for every frame; back-projection SAR algorithm is performed on RVA cubes generated from every K chirps by 3D-FFT.

SAR coherent processing.

- Discussion on the limit of pulse repetition frequency (PRF), coherent processing interval (CPI), and region of interest (ROI) to ensure system performance.

- Validation of the proposed MIMO-SAR algorithm by both simulations and real data.

3.2 Signal Model

The signal model used in this work is consistent with the description provided in Chapter 2.2. As previously mentioned, we consider an FMCW radar system operating in a MIMO configuration. The received radar signal is modeled as a set of post-demodulated I-Q samples, which capture the reflection and scattering information from the surrounding environment.

3.3 MIMO-SAR Algorithm

The hierarchical MIMO-SAR algorithm workflow is divided into 2 stages - MIMO processing, and SAR processing - to reduce the computation load, as shown in Fig. 3.1. The load

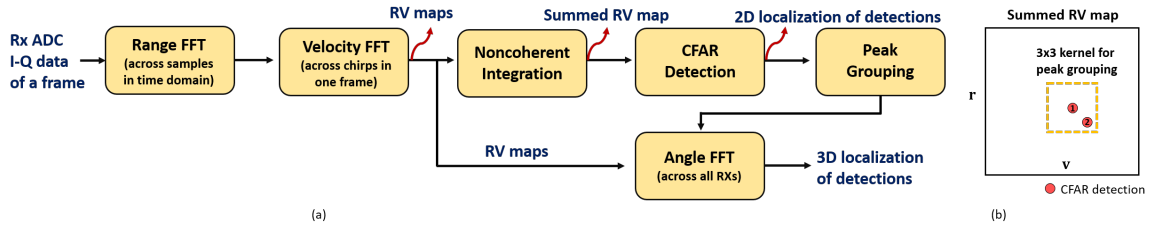

Figure 3.2: (a) The block diagram for CFAR-based target detection and localization algorithm; (b) Illustration for peak grouping.

decrease is achieved by both spatially focusing on the Regions-of-Interest (ROI) provided by MIMO processing and temporally reducing the SAR sampling rate (or PRF) using RVA cubes. Specifically, MIMO processing is performed for every frame to localize the detected targets in the field of view such that we can narrow down the ROI for subsequent SAR imaging. For SAR processing, we implement the Back-projection algorithm [73, 81] on the RVA data cubes (provided by 3D-FFT on every K chirps[1]) to obtain high-resolution imaging.

3.3.1 MIMO Processing

MIMO radar can provide an estimation of range, Doppler (radial) velocity, and azimuth angle for the detected targets in the field of view, which confers two advantages to our MIMO-SAR algorithm compared to traditional SAR processing: 1) Initial target localization enables a hierarchical approach whereby selecting ROI within imaging plane for subsequent SAR processing effectively reduces computation complexity; 2) Analyzing estimated Doppler velocities and azimuth angles for detected stationary targets enables radar ego-motion estimation, which is necessary for SAR phase compensation along trajectory (Section 3.3.3).

The computation load of SAR imaging increases in direct proportion to the potential target area of the SAR image plane. Classic SAR algorithms coherently process each cell in the SAR image plane. This is redundant since in practice, most of the energy (or

[1]The MIMO localization and radar odometry are operated every frame because the accuracy and resolution of velocity estimation are greater with longer observation [50]. However, the SAR image is updated once for every K chirps (a subset of the frame) to avoid the azimuth aliasing.

useful information) is concentrated in a subset. Therefore, to reduce the computational complexity, we propose to select the ROIs in SAR images based on the localization of detections provided by constant false-alarm rate (CFAR) detector [48, 85], and then perform subsequent processing only on the selected regions. The CFAR-based target detection and localization algorithm operates on the raw radar I-Q samples vide Eq. (2.5) for each frame as presented in Fig. 3.2(a). First, the Range and Velocity FFTs are performed on I-Q data in a frame to obtain the range-velocity (RV) map for initial target detection. The RV maps from all receivers are integrated non-coherently (i.e., sum the magnitude RV maps) to increase the signal-to-noise ratio (SNR) of the resulting RV map [86].

Post summing, the cell-averaging CFAR [48] algorithm is applied to detect targets and obtain their 2D (range and velocity) localization. During the CFAR detection process, each cell/bin is evaluated for the presence or absence of a target using a threshold, and the threshold adapts itself according to the noise power estimate within a sliding window. Thereafter, peak grouping for all CFAR-detected targets is done based on their 2D localization - see Fig. 3.2(b) - by checking if the detected amplitude is greater than that of its neighbors. For an illustration of Fig. 3.2(b), in the 3×3 kernel centered at CFAR detection 1, CFAR detection 2 will be discarded if it is with smaller amplitude. In the end, we calculate the Angle FFT[2] for each detected target across the RV maps of all receivers (in the virtual array formed by TDM-MIMO) to estimate its azimuth angle.

Assume a group of targets is detected from the first frame's radar data with the above algorithm. For each detection with range and azimuth angle (r, θ), its relative Cartesian position $(x_\mathrm{r}, y_\mathrm{r})$ with respect to radar is given by $x_\mathrm{r} = r \sin\theta$, $y_\mathrm{r} = r \cos\theta$. The relative position $(x_\mathrm{r}, y_\mathrm{r})$ is transformed to the position (x, y) in global geometry plane (see Fig. 3.3) by adding radar trajectory$(x_\mathrm{s}(t), y_\mathrm{s}(t))$, i.e., $(x, y) = (x_\mathrm{r}, y_\mathrm{r}) + (x_\mathrm{s}(t), y_\mathrm{s}(t))$. Based on the calculated global positions of targets, say $\{(x_1, y_1), (x_2, y_2)\}$ for example, we then can define

[2]Before this Angle FFT, we compensate the motion-induced phase error for TDM-MIMO, using the detected Doppler velocity.

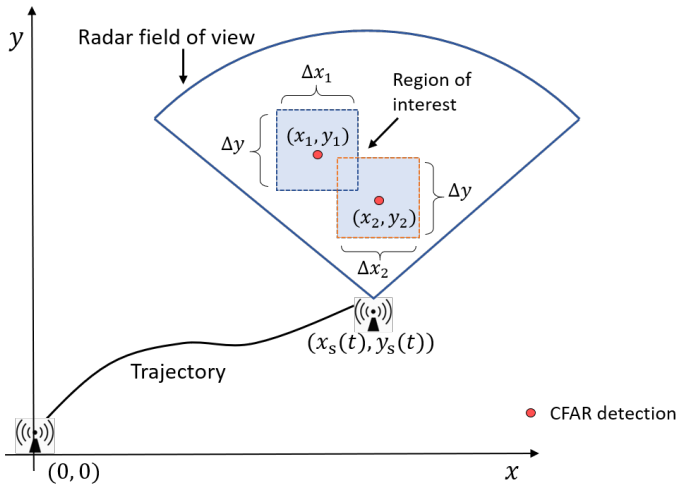

Figure 3.3: A example that shows how to define the ROI based on the localization of 2 detected targets (x_1, y_1), (x_2, y_2).

the ROI U_g in geometry plane that need to be imaged:

$$\Big\{(x', y') \,|\, \{x_1 - \frac{\Delta x_1}{2} \leq x' \leq x_1 + \frac{\Delta x_1}{2},\, y_1 - \frac{\Delta y}{2} \leq y' \leq y_1 + \frac{\Delta y}{2}\} \\ \cup \{x_2 - \frac{\Delta x_2}{2} \leq x' \leq x_2 + \frac{\Delta x_2}{2},\, y_2 - \frac{\Delta y}{2} \leq y' \leq y_2 + \frac{\Delta y}{2}\}\Big\} \tag{3.1}$$

where the overall regions are the union of two small regions that are centered on (x_1, y_1), (x_2, y_2) respectively. The Δx_1, Δx_2, Δy are the side length of two small regions. For each region, we use a fixed side length Δy along the y-axis and customize the side length along the x-axis for defined azimuth scope $\Delta\theta$ (i.e., $\Delta x = r\Delta\theta$).

The radar trajectory $(x_s(t), y_s(t))$ is required for defining ROI U_g (indicated above), as well as for the phase compensation for coherent SAR processing. In Section 3.3.3, we explain the radar odometry algorithm that takes the radial velocities and azimuth angles of detected targets as input to estimate the moving trajectory of ego-radar.

3.3.2 Back-projection SAR Processing

In the last section, we use the MIMO processing to define ROI U_g in the geometry plane that needs to be imaged. In this section, we would operate the Back-projection algorithm on ROI

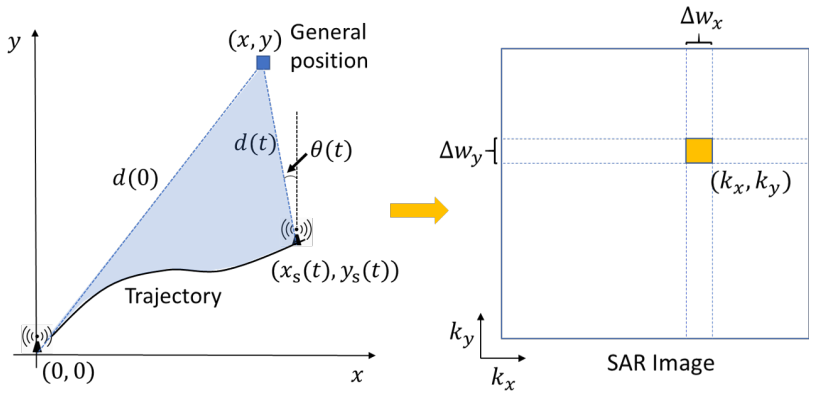

Figure 3.4: The transformation from a general position (x, y) in geometry plane to the pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ in SAR image plane.

to get a high-resolution image. Firstly, the region U_{g} in the geometry plane is transformed to a grid of pixels U_{s} in the SAR image plane. As illustrated in Fig. 3.4, a general position (x, y) in geometry plane corresponds to the pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ in the SAR image plane:

$$x = k_{\mathrm{x}}\Delta w_{\mathrm{x}},\ y = k_{\mathrm{y}}\Delta w_{\mathrm{y}} \tag{3.2}$$

where Δw_{x}, Δw_{y} are the pixel width and height.

For each pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ in the U_{s}, we coherently sum up its corresponding radar measurement for multiple snapshots using matched filter response. Such an algorithm updates the SAR image for every snapshot and reuses the previous calculations when updating with any new measurement. As presented in Fig. 3.1, we select one snapshot from every K chirps with a 3D-FFT algorithm to reduce SAR sampling rates. The parameter K is carefully chosen based on Section 3.4.1 to avoid azimuth aliasing in defined scope $\Delta\theta$. Using 3D-FFT[3], the radar I-Q samples of K chirps are converted to an RVA data cube where the corresponding radar measurement for pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ can be found. That is, for pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$, we calculate its relative distance d and azimuth angle θ to radar, and use them (d, θ) to localize the corresponding radar measurement in RVA data cube S_{RVA} (i.e., a radar measurement is

[3]Note that we use small FFT points here to avoid the migration [87] and also to reduce the computation complexity.

represented by a small red cube in Fig. 3.1). For example, we consider a train of K chirps that start from t (e.g., $t = nT_{\mathrm{c}} + pT_{\mathrm{f}}$ for chirp n of frame p). At time t, the relative distance $d(t)$ and azimuth angle $\theta(t)$ for pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ is given by:

$$\begin{aligned} d(t) &= \sqrt{(x_{\mathrm{s}}(t) - k_{\mathrm{x}}\Delta w_{\mathrm{x}})^2 + (y_{\mathrm{s}}(t) - k_{\mathrm{y}}\Delta w_{\mathrm{y}})^2} \\ \theta(t) &= \arcsin\left(\frac{x_{\mathrm{s}}(t) - k_{\mathrm{x}}\Delta w_{\mathrm{x}}}{d(t)}\right) \end{aligned} \tag{3.3}$$

The calculated $d(t)$ and $\theta(t)$ are mapped to the range bin m_{r} and angle bin m_θ of the RVA cube respectively [36] to localize the radar measurement: $m_{\mathrm{r}} = \lfloor \frac{2M_{\mathrm{r}}S_{\mathrm{w}}d(t)}{c_0 f_{\mathrm{s}}} \rfloor$, $m_\theta = \lfloor \frac{M_\theta h(x_{\mathrm{s}}(t) - k_{\mathrm{x}}\Delta w_{\mathrm{x}})}{\lambda d(t)} \rfloor$. The velocity bin m_{v} is chosen to have the max-intensity velocity component at the location $(m_{\mathrm{r}}, m_\theta)$. Therefore, the corresponding radar measurement for pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ at time t is given by $S_{\mathrm{RVA},t}(m_{\mathrm{r}}, m_{\mathrm{v}}, m_\theta)$. For a coherent summation of the radar measurements from multiple snapshots, their phase must be compensated by a matched filter response $H(t)$ for the relative distance $d(t)$ [73] [81].

$$H(t) = \exp\left(j2\pi f_{\mathrm{c}} \frac{2d(t)}{c_0}\right) \tag{3.4}$$

The last step for Back-projection SAR processing is to coherently sum the compensated radar measurements from all observed snapshots. For example, for N_{f} frames as input where each frame contains N_{c}/K snapshots, there are $N_{\mathrm{f}}N_{\mathrm{c}}/K$ snapshots in total. We denote the start time for the l^{th} snapshot of frame p by $t_{l,p} = (l-1)KT_{\mathrm{c}} + pT_{\mathrm{f}}$. Then the Back-projection processing for pixel $(k_{\mathrm{x}}, k_{\mathrm{y}})$ is given by:

$$I\left(k_{\mathrm{x}}, k_{\mathrm{y}}\right) = \sum_{p=0}^{N_{\mathrm{f}}} \sum_{l=0}^{N_{\mathrm{c}}/K} H^*(t_{l,p}) \cdot S_{\mathrm{RVA},t_{l,p}}(m_{\mathrm{r}}, m_{\mathrm{v}}, m_\theta) \tag{3.5}$$

To form an image using this method, we apply Eq. (3.5) for each pixel in ROI U_{s}, and leave the rest of the pixels in the SAR image blank.

3.3.3 Ego-motion Estimation with Radar Odometry

As mentioned earlier, one benefit of FMCW MIMO radar is access to an accurate estimate of the target's radial Doppler velocity and azimuth angle within a frame [25]. This in turn enables radar odometry since the sensor's velocity can be estimated by analyzing the

relationship between the radial Doppler velocities and azimuth angles of all static targets in the field of view [1, 84]. As illustrated in Fig. 3.5, if a radar sensor is moving with v_{s}, then all stationary targets move with relative velocity (blue line) equal to the sensor's speed with opposite heading. Given that Doppler radar only measures the radial velocity component (green line) of the target, we can reconstruct the sensor's velocity components along the x-axis and y-axis $(v_{\mathrm{x}}, v_{\mathrm{y}})$ by analyzing the velocity profile of at least two stationary targets [1, 84].

For N_{d} detected stationary targets in a frame, the velocity profile of these targets is given by Eq. (3.6), where $v_{\mathrm{r},i}$ and θ_i are the measured radial Doppler velocity and azimuth angle for i^{th} stationary target [1].

$$\begin{bmatrix} v_{\mathrm{r},1} \\ \vdots \\ v_{\mathrm{r},N_{\mathrm{d}}} \end{bmatrix} = - \begin{bmatrix} \sin(\theta_1) & \cos(\theta_1) \\ \vdots & \vdots \\ \sin(\theta_{N_{\mathrm{d}}}) & \cos(\theta_{N_{\mathrm{d}}}) \end{bmatrix} \begin{bmatrix} v_{\mathrm{x}} \\ v_{\mathrm{y}} \end{bmatrix} \tag{3.6}$$

We solve Eq. (3.6) via the minimum mean-square-error (MMSE) estimator [88] to obtain the frame-level radar velocity $\hat{\boldsymbol{v}}_{\mathrm{s}} = (\hat{v}_{\mathrm{x}}, \hat{v}_{\mathrm{y}})$. We denote the estimated sensor velocity for frame p by $\hat{\boldsymbol{v}}_{\mathrm{s},p} = (\hat{v}_{\mathrm{x},p}, \hat{v}_{\mathrm{y},p})$ and assume that radar moves with constant velocity within frame [46]. Then the position of radar $(x_{\mathrm{s}}(t),\, y_{\mathrm{s}}(t))$ at time t is obtained by integrating the estimated sensor velocity profile. Considering the chirp n of frame p that starts at time $t = nT_{\mathrm{c}} + pT_{\mathrm{f}}$, the corresponding radar position estimation $(\hat{x}_{\mathrm{s}}(t),\, \hat{y}_{\mathrm{s}}(t))$ is given by:

$$\begin{aligned} \hat{x}_{\mathrm{s}}(t) &= \sum_{u=0}^{p-1} \hat{v}_{\mathrm{x},u} T_{\mathrm{f}} + \hat{v}_{\mathrm{x},p} n T_{\mathrm{c}} \\ \hat{y}_{\mathrm{s}}(t) &= \sum_{u=0}^{p-1} \hat{v}_{\mathrm{y},u} T_{\mathrm{f}} + \hat{v}_{\mathrm{y},p} n T_{\mathrm{c}} \end{aligned} \tag{3.7}$$

where the initial location of radar is set to the origin without loss of generality, T_{f} is the frame duration, and T_{c} is the chirp duration.

In reality, moving objects are expected in an experimental scenario, resulting in mixing detection of stationary objects and moving objects. Therefore, we employ Random Sample Consensus (RANSAC) algorithm [89] prior to radar odometry to separate out needed stationary targets and determine N_{d} [1]. RANSAC is an iterative method for optimal

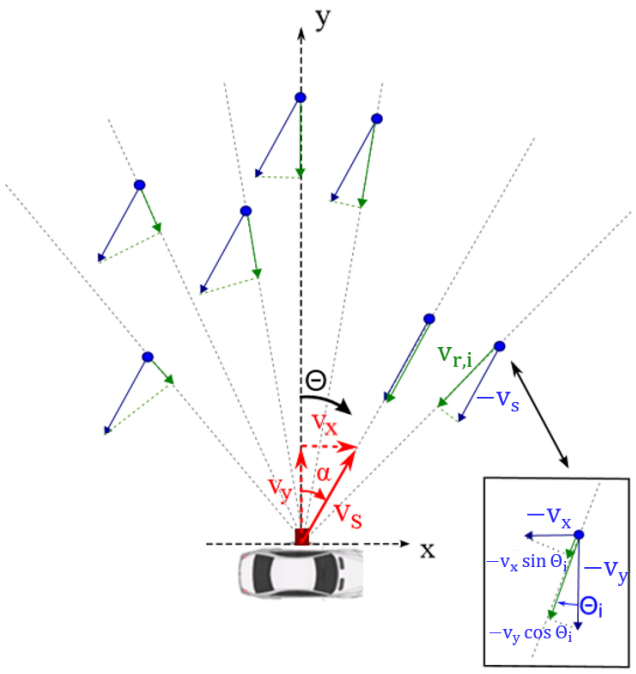

Figure 3.5: Illustration for radar odometry algorithm [1].

extraction of inliers (corresponding to stationary targets) that fit model Eq. (3.6) very well and separation of outliers (moving targets or clutter) by random sampling of observed data [89].

3.4 System Parameter Requirement

3.4.1 Pulse Repetition Frequency (PRF)

The along-track sampling/update rate is a significant limitation of SAR imaging. Here, we use the PRF to represent the sampling rate. The PRF must be greater than the azimuth signal bandwidth to avoid azimuth resolution ambiguities caused by aliasing [90]. By the Nyquist sampling theorem, it is usual to ensure that the SAR element spacing Δu is less than $\lambda/2$ to avoid the aliasing lobes [91, 92]. For the moving radar with velocity v_s, wavelength λ, PRF f_p, the azimuth sampling constraint can be expressed as: $v_s \cdot \frac{1}{f_p} < \frac{\lambda}{2}$. This requires either increasing the PRF (reduce K in Fig. 3.1) or limiting the maximum velocity of source

radar.

For the proposed MIMO-SAR algorithm, the above azimuth sampling requirement might be relaxed by considering the potential aliases only within a small region (as shown in Fig. 3.3). From [91], the minimum required sampling space Δu for imaging the defined small region that has azimuth scope $\Delta\theta$ without alias is given by $\Delta u < \lambda / \left(2 \sin \frac{\Delta\theta}{2}\right)$. Hence, we can obtain the sampling constraint for moving radar with MIMO-SAR:

$$v_{\mathrm{s}} \cdot \frac{1}{f_{\mathrm{p}}} < \frac{\lambda}{2 \sin \frac{\Delta\theta}{2}} \tag{3.8}$$

For example, the along-track sampling for 77 GHz radar should be finer than 2 mm. Given $f_{\mathrm{p}} = 556\,\mathrm{Hz}$ (SAR updating every 1.8 ms, $K = 20$ in Fig. 3.1), radar must travel no faster than 1.1 m/s, which is unrealistic for automotive scenario. Limiting to the azimuth scope $\Delta\theta = 5°$ for each small region, the radar should move no faster than 25.1 m/s. In practice, the chosen PRF is a little higher than the minimum requirement for oversampling.

3.4.2 Coherent Processing Interval (CPI)

The CPI for SAR processing is defined as the total duration over which the received echo can be coherently processed. We assume that the coherent phase course can no longer be guaranteed when the matched filter response Eq. (3.4) has a phase error greater than threshold $\pi/2$, resulting in a loss of image quality. We analyze two sources of the phase error in Eq. (3.4): system clock error and velocity estimation error from ego-motion.

The matched filter response is $\exp\left(j 2\pi f_{\mathrm{c}} \frac{2d}{c_0}\right)$, where the inaccurate carrier frequency f_{c} and distance d can result in the phase error. If the system clock provided by the local oscillator has an error, there will be a deviation on both f_{c} and the measurement timing (therefore affecting d). According to Eq. (3.3) Eq. (3.7), any velocity error in estimating radar's ego motion will also affect d. Based on [93], the maximum clock error for TI's on-board 40 MHz oscillator is 200 ppm, which is very small compared to the potential velocity error. We therefore only consider the effect of velocity error on CPI in the following analysis.

We assume the velocity errors $\epsilon_{v_{\mathrm{x}}}$, $\epsilon_{v_{\mathrm{y}}}$ for any frame are independent Gaussian variables, i.e., $\epsilon_{v_{\mathrm{x}}} \sim \mathcal{N}(0, \sigma^2)$, $\epsilon_{v_{\mathrm{y}}} \sim \mathcal{N}(0, \sigma^2)$. Then the distance deviation Δd at frame p is the

accumulation of the previous velocity error components projected to the radial direction θ_i.

$$\Delta d_p = \left| \sum_{i=0}^{p-1} \left(\cos\theta_i \cdot \epsilon_{v_{\mathrm{x}},i} + \sin\theta_i \cdot \epsilon_{v_{\mathrm{y}},i} \right) \right| \cdot T_{\mathrm{f}} \tag{3.9}$$

Since $\epsilon_{v_{\mathrm{x}},i}, \epsilon_{v_{\mathrm{y}},i}$ are all independent Gaussian variables, we have:

$$\begin{aligned} \cos\theta_i \cdot \epsilon_{v_{\mathrm{x}},i} + \sin\theta_i \cdot \epsilon_{v_{\mathrm{y}},i} &\sim \mathcal{N}(0, \sigma^2) \\ \sum_{i=0}^{p-1} \left(\cos\theta_i \cdot \epsilon_{v_{\mathrm{x}},i} + \sin\theta_i \cdot \epsilon_{v_{\mathrm{y}},i} \right) &\sim \mathcal{N}(0, p\sigma^2) \end{aligned} \tag{3.10}$$

Therefore, Δd_p follows the folded normal distribution [94] with expectation $\mathbb{E}\left[\Delta d_p\right] = \sqrt{\frac{2p}{\pi}}\sigma T_{\mathrm{f}}$. Based on Eq. (3.4), the relative phase error caused by $\mathbb{E}\left[\Delta d_p\right]$ can be expressed as $\Delta\phi_p = 4\pi f_{\mathrm{c}}\,\mathbb{E}\left[\Delta d_p\right]/c_0$. When $\Delta\phi_p$ exceeds a threshold ϕ_{Th}, we define the limits of CPI. Thus, the number of frames N_{f} within CPI is given by:

$$N_{\mathrm{f}} = \frac{1}{2\pi}\left(\frac{c_0 \cdot \phi_{\mathrm{Th}}}{4 f_{\mathrm{c}} \sigma T_{\mathrm{f}}}\right)^2 \tag{3.11}$$

For example, given $\phi_{\mathrm{Th}} = \pi/2$, $f_{\mathrm{c}} = 77\,\mathrm{GHz}$, $T_{\mathrm{f}} = 33.3\,\mathrm{ms}$, $\sigma = 0.005$, the CPI $N_{\mathrm{f}} \approx 14$ frames.

3.4.3 Region of Interest (ROI)

The definition of ROI in Section 3.3 is important for balancing computational efficiency and overall imaging quality. Each ROI square is defined by two parameters: side length along y-axis Δy and azimuth scope $\Delta\theta$. With smaller Δy and $\Delta\theta$, the ROI covers a smaller area for imaging, thus resulting in a lower computation load. Besides, Equation Eq. (3.8) implies that smaller $\Delta\theta$ gives a lower PRF requirement as well for avoiding azimuth aliasing in ROI. However, it is possible to miss interested targets in imaging results if we choose too small Δy or $\Delta\theta$, i.e., when the resulting ROI is not big enough to cover the undetected targets of the CFAR algorithm. In the following experiments, we use a sensitive CFAR detector with a probability of false alarm 10^{-4} and select medium-size ROI with $\Delta y = 0.9\,\mathrm{m}$ and $\Delta\theta = 5^\circ$ accordingly to achieve the necessary balance.

---

[4] T_{c} is equal to single chirp duration times the number of Tx (i.e., N_{Tx}) under TDM MIMO configuration.

Table 3.1: system parameter calculation (based on [25]) and configurations for mimo-sar processing

Parameter	Calculation Equation	Configuration	Value
Range resolution (R_{res})	$R_{\text{res}} = \frac{c_0}{2B} = 0.447\,\text{m}$	Frequency (f_{c})	77 GHz
Velocity resolution (V_{res})	$V_{\text{res}} = \frac{\lambda}{2N_{\text{c}}T_{\text{c}}} = 0.0848\,\text{m/s}$	Sweep Bandwidth (B)	335 MHz
Angle resolution (θ_{res})	$\theta_{\text{res}} = \frac{\lambda}{N_{\text{Tx}}N_{\text{Rx}}h\cos\theta} \approx 15^\circ$	Sweep slope (S_{w})	21 MHz/µs
Max operating range (R_{max})	$R_{\text{max}} = \frac{f_{\text{s}}c_0}{2S_{\text{w}}} = 28.5\,\text{m}$	Sampling frequency (f_{s})	4000 ksps
Max operating velocity (V_{max})	$V_{\text{max}} = \frac{\lambda}{4T_{\text{c}}} = 10.82\,\text{m/s}$	Num of chirps in one frame (N_{c})	255
Max operating angle (θ_{max})	$\theta_{\text{max}} = \sin^{-1}\left(\frac{\lambda}{2h}\right) = 90^\circ$	Num of samples of one chirp (N_{s})	64
		Num of transmitters, receivers (N_{Tx}, N_{Rx})	2, 4
		Duration of chirp[4] and frame (T_{c}, T_{f})	90 µs, 33.3 ms

3.5 Simulations

3.5.1 Imaging Point Targets

Simulation Studies via MATLAB: Assume two static point targets in the field of view at the same range (5 m) but at angles $(-0.5^\circ, 0.5^\circ)$ that are within a current angular beam-width (hence cannot be separated by conventional 3D-FFT processing). The initial location of the radar sensor is $(0, 0)$, and it is assumed to move with velocity $\left(v_{\text{x}}, v_{\text{y}}\right) = \left(1\,\text{m/s}, 0\right)$. Other system configurations for radar are presented in Table 3.1. We use MATLAB to simulate the TDM-MIMO radar with the I-Q samples post-demodulated at the receiver.

Range-Angle Imaging for MIMO Radar: To compare the imaging capability of different algorithms, we show the range-angle image [25] for MIMO radar data. The range-angle maps are obtained by operating Range and Angle FFTs on the simulated I-Q samples of the first frame. We then average the amplitude range-angle maps of all chirps to get the final image. As shown in Fig. 3.6(a), the generated range-angle map presents the imaging for two point-targets with very close azimuth angles. Results tell that two point-targets can not be separated with the current MIMO beam width (around 15° resolution).

MIMO-SAR Imaging: Using the MIMO operation (CFAR detection and peak grouping),

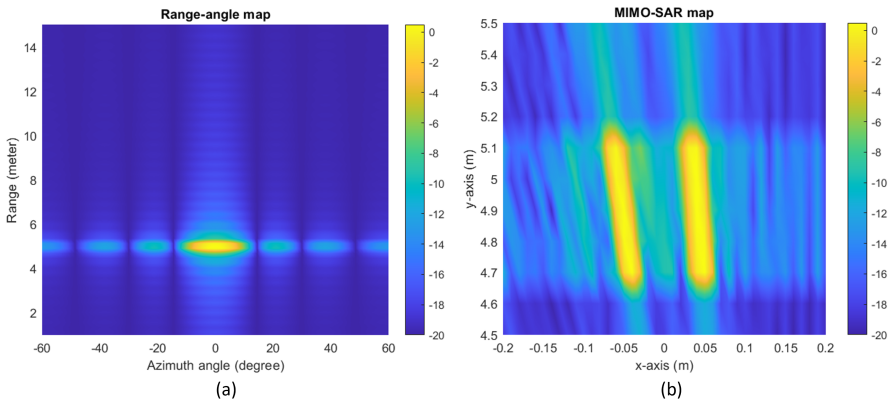

Figure 3.6: Imaging two close point targets: (a) Range-angle map imaging for the first frame; (b) MIMO-SAR imaging (magnitude I of Eq. (3.5)) for 13 frames.

Table 3.2: Parameter values for MIMO-SAR algorithm

Parameter	Value
Num of chirps for SAR processing (K)	20
Pixel size of SAR image (k_{x}, k_{y})	0.01 m, 0.1 m
y-axis side length for each ROI (Δy)	0.9 m
Azimuth scope for each ROI ($\Delta\theta$)	5°
Range FFT points	64
Velocity FFT points for target detection	256
Velocity FFT points for SAR processing	20
Angle FFT points for target localization	128
Angle FFT points for SAR processing	16

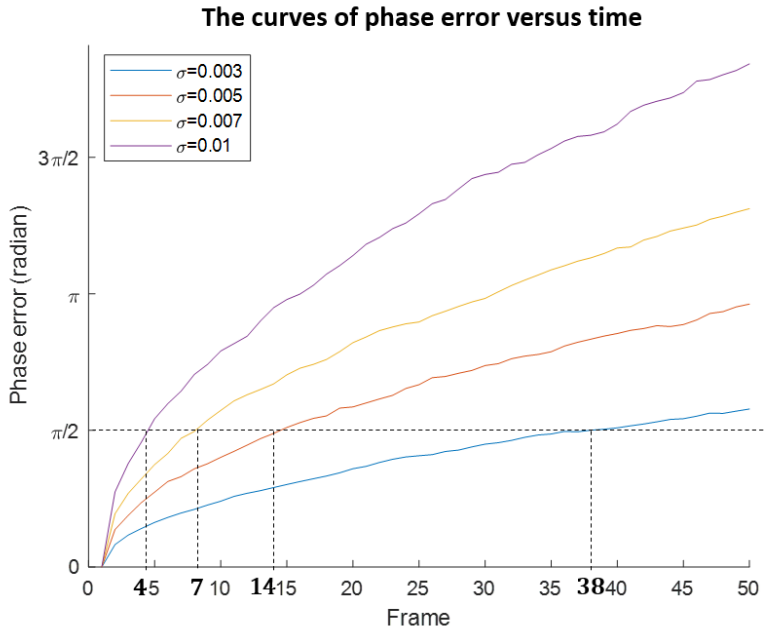

Figure 3.7: The curves of phase error versus time, for velocity error $\sigma = 0.003$, 0.005, 0.007, 0.01.

we select the ROI that covers two point-targets. The selected region covers the $-0.2\,\text{m} \sim 0.2\,\text{m}$ (4.6° coverage) in x-axis and $4.5\,\text{m} \sim 5.5\,\text{m}$ in y-axis. The spacing of the x-axis and y-axis are 0.01 m and 0.1 m respectively. We assume that radar velocity is known at the receiver in this simulation, and perform the Back-projection SAR processing on each pixel of ROI for 13 frames. The parameters for MIMO-SAR are set according to Table 3.2. As shown in Fig. 3.6(b), the resulting MIMO-SAR image shows a clear separation between two nearby point targets in azimuth, thereby validating that the MIMO-SAR imaging algorithm can effectively increase azimuth angle resolution, and achieve separation of close point targets.

3.5.2 Verifying Predicted CPI for MIMO-SAR

From Eq. (3.11), the CPI for SAR processing is mainly determined by the root-mean-square velocity error σ of radar's ego-motion. With increasing time, the accumulated velocity error causes an increase in phase error in the matched filter response Eq. (3.4), thus limiting the

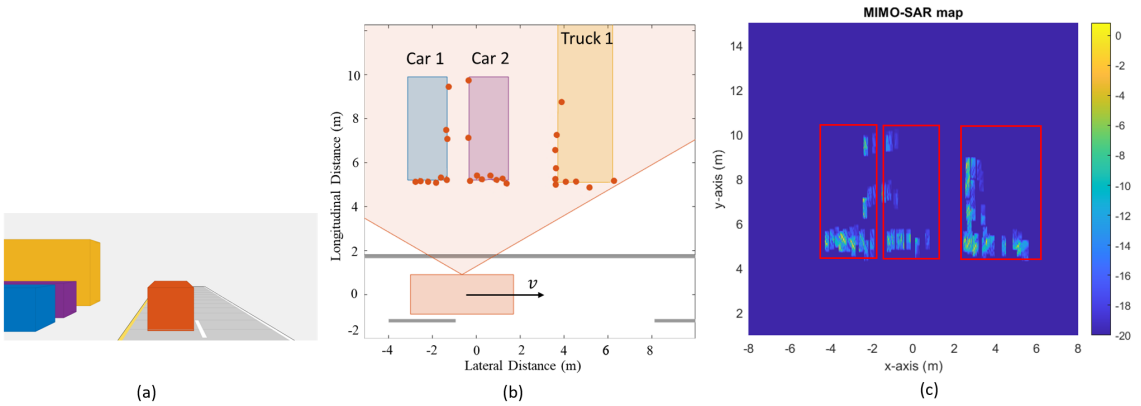

Figure 3.8: MIMO-SAR simulation for the parking lot scenario: (a) The ego-centric view of the environment with one driving car, two parked cars, and 1 truck; (b) The bird-view for the parking lot environment; The red points are the point-representation of the extended cars and truck; (c) MIMO-SAR imaging result (magnitude I of Eq. (3.5)) for parking lot scenario where we use three rectangles to cover the imaging cars and truck.

CPI. To validate the predicted CPI, we reuse the simulation environment in Section 3.5.1, add r.m.s. ego-motion velocity error σ, and evaluate the phase error of a pixel in ROI versus time. This simulation was averaged 1000 times to plot the mean phase error vs. time curves for different velocity error values.

We show the curves for velocity error 0.003, 0.005, 0.007, and 0.01 respectively in Fig. 3.7, and mark the frame points that attain the phase error threshold $\pi/2$. For example, given $\sigma = 0.003$, the calculated CPI from Eq. (3.11) is 38 frames. It shows that in the blue curve of Fig. 3.7, the phase error of frame 38 is nearest to $\pi/2$. Similarly, we validate that the CPIs for velocity error 0.005, 0.007, 0.01 are 14, 7, 4 frames respectively.

3.5.3 MIMO-SAR Imaging for Simulated Parking Lot Scenario

Simulation Environment: We simulate the driving environment for a parking lot with the MATLAB Automated Driving Toolbox [95] to validate the MIMO-SAR algorithm. As shown in Fig. 3.8(a), this simulation environment contains 1 (self) driving car, 2 parked cars, and 1

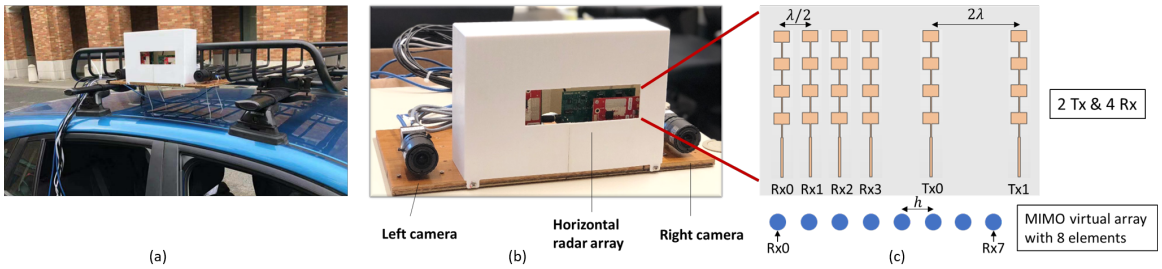

Figure 3.9: (a) Integrated radar-camera data capture platform mounted on one vehicle; (b) FLIR cameras and TI radar board with 1-D horizontal array; (c) Horizontal antenna array includes 2 Tx with distance 2λ and 4 Rx with distance $\lambda/2$, yielding virtual MIMO array with 8 Rx elements.

parked truck. The driving car is with constant moving velocity 4 m/s, and integrated with a radar sensor on its left side to observe the parked vehicles. Each extended vehicle is modeled as a group of point targets (see Fig. 3.8(b)) in the radar's field of view. We then simulate the radar I-Q samples for these point targets with parameters configured as Table. 3.1.

MIMO-SAR Imaging: We perform the MIMO-SAR algorithm on simulated I-Q samples assuming that the moving velocity of radar is unknown. That is, before operating the Back-projection SAR, we use the radar odometry to estimate the radar trajectory. The parameters for MIMO-SAR are set according to Table. 3.2. The MIMO-SAR imaging result for the data from 3 frames is shown in Fig. 3.8(c) where we use three rectangles to highlight the parked cars and 1 truck. The synthetic aperture for 3 frames is around 0.4 m ($4\,\text{m/s} \times 3 \times 33.3\,\text{ms} = 0.4\,\text{m}$), which is much larger than the intrinsic MIMO aperture ($4\lambda \approx 16\,\text{mm}$). It turns out that MIMO-SAR is effective for imaging the boundary of extended objects and separating close objects with large synthetic apertures.

3.6 Experimental Results

3.6.1 Experiment Platform Setup

A set of camera images along with corresponding radar raw data (I-Q samples post-demodulated at the receiver) [36] have been collected by a vehicle-mounted platform (see

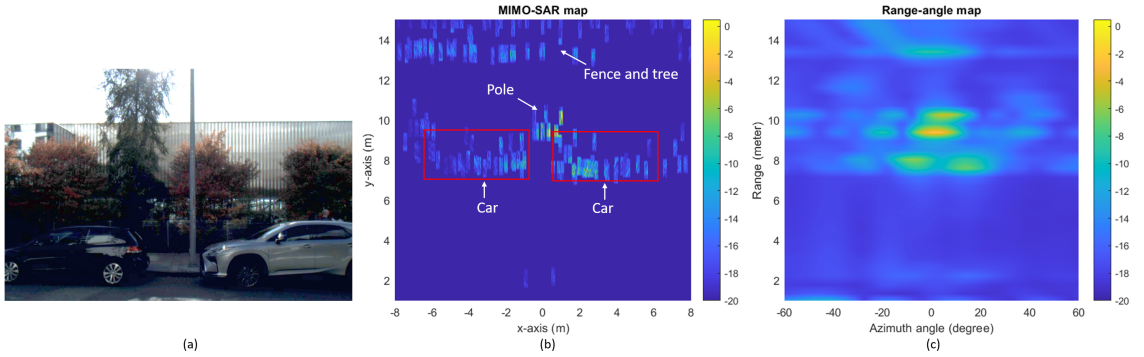

Figure 3.10: MIMO-SAR imaging for roadside experiment 1: (a) The camera image for the imaging environment with two parallelly parked cars; (b) MIMO-SAR imaging result (magnitude I of Eq. (3.5)) where we use two rectangles to cover the parked cars; (c) Range-angle map imaging for single-frame radar data.

Fig. 3.9(a)). The platform is placed on one side to facilitate sideways data collection whereby roadside objects are illuminated. The data collection platform - see Fig. 3.9(b) - consists of 2 FLIR cameras (left and right) and a TI AWR1843 EVM radar [64]. As shown in Fig. 3.9(c), the AWR1843 radar is integrated with 2 Tx and 4 Rx[5], which forms an 8-element horizontal virtual array with TDM-MIMO. We configure the parameters of this radar according to Table. 3.1. The binocular cameras are synchronized with radar to provide visualization for imaging scenarios. Since there is no external navigation sensor (IMU or GPS) on the platform, we include radar odometry with RANSAC for all experiments below to estimate the necessary radar motion trajectory.

3.6.2 Experiments

Roadside Experiment 1: The scene for the first experiment contains 2 parallel parked cars and the background (fence and tree), as shown in Fig. 3.10(a). Similar to the above, the MIMO-SAR algorithm is operated on the collected radar I-Q samples lasting for 3 frames. We present the MIMO-SAR map in Fig. 3.10(b) where the pole and two closely parked cars

[5]The 3rd elevation transmitter on AWR1843 radar board is not used.

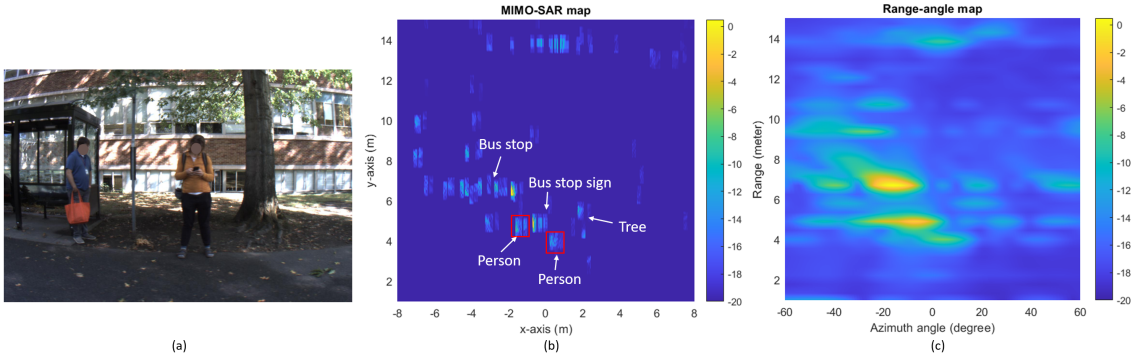

Figure 3.11: MIMO-SAR imaging for roadside experiment 2: (a) The camera image for the imaging environment with two static people; (b) MIMO-SAR imaging result (magnitude I of Eq. (3.5)) where we use two rectangles to cover the standing persons; (c) Range-angle map imaging for single-frame radar data.

are well-separated.

Roadside Experiment 2: The second experiment shown in Fig. 3.11(a) is going to image 2 standing persons and background (bus stop and tree). Its MIMO-SAR imaging result from 3 frames is presented in Fig. 3.11(b), from which we can identify the persons from the strong-reflection background (i.e., bus stop and the sign between two persons).

The above experiments validate the effectiveness of the MIMO-SAR algorithm for imaging closely located objects in the foreground (extended vehicles and standing persons) among a sophisticated background (building, tree, bus stop, fence, etc). To show the performance gain provided by a large synthetic aperture, we compare each MIMO-SAR imaging result with its range-angle map counterpart. The range-angle maps - see Fig. 3.10(c), and Fig. 3.11(c) - are obtained from the I-Q samples of a single frame, using the method in Section 3.5.1. Based on the analysis of antenna aperture size in Section 3.5.3, the synthetic aperture of vehicular-based MIMO-SAR for 3 frames (~ 100 ms) is many times larger than the intrinsic MIMO aperture, resulting in much greater azimuth angle resolution than conventional range-angle imaging.

Table 3.3: computational complexity analysis

Method	FLOP	Run time per frame
MIMO-SAR algorithm	7.7326×10^7	1.17 s
Back-projection algorithm [73]	7.1532×10^9	134.19 s
Range-angle imaging	1.4894×10^7	0.23 s

3.7 Discussion

3.7.1 Computation Complexity Analysis

We compare the computational complexity of different radar imaging algorithms using two metrics: the number of float point operations (FLOP) and running time. We run algorithms for the roadside experiment 1 to calculate their FLOPs and averaged running time per frame. For the MIMO-SAR algorithm, we don't take the complexity of radar odometry into account for a fair comparison. The comparison baselines are the range-angle imaging indicated above, and the Back-projection algorithm - a traditional SAR imaging algorithm that only takes the radar I-Q samples of one receiver as input [73]. We let the back-projection algorithm update the whole SAR image at every chirp to satisfy the azimuth sampling constraint. We present the calculated results for three algorithms in Table. 3.3. First, from the comparison between MIMO-SAR and range-angle imaging, we conclude that the performance gain of MIMO-SAR comes at the expense of around 5 times heavier computation load. Second, it tells that the MIMO-SAR algorithm needs almost **93 times fewer** FLOPs compared to the Back-projection algorithm [73]. While performing on the MATLAB R2019b with an Intel i7-9750H CPU, MIMO-SAR needs 1.17 s running time per frame that is **115 times faster** than the back-projection algorithm. The above results confirm our claimed statement - the hierarchical MIMO-SAR algorithm drastically reduces the computational complexity.

The running time for the MIMO-SAR algorithm (1.17 s per frame) can be *further reduced* to achieve real-time capability. Since the Back-projection processing is suitable for easy

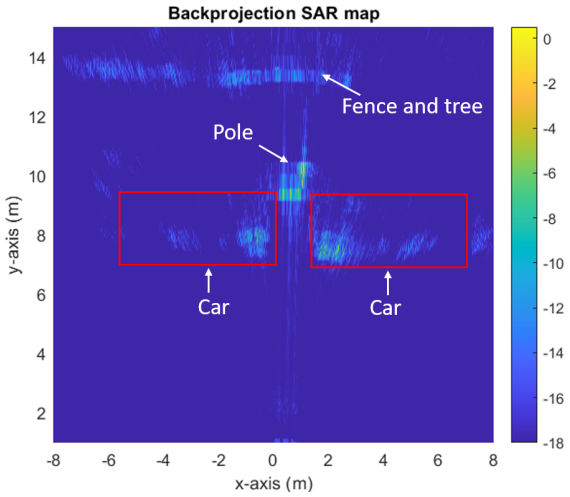

Figure 3.12: Back-projection SAR imaging result using 3 frames for roadside experiment 1 in Fig. 3.10.

parallel implementation[6] utilizing GPUs [73, 96], the calculation time for MIMO-SAR can be efficiently distributed. Besides, some adjustments of parameter values, such as increasing the pixel size k_{x}, k_{y} of SAR image, and decreasing the side length Δx, Δy of the interested region, would reduce the computation load linearly, which allows a trade-off to achieve real-time capability at the cost of image quality degradation.

3.7.2 Strengths and Limitations

To summarize, the proposed MIMO-SAR algorithm enables high azimuth resolution imaging for closely located objects in the foreground with much-reduced algorithm run time. Experiments in Section 3.6 show that vehicular-based MIMO-SAR can synthesize a large aperture by exploiting vehicle motion, resulting in better azimuth angle resolution and separation ability (of close targets) than conventional range-angle imaging. Meanwhile, the proposed hierarchical MIMO-SAR algorithm requires 1.17 s average run time per frame in Experiment 1, which is 115 times faster compared to its counterpart Back-projection algorithm. Note

[6]The calculations are performed pixel-wise on the SAR image, so they can be distributed in GPUs. In addition, the calculations for different snapshots can be done parallelly before the coherent summation.

that MIMO-SAR preserves the high-resolution imaging quality while reducing computation load, since we exploit all available information on ROIs, i.e., using the virtual aperture created by MIMO, and condensing multiple chirps for each update. To illustrate it, we plot the Back-projection SAR imaging result for roadside experiment 1 and show it in Fig. 3.12 for comparison with the corresponding MIMO-SAR map in Fig. 3.10(b). Results indicate that the MIMO-SAR map enables to visually separate the cars and poles just as the Back-projection SAR map, though with some losses to the image SNR due to lower PRF.

One limitation of the MIMO-SAR algorithm is that it is modeled for stationary targets only and we can not apply it directly for moving targets such as pedestrians and other vehicles. However, it is possible to address this problem by introducing ISAR concepts [76, 77] to MIMO-SAR that incorporates the target motion.

3.8 Summary

In this chapter, we proposed a new MIMO-SAR algorithm for FMCW automotive radar to get high-resolution imaging. MIMO-SAR incorporates a MIMO processing stage to narrow down the ROI for subsequent finer-resolution SAR processing, which drastically reduces the computation load. Besides, we integrated a radar odometry algorithm to estimate the requisite ego-radar trajectory without relying on additional IMU sensors. We validated the MIMO-SAR by both simulations (via MATLAB) and experiments using a vehicle-mounted side-view radar. For future work, we will continue to explore better ways of radar imaging and make the high-resolution radar images apply to practical use.

Chapter 4

STATIC BACKGROUND REMOVAL IN VEHICULAR RADAR

4.1 Motivation

Autonomous driving systems that rely on multi-sensor fusion and scene perception are key to achieving future L4 and L5-level vehicular automation [30, 36, 37, 97, 98]. However, image understanding in sophisticated environments such as city roads with dense multi-purpose traffic, remains a significant current bottleneck. In these scenarios which contain both moving and static objects/backgrounds, the removal of static objects is a well-known approach to enhance moving target indication (MTI) [99–102].

MTI is a mature technology derived from air-borne radar systems [99] that has been used to detect and track moving targets while filtering out clutter - the unwanted echoes from the background and other stationary objects. To improve MTI in air-to-air and air-to-ground surveillance scenarios in the presence of clutter and other interference, space-time adaptive processing (STAP) is a long-established signal processing technique [99] that utilizes 2-dimensional joint adaptive filtering in the spatial and temporal domains to maximize the signal-to-interference ratio (SIR) [100]. Optimal filters for STAP-based MTI generally require prior knowledge of clutter statistics of the cell under test [101]. Adaptive methods such as sample matrix inversion (SMI) iteratively estimate the true interference covariance matrix based on received sample inputs [102]. It has been shown that SMI requires twice the degree of freedom (DOF) training samples to ensure adequate SIR and requires training samples containing only clutter that share the same statistics as the clutter in cell-under-test (i.e., implicitly assume homogeneous environments), and are independent and identically distributed (IID) [103]. To address the challenges of acquiring training data volume and computation complexity in computing the clutter covariance matrix for real-time operations, various pre and post-Doppler STAP methods have been developed [104–108]. These explore the trade-off between sub-optimal filtering with lower complexity and the

Table 4.1: Discussion of STAP Methods

Methods	Assumption	Processing	DOF	Pros & Cons
optimal STAP [109]	clutter echoes present in all array channels and pulses.	STAP on all array channels and pulses	$N_c N_h$	optimal filtering, heavy computation
adjacent-bin post-Doppler STAP [110]	Post Doppler processing, clutter echoes occupy only a few Doppler bins while present in all pulses.	Initial Doppler processing, and STAP on array channels and remaining adjacent Doppler bins	$N_h L$	suboptimal filtering, reduced computation
joint domain localized STAP [111]	Post Doppler processing and beamforming, clutter echoes occupy only a few Doppler and angle bins.	Initial Doppler processing + beamforming, and STAP on remaining adjacent angle bins and Doppler bins	HL	suboptimal filtering, further reduced computation

resulting degradation of clutter suppression performance.

4.1.1 Prior Art: STAP Approaches

The displaced phase center antenna (DPCA) is a deterministic STAP method [106] suited for synthetic aperture radar (SAR) scenarios, that requires specific conditions to hold. It relies on identical channels between TX locations and target, with antenna phase centers displaced precisely along the flight axis, and a pulse repetition frequency that matches the platform velocity [108]. By utilizing a shifted array aperture to compensate for platform motion, DPCA removes clutter through the subtraction of returns corresponding to consecutive

pulses. However, meeting these requirements in practice can be challenging due to channel or phase errors. In addition to DPCA, there are other STAP algorithms that aim to reduce computation by reducing the total DOF (*product of spatial and temporal DOF*). Among them, post-Doppler methods efficiently reduce the temporal DOF by performing Doppler processing before applying STAP [109]. One such method is the adjacent-bin post-Doppler STAP [110], which adaptively processes the spatial returns from adjacent Doppler bins. This approach divides the $N_c N_h$-DOF space-time filtering problem into N_c separate $N_h L$-DOF space-time filtering problems, where N_c, N_h, L are dimensions of time, space, and the number of adjacent Doppler bins ($L \ll N_c$), respectively[1]. Another technique is joint domain localized STAP [111], which incorporates additional beamforming in the spatial domain and applies STAP on adjacent Doppler and angle bins. This method divides the $N_c\, N_h$-DOF space-time filtering problem into $N_c N_h$ separate HL-DOF space-time filtering problems, where H is the number of adjacent angle bins selected ($H \ll N_h$), and we assume the beamforming does not change the spatial dimension. A systematic discussion of various STAP methods in terms of their processing procedures, assumptions, resulting DOF, and pros & cons are presented in Table. 4.1.

The STAP methods have also found widespread application in ground-penetrating and through-the-wall radars, where background removal is commonly referred to as the process of removing ground-reflected clutter and walls respectively to reveal masked targets [112]. Various techniques have been employed including coherent background subtraction, mean subtraction [113,114], frame differencing [115], and component selection using singular-value-decomposition (SVD) [116] and principal component analysis (PCA). Coherent background subtraction assumes knowledge of the wall characteristics and requires access to a reference scene, which can be challenging to obtain in automotive applications. On the other hand, frame differencing and mean subtraction take advantage of the clutter's time and angle invariance. These methods calculate the clutter signal by averaging over time and aspect angle to mitigate its effects. These revisions aim to enhance clarity, conciseness, and cohesiveness while preserving the original content and reference format. Recent advancements in deep

[1]We assume Doppler processing/beamforming does not change the size of the time/space dimension in the DOF analysis.

learning (DL) have also contributed to the field, providing new ideas for radar image processing in automotive radars [25, 36, 117–119], as well as clutter suppression in airborne radars [103]. In [103], a novel DL framework is proposed to suppress clutter in the angle-Doppler spectrum, addressing the challenges of insufficient training samples and low detection probability in non-homogeneous clutter environments.

4.1.2 Issues with Existing Methods

The above 'traditional' methods are not readily applicable to vehicular radars for several reasons. First, unlike the air-borne scenario, the automotive environment is non-homogeneous, meaning that the type of objects or background can vary across the field of view. This poses a challenge for STAP as it may not have enough qualified training samples that are IID (i.e., twice the system's DOF from adjacent range bins), resulting in significant performance loss. Second, the motion of the source vehicle leads to rapidly changing clutter statistics for both the ground and stationary targets (trees, buildings, etc.). Therefore, the clutter removal algorithms designed for ground-penetrating radar or through-the-wall radar are not directly applicable to vehicular radar. Third, vehicular radars require real-time processing for environment perception implying that algorithms with heavy computational requirements such as STAP, cannot be directly deployed in-situ on vehicles and will need modifications.

4.1.3 Contributions

Recent advancements in 77 GHz FMCW radars have demonstrated highly accurate object detection and localization capabilities, regardless of environmental conditions [25, 27, 36, 120]. In particular, these radars excel in fine-resolution Doppler velocity measurement [25], presenting an opportunity for background removal in automotive scenarios. The Doppler velocities of static objects are determined by their azimuth and elevation angles as well as the radar's instantaneous velocity and direction of motion [37]. This allows for the localization of background objects by identifying their "specific Doppler velocity profile" and removal via filtering in the azimuth-elevation-Doppler domain. This approach offers several advantages over temporal and spatial domain filtering methods when dealing with the challenges posed

by automotive scenarios. It does not rely on any training data and is applicable to both moving and stationary radar; finally (and perhaps critically) it operates efficiently without the need for heavy STAP-like covariance matrix computations.

While our proposed method also applies Doppler processing and beamforming prior to background removal, there is a fundamental difference vis-a-vis classic STAP methods. STAP methods estimate the clutter covariance matrix from space-time samples to obtain the filter weights that maximize the SIR for specific Doppler-angle pair. The Doppler processing and beamforming steps in post-Doppler STAPs concentrate the clutter in a reduced number of Doppler/angle bins, effectively reducing the DOF for clutter covariance estimation. In contrast, our proposed method directly suppresses background clutter by identifying the Doppler frequency corresponding to each azimuth-elevation angle and applying notch filtering to remove the clutter components from the input. The fine Doppler velocity resolution of vehicular radar gives more accurate Doppler estimation than before, which makes it possible to locate and notch the clutters from the Doppler domain without applying STAP. Thus, the proposed method avoids the computation of the clutter covariance matrix, allowing for faster and more efficient processing.

In this chapter, the presented background removal algorithm for FMCW-based MIMO radar consists of a radar signal preprocessing step followed by 3D ego-motion estimation, and notch filter-based background removal in the azimuth-elevation-Doppler domain. The contributions of our work can be summarized as follows:

- We develop a model for the received signal of FMCW MIMO radar with time-division multiplexing specifically for point targets in automotive scenarios. Using this model, we propose a signal processing framework that extracts 4D point clouds from the raw radar signal.

- We propose a 3D ego-motion estimation algorithm that takes radar point clouds as input and accurately estimates the radar ego velocity along the x, y, z axes. Our algorithm incorporates the consideration of measurement error and tackles the issue of Doppler ambiguity that can arise in massive MIMO systems, offering a robust solution.

- We propose a background removal algorithm that identifies Doppler frequencies corresponding to the background via the "specific Doppler velocity profile" and removes the background via notch filtering in the azimuth-elevation-Doppler domain.

- To evaluate algorithm performance, we conduct tests using simulated data from the MATLAB Automated Driving toolbox. Additionally, we validate our approach through extensive experiments with real-world data collected from Texas Instrument (TI)'s cascaded-chip radar board under practical driving scenarios.

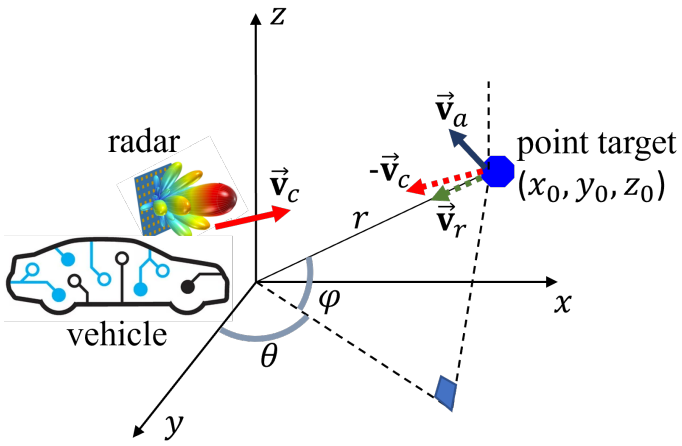

Figure 4.1: Model scenario - a moving radar and a moving point target in Global Cartesian coordinates for a single frame at $t = 0$, with radar at origin and target at (x_0, y_0, z_0). Radar (target) moves with velocity $\vec{\mathbf{v}}_c$ ($\vec{\mathbf{v}}_a$), respectively. The target exhibits a relative Doppler velocity $\vec{\mathbf{v}}_r$ with respect to the radar over this frame.

4.2 Signal Model

In this section, we model one frame of the FMCW radar return signal for a general point target in autonomous scenarios. We assume that each radar TX transmits a sequence of N_c chirps with duration T_c in a frame. With carrier frequency f_c and chirp slope S_w, amplitude

A_{T}, and initial phase ϕ_0, the transmit signal of the FMCW radar during time t within a frame is given by $s_{\mathrm{T}}(t) = A_{\mathrm{T}} \cos\left(2\pi\left(f_{\mathrm{c}}t + \frac{1}{2}S_{\mathrm{w}}t^2\right) + \phi_0\right)$.

For target at distance r from radar, the received reflected signal $s_{\mathrm{R}}(t)$ incurs round-trip delay $\tau = 2r/c_0$, i.e., $s_{\mathrm{R}}(t) = \frac{A_{\mathrm{R}}}{A_{\mathrm{T}}} s_{\mathrm{T}}(t-\tau)$, where c_0 is the speed of light and A_{R} is the received signal amplitude. The received signal is mixed with the transmit signal at the receiver to produce the difference intermediate frequency (IF) signal $s_{\mathrm{IF}}(t)$:

$$s_{\mathrm{IF}}(t) = \frac{A_{\mathrm{T}}A_{\mathrm{R}}}{2}\left\{\cos\left[2\pi\left(S_{\mathrm{w}}\tau t + f_{\mathrm{c}}\tau - \frac{1}{2}S_{\mathrm{w}}\tau^2\right)\right]\right\} \tag{4.1}$$

In the global coordinate system of Fig. 4.1, the vehicle-mounted radar moves from the origin at $t = 0$ with ego velocity $\vec{\mathbf{v}}_c(t) = (v_x(t), v_y(t), v_z(t))$ at time t. For frame-by-frame modeling, we assume that the velocities of the radar and any targets may be assumed constant over a frame duration (typically milliseconds). Therefore, we simplify the time-dependant variables $\vec{\mathbf{v}}_c(t)$ by $\vec{\mathbf{v}}_c$ etc. in the following analysis. Fig. 4.1 shows a point target located at a range r, azimuth angle θ, and elevation angle φ (corresponding to Cartesian-coordinates location $(x_0, y_0, z_0) = (r\cos\varphi\sin\theta, r\cos\varphi\cos\theta, r\sin\varphi)$), moving with velocity $\vec{\mathbf{v}}_a = (v_{a,x}, v_{a,y}, v_{a,z})$ in one frame. The target exhibits a relative Doppler velocity (the velocity along the radial direction) $\vec{\mathbf{v}}_r$ with respective to the radar; the amplitude of $\vec{\mathbf{v}}_r$ (denoted by v_r) is obtained by projecting the inverse platform velocity $-\vec{\mathbf{v}}_c$ and target velocity $\vec{\mathbf{v}}_a$ onto the radial direction as follows:

$$v_r = \big((v_{a,y} - v_y)\cos\theta + (v_{a,x} - v_x)\sin\theta\big)\cos\varphi + (v_{a,z} - v_z)\sin\varphi \tag{4.2}$$

If $\vec{\mathbf{v}}_a = \mathbf{0}$ (target is stationary), the relative Doppler velocity amplitude is represented by:

$$v_r = -(v_y\cos\theta + v_x\sin\theta)\cos\varphi - v_z\sin\varphi \tag{4.3}$$

FMCW radar sends a sequence of chirps within a frame, and the round-trip delay at each chirp varies (slightly) due to relative motion. By decomposing the t into fast time t_f (time within a chirp) and slow time n (the chirp index, i.e. time between chirps), i.e., $t = nT_c + t_f$, we can represent the round-trip delay for the target at n-th chirp τ_n using its Doppler velocity v_r and the round-trip delay for the first chirp $\tau_0 = 2\sqrt{x_0^2 + y_0^2 + z_0^2}/c_0$:

$$\tau_n = \tau_0 - \frac{2v_{\mathrm{r}}nT_{\mathrm{c}}}{c_0} \tag{4.4}$$

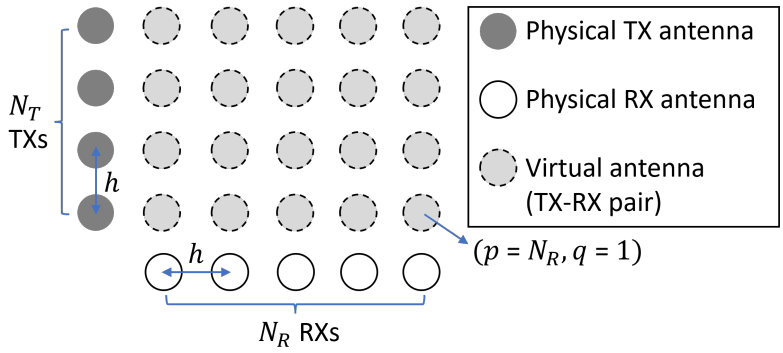

Figure 4.2: A TDM-MIMO setup with N_T TXs and N_R RXs and the resulting $N_T N_R$-element virtual array. Each element in the virtual array is formed by a TX-RX pair.

Substituting Eq. (4.4) into Eq. (4.1) and then sampling Eq. (4.1) in fast time t_f with frequency f_s, we can get the (post analog-to-digital conversion (ADC)) sampled IF signal with chirp index n and fast-time sample index m:

$$s_{\mathrm{IF}}(m,n) = \frac{A_\mathrm{T} A_\mathrm{R}}{2} \exp\left(j2\pi\left(S_\mathrm{w}\tau_n \frac{m}{f_\mathrm{s}} + f_\mathrm{c}\tau_n - \frac{1}{2}S_\mathrm{w}\tau_n^2\right)\right) \tag{4.5}$$

Under the time division multiplexed (TDM)-MIMO setup depicted in Fig. 4.2, we consider a radar system consisting of N_T transmitters (TXs) and N_R receivers (RXs), where TXs and RXs are arranged vertically and horizontally, separated by identical separation h. The TDM-MIMO scheme ensures the orthogonality of the transmit signals by sequentially transmitting chirps from each TX and allows recovery of the individual transmitted signals at the receive array. The measurements at the physical receive array corresponding to each orthogonal TX waveform can then be stacked as a matrix to effectively create a 'virtual' array [51]. For the setup in Fig. 4.2, the measurement at the (p,q)-th virtual array location is the signal transmitted from q-th TX and received by p-th RX, where $p \in \{1,2,\ldots,N_R\}$ and $q \in \{1,2,\ldots,N_T\}$. Thus TDM-MIMO operation results in a 2D array of $N_T N_R$ measurements, where the corresponding virtual MIMO antenna locations result from vector superposition of the TX and RX antenna locations [51].

From array theory [27], for a far-field target with azimuth angle θ and elevation angle φ, the IF signal at the (p,q)-th virtual array location contains the additional phase terms

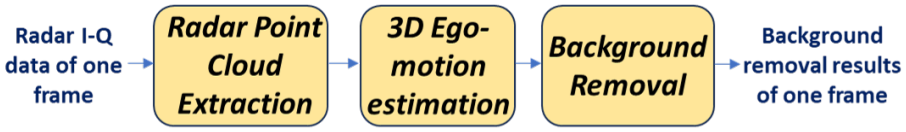

Figure 4.3: The overall workflow for the proposed system is operated in a frame-by-frame manner. The radar I-Q data for each frame is processed by the radar point cloud extraction module, 3D ego-motion estimation module, and background removal module to get the post-background removal results.

induced by the target's angles: $2\pi ph\cos\theta/\lambda$ and $2\pi qh\sin\varphi/\lambda$. In the TDM-MIMO setup, the duration of transmitting one chirp by N_T TXs is extended to $N_T T_c$ (which differs from the chirp duration T_c in the single TX case), resulting in a new round-trip delay τ_n' for the target during the n-th chirp:

$$\tau_n' = \tau_0 - \frac{2v_{\mathrm{r}} n N_T T_{\mathrm{c}}}{c_0} \tag{4.6}$$

In summary, the IF signal for chirp n, fast-time sample m, and (p,q)-th virtual antenna in TDM-MIMO setup is given by:

$$s_{\mathrm{IF}}(m,n,p,q) = \frac{A_{\mathrm{T}}A_{\mathrm{R}}}{2}\exp\left(j2\pi\left(S_{\mathrm{w}}\tau_n'\frac{m}{f_{\mathrm{s}}} + f_{\mathrm{c}}\tau_n' - \frac{1}{2}S_{\mathrm{w}}{\tau_n'}^2 + \frac{ph\cos\theta}{\lambda} + \frac{qh\sin\varphi}{\lambda}\right)\right) \tag{4.7}$$

4.3 System Design: Background Removal

The designed system for background removal in radar images is operated in a frame-by-frame manner (i.e., the data for each frame is processed separately), as shown in Fig. 4.3. It starts by processing the radar I-Q data $s_{\mathrm{IF}}(m,n,p,q)$ of one frame and extracting the point clouds with $(r, v_{\mathrm{r}}, \theta, \varphi)$ values as shown in Fig. 4.4. The extracted point clouds are used as input for 3D radar ego-motion estimation to solve the ego velocity (v_x, v_y, v_z) for that frame. To account for the presence of both stationary and moving targets in the point clouds, we incorporate the random sample consensus (RANSAC) [121] into the ego velocity estimation to separate the point clouds of stationary targets. Moreover, the proposed 3D ego-motion model takes into consideration the Doppler ambiguity and the measurement errors of Dopplers and

angles, enhancing the robustness and accuracy of the estimation. Once the radar ego velocity (v_x, v_y, v_z) is determined, a background removal algorithm that identifies Doppler frequencies corresponding to the background via the "specific Doppler velocity profile" and removes the background via notch filtering in the azimuth-elevation-Doppler domain is implemented on the radar images. In the following, we introduce the three main components of the system design: radar point cloud extraction, 3D radar ego-motion estimation, and static background removal.

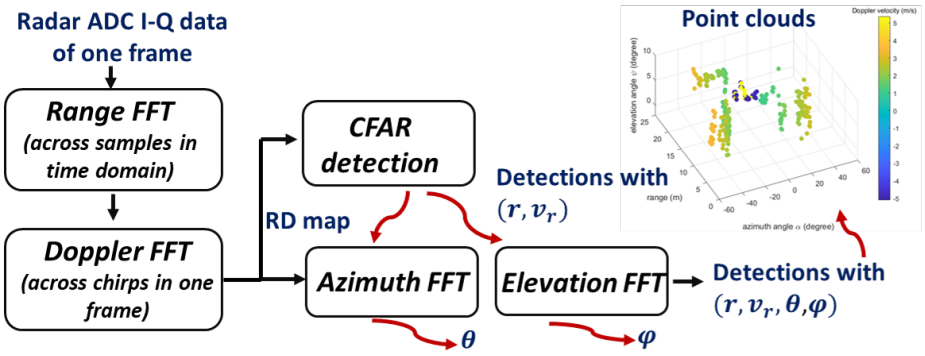

Figure 4.4: Workflow for processing one frame of radar data to extract radar point clouds: a collection of detections with corr. $(r, v_{\mathrm{r}}, \theta, \varphi)$ values.

4.3.1 Radar Point Cloud Extraction

The point cloud extraction process is accomplished using basic fast Fourier transform (FFT) processing and cell-averaging constant-false-alarm-rate (CA-CFAR) detection techniques. The detailed workflow is illustrated in Fig. 4.4, where the input to the workflow is the ADC I-Q data, and the output is a collection of detections represented by their $(r, v_{\mathrm{r}}, \theta, \varphi)$ values that constitute '*point clouds*'.

The signal processing architecture is based on the approach presented in our prior work [25], with an additional step for elevation angle estimation. The first step estimates the range and Doppler velocity via two sequential FFTs - Range FFT and Doppler FFT - as shown[25]. The Range FFT is performed across the samples in the fast time domain to extract the

round-time delay term $S_{\mathrm{w}}\tau_n'/f_{\mathrm{s}}$ in Eq. (4.7), and the Doppler FFT processes the slow time (chirps) to extract the Doppler-induced phase term $-4\pi f_{\mathrm{c}} v_{\mathrm{r}} N_T T_{\mathrm{c}}/c_0$ in Eq. (4.7) for each range bin. The resulting Range-Doppler (RD) map is then processed by the CA-CFAR algorithm [48] to detect peaks and obtain their ranges and Doppler velocities (r, v_{r}). During the CA-CFAR detection process, each cell in the RD map is evaluated for the presence or absence of a target using a threshold, where the threshold is set according to the noise power estimate within the training cells.

Thereafter, the azimuth and elevation angles (θ, φ) are estimated for each detection at (r, v_{r}) via an angle of arrival (AoA) estimator. While there are many approaches to AoA estimation, we adopt the FFT-based (conventional) AoA estimator. For the specific range-Doppler cell corresponding to (r, v_{r}), we use FFTs to process the measurements of the virtual array (shown in Fig. 4.2) along the horizontal or vertical direction to extract the phase term $2\pi h \cos\theta/\lambda$ or $2\pi h \sin\varphi/\lambda$ in Eq. (4.7), thus obtaining estimates for (θ, φ). For non-stationary targets, the motion-induced phase errors at each virtual antenna location should be compensated for different TXs before the AoA FFT [52]. In general, for a uniform linear transmit array with N_T TXs, the received phases at the virtual elements corresponding to the i-th TX are corrected via rotation of $(i-1)\Delta\phi_v/N_T$, where $\Delta\phi_v = -4\pi f_{\mathrm{c}} v_{\mathrm{r}} N_T T_{\mathrm{c}}/c_0$ from prior Doppler estimation [52].

4.3.2 3D Radar Ego-motion Estimation

We assume that there are N_{total} detections (indexed by i) in the extracted point clouds, identified by $(r_i, v_{r,i}, \theta_i, \varphi_i)$, a sub-set of which belong to the class of stationary targets, which is initially unknown. Representing the stationary subset size by N, from Eq. (4.3), the relationship between the Doppler velocities and angles of N stationary targets $(i = 1, 2, \ldots, N)$ can be expressed as follows:

$$\begin{bmatrix} v_{r,1} \\ \vdots \\ v_{r,N} \end{bmatrix} = - \begin{bmatrix} \sin\theta_1 \cos\varphi_1 & \cos\theta_1 \cos\varphi_1 & \sin\varphi_1 \\ \vdots & \vdots & \vdots \\ \sin\theta_N \cos\varphi_N & \cos\theta_N \cos\varphi_N & \sin\varphi_N \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \tag{4.8}$$

For more accurate modeling, we assume errors are associated with azimuth angle measure-

ment θ_i, elevation angle measurement φ_i, and Doppler velocity measurement $v_{r,i}$. Denoting the ground-truth values for the azimuth angle, elevation angle, and Doppler velocity by Θ_i, Φ_i, and $V_{r,i}$, the measurement errors given by $\Theta_i - \theta_i$, $\Phi_i - \varphi$ and $V_{r,i} - v_{r,i}$, respectively, are assumed to follow zero-mean Gaussian distributions [122] with no mutual coupling. We next build an orthogonal distance regression (ODR) solver for Eq. (4.8). that estimates(v_x, v_y, v_z) as well as $(\Theta_i, \Phi_i, V_i, i = 1, \ldots, N)$ based on the measurement error model above. For simplicity, we denote the vector of (v_x, v_y, v_z) by $\mathbf{v}_\mathrm{c}$, the set $\{\Theta_i, i = 1, \ldots, N\}$ by $\mathbf{\Theta}$, the set $\{\Phi_i, i = 1, \ldots, N\}$ by $\mathbf{\Phi}$, and the set $\{V_i, i = 1, \ldots, N\}$ by $\mathbf{V}$ in the following. The objective function and constraint for the ODR estimator are given by [123]:

$$\begin{aligned} &\underset{\mathbf{v}_\mathrm{c}, \mathbf{\Theta}, \mathbf{\Phi}, \mathbf{V}}{\arg\min} \sum_{i=1}^{N} \eta_1^2(\Theta_i - \theta_i)^2 + \eta_2^2(\Phi_i - \varphi_i)^2 + (V_i - v_{r,i})^2 \\ &\qquad s.t.\ V_i = -\sin\Theta_i \cos\Phi_i v_x - \cos\Theta_i \cos\Phi_i v_y - \sin\Phi_i v_z,\ \forall i = 1, \ldots, N \end{aligned} \tag{4.9}$$

where η_1^2, η_2^2 are the relative weights for the first two terms (assuming the weight for the third term is 1). η_1^2, η_2^2 are determined by the ratios of the variances of the Doppler velocity measurement errors (σ_v^2), azimuth angle measurement errors (σ_θ^2), and elevation angle measurement errors (σ_φ^2), as follows:

$$\eta_1^2 = \frac{\sigma_v^2}{\sigma_\theta^2},\ \eta_2^2 = \frac{\sigma_v^2}{\sigma_\varphi^2}. \tag{4.10}$$

In practice, we determine the variance of measurement errors based on the theoretical resolution and substitute it into Eq. (4.10) to calculate the weights η_1^2, η_2^2.

The constraint in Eq. (4.9) can be utilized to eliminate the term of $(V_i - v_{r,i})^2$ from the objective function (thus also eliminate unknowns $\mathbf{V}$) by substitution, thereby transforming the original constrained optimization problem in Eq. (4.9) into an unconstrained minimization problem [123, 124], as follows:

$$\underset{\mathbf{v}_\mathrm{c}, \mathbf{\Theta}, \mathbf{\Phi}}{\arg\min} \sum_{i=1}^{N} \eta_1^2(\Theta_i - \theta_i)^2 + \eta_2^2(\Phi_i - \varphi_i)^2 + [\sin\Theta_i \cos\Phi_i v_x + \cos\Theta_i \cos\Phi_i v_y + \sin\Phi_i v_z + v_{r,i}]^2 \tag{4.11}$$

The new optimization problem has unknowns $\mathbf{v}_\mathrm{c}, \mathbf{\Theta}, \mathbf{\Phi}$ and aims to minimize the sum of squares of the orthogonal distances of measurements to the fitting hyperplane. To solve

this, we propose an algorithm consisting of three steps, summarized in Algorithm 1 outlined below.

Step 1. Obtain an initial solution for $\mathbf{v}_\mathrm{c}, \boldsymbol{\Theta}, \boldsymbol{\Phi}$:
The unknowns $\boldsymbol{\Theta}, \boldsymbol{\Phi}$ is initialized using the measurement values $\{\theta_i, i = 1, \ldots, N\}$, $\{\varphi_i, i = 1, \ldots, N\}$, and initial value of $\mathbf{v}_\mathrm{c}$ is determined by solving Eq. (4.8) via the RANSAC algorithm and least-square regression (LSR). Let $\mathbf{P} = \begin{bmatrix} \sin\theta_1 \cos\varphi_1 & \cos\theta_1 \cos\varphi_1 & \sin\varphi_1 \\ \vdots & \vdots & \vdots \\ \sin\theta_N \cos\varphi_N & \cos\theta_N \cos\varphi_N & \sin\varphi_N \end{bmatrix}$ and $\mathbf{Q} = \begin{bmatrix} v_{r,1} \\ \vdots \\ v_{r,N} \end{bmatrix}$, then the Least Squares solution to Eq. (4.8) is $\widehat{\mathbf{v}}_\mathrm{c} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{Q}$. In automotive scenarios, the elevation angle φ_i is typically small for long-range targets and consequently, the last column of $\mathbf{P}$ has values close to zero. This can lead to near- singular matrix of $\mathbf{P}^T\mathbf{P}$ and numerical difficulties in computing the inverse of $\mathbf{P}^T\mathbf{P}$. To address this, we adopt the Moore-Penrose pseudo-inverse denoted by $(\cdot)^\dagger$ to obtain a more reliable solution, i.e.:

$$\widehat{\mathbf{v}}_\mathrm{c} = (\mathbf{P}^T\mathbf{P})^\dagger\mathbf{P}^T\mathbf{Q} \tag{4.12}$$

where the unique pseudoinverse exists for any matrix (even singular) [125]. An efficient way to compute the pseudoinverse is by using SVD [126]. If $\mathbf{P}^T\mathbf{P} = \mathbf{W}\boldsymbol{\Sigma}\mathbf{U}^T$ is the SVD of $\mathbf{P}^T\mathbf{P}$, then $(\mathbf{P}^T\mathbf{P})^\dagger = \mathbf{U}\boldsymbol{\Sigma}^\dagger\mathbf{W}^T$. For a rectangular diagonal matrix such as $\boldsymbol{\Sigma}$, its pseudoinverse $\boldsymbol{\Sigma}^\dagger$ is obtained by taking the reciprocal of each non-zero element on the diagonal, leaving the zeros in place, and transposing the matrix [126].

A key underlying issue when solving for (v_x, v_y, v_z) in Eq. (4.8) is identifying or separating the subset of stationary detections from the set of all detections. For this, we use the RANSAC algorithm [1,121] on the input point clouds to find the inliers that achieve the best consensus for estimates obtained from Eq. (4.8). Following the workflow in Fig. 4.5, RANSAC randomly selects a subset of points from the dataset and fits the model, then identifies all data points in the full set that are consistent with the model within a certain threshold - these are inliers, while the remaining points are outliers [121]. The above two steps are iteratively repeated

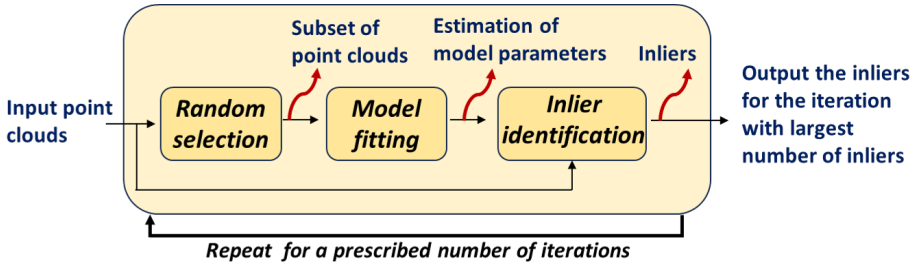

Figure 4.5: Workflow of RANSAC algorithm for identifying the inliers (the static subset) from the input point clouds $(r, v_{\mathrm{r}}, \theta, \varphi)$.

for a prescribed number of trials. After all trials, the model with the highest number of inliers is selected and the corresponding inliers are output. Increasing the number of trials improves the robustness and accuracy of the identified inliers because a more accurate model estimation can be explored through random sampling. In our case, the obtained inliers represent the stationary targets, while the outliers correspond to the moving targets. The identified inliners are then plug into Eq. (4.12) to get $\widehat{\mathbf{v}}_{\mathrm{c}}$ as the initial value of $\mathbf{v}_{\mathrm{c}}$.

Step 2. Using the initial solution, we employ the modified trust-region Levenberg-Marquardt-type (MTRLM) algorithm (from the ODRPACK library [127]) to solve Eq. (4.11) and obtain an improved estimation $\mathbf{v}_{\mathrm{c}}^{*}, \boldsymbol{\Theta}^{*}, \boldsymbol{\Phi}^{*}$.

The MTRLM algorithm combines the Levenberg-Marquardt method with the trust-region approach to improve the efficiency and robustness of solving nonlinear least squares problems [127]. After initialization, the MTRLM algorithm iterates the following steps until convergence: a) Compute the Jacobian and Hessian matrixes of the parameters, b) Determine the step size using the trust-region concept and the Jacobian and Hessian matrixes, c) Update the parameters of the model using the step size determined in the previous step, and d) Check for convergence criteria [127]. By following this iterative process, the MTRLM algorithm refines the initial solution and determines the best-fit solution $(\mathbf{v}_{\mathrm{c}}^{*}, \boldsymbol{\Theta}^{*}, \boldsymbol{\Phi}^{*})$ to the problem in Eq. (4.11).

Step 3. After the MTRLM algorithm converges, output the optimized radar ego velocity $\mathbf{v}_{\mathrm{c}}^{*}$.

Algorithm 1 3D Ego-motion Estimation Algorithm

procedure ESTIMATE
    # **Step 1**
    initialize $\mathbf{\Theta}, \mathbf{\Phi}$ using measurements θ_i, φ_i.
    initialize $\mathbf{v}_\mathrm{c}$ using the solution to Eq. (4.8) $\widehat{\mathbf{v}}_\mathrm{c}$ determined by RANSAC and LSR.
    # **Step 2**
    use MTRLM algorithm to iteratively optimize parameters $\mathbf{v}_\mathrm{c}, \mathbf{\Theta}, \mathbf{\Phi}$ in Eq. (4.11) until convergence.
    when coverage, obtain $\mathbf{v}_\mathrm{c}^*, \mathbf{\Theta}^*, \mathbf{\Phi}^*$.
    # **Step 3**
    output $\mathbf{v}_\mathrm{c}^*$.
end procedure

4.3.3 Dealing with Doppler Ambiguity: Heurestics

From [25], with N_T TXs and a chirp duration of N_c, the maximum unambiguously measurable Doppler velocity of radar is given by:

$$v_\mathrm{max} = \frac{\lambda}{4N_T T_c} \tag{4.13}$$

where $N_T T_c$ is the pulse repetition interval in TDM-MIMO. The current 4D imaging radar trend [128] is towards increasing the density of MIMO arrays. Hence, as the number of TXs N_T increases, it leads to a proportional reduction in the maximum unambiguously measurable velocity v_max [25], causing Doppler ambiguity whenever the true Doppler exceeds v_max.

In majority of urban street driving scenarios with a maximum speed limit of 25 mph (i.e., 11.2 m/s) [129] for ego vehicle (with radar mounted), as long as the radar pulse repetition interval is smaller than 87 us, Eq. (4.13) suggests that a desired v_max (greater than 11.2 m/s) is achieved, resulting in no-ambiguity measurements of stationary targets' Doppler velocities. This implies that by appropriate choice of pulse repetition interval, the Doppler ambiguity issue can be managed and the 3D ego-motion estimation method in Section 4.3.2 applied.

Nonetheless, a generalization of the proposed 3D ego-motion estimation method to include a Doppler ambiguity is necessary, and we propose a heuristic solution next.

Since ambiguity implies that true Doppler velocities beyond the range $[-v_{\max}, v_{\max}]$ will alias to an observed lower Doppler frequency within the range, we propose adding $2k\,v_{\max}$, where $k \in \mathbb{Z}$ is an integer, under the assumption that all Doppler measurements from static sources are aliased with the same k value. Hence we introduce an additive alias term $2kv_{\max}\mathbf{1}$ ($\mathbf{1}$ is a column vector of ones with size N) to the right-hand side of the Eq. (4.8) as follows:

$$\begin{bmatrix} v_{r,1} \\ \vdots \\ v_{r,N} \end{bmatrix} = - \begin{bmatrix} \sin\theta_1 \cos\varphi_1 & \cos\theta_1 \cos\varphi_1 & \sin\varphi_1 \\ \vdots & \vdots & \vdots \\ \sin\theta_N \cos\varphi_N & \cos\theta_N \cos\varphi_N & \sin\varphi_N \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} + 2kv_{\max}\mathbf{1} \tag{4.14}$$

While k is unknown, bounds on possible k values are readily determined, depending on $v_{\max}$ and the maximum vehicle driving speed. For example, if $v_{\max} = 5\,\text{m/s}$ and the maximum vehicle driving speed is 11.2 m/s (the speed limit for urban street driving), then $k \in \mathcal{K} = \{-1,\, 0,\, \text{or}\, 1\}$. This is because when k is -1 or 1, the maximum/minimum Doppler velocity of any stationary target can be folded into the range of $[-v_{\max}, v_{\max}]$ by adding $2kv_{\max}$[2].

We propose a heuristic approach to solve the best k as well as the radar ego velocity (v_x, v_y, v_z) or $\mathbf{v}_c$ from Eq. (4.14). That is, for each $k \in \mathcal{K}$, we substitute it into Eq. (4.14) and then apply the 3D ego-motion estimation algorithm proposed in Section 4.3.2 to solve (v_x, v_y, v_z). We iterate all k and determine the optimal k that yields the maximum number of inliers (from the RANSAC output in step 1 of Algorithm 1). In other words, we choose the k value that maximizes the number of data points that satisfies Eq. (4.14). Consequently, the best-fit radar ego velocity $\mathbf{v}_c^*$ would be the one corresponding to the optimal k.

4.3.4 Static Background Removal from Radar Image in Azimuth-Elevation-Doppler Domain

After obtaining the estimated radar ego-motion velocity $\mathbf{v}_c^*$, we leverage this information to remove the reflection of the static background from the radar images. Radar images are generated from FFT-based spectrograms, using the techniques described in [25, 36]. Initially,

[2] $k = 0$ represents the no-ambiguity case when the actual moving speed of radar lies within the bound.

we assume no Doppler ambiguity and background removal is implemented via an inverse approach to Eq. (4.8), wherein we compute the expected Doppler velocity v_r for each feasible azimuth-elevation angle pair (θ, φ) that satisfies the relationship in Eq. (4.8). We then utilize a 3D notch filter to eliminate the Doppler component from the radar image in the Doppler domain.

Notch filtering refers to removing or zeroing out specific components in the Doppler spectrum by element-wise multiplication with the frequency response of the filter. The frequency response of a 3D notch filter is obtained by combining the frequency responses of three one-dimensional (1D) filters. First, we design three separate second-order infinite impulse response (IIR) notch filters [130] for rejecting the frequency corresponding to azimuth angle θ, elevation angle φ, and Doppler v_r. Second, the frequency response of each 1D filter is expanded to 3D by replicating itself along the other two dimensions. Third, we take the point-wise minimum among the three expanded 3D filters to find the desired frequency response for the 3D notch filter. This methodology offers a simplified yet effective means of designing the 3D notch filter for background removal in the radar image.

To remove the static background, we perform element-wise multiplication between the radar image S_{input} and the frequency response of the 3D notch filter $H(\theta_i, \varphi_i, v_{r_i})$ constructed above for each angle pair (θ_i, φ_i) and corresponding Doppler velocity v_{r_i}. Assuming there are a total of M angle pairs covering all azimuth and elevation angles, the background removed radar image S_{output} can be calculated as follows:

$$S_{\text{output}} = S_{\text{input}} \otimes \prod_{i=1}^{M} |H(\theta_i, \varphi_i, v_{r,i})| \tag{4.15}$$

where $\otimes$ represents the element-wise multiplication.

Remark 1: The computation required for the background removal step can be significantly reduced by utilizing a fixed 3D notch filter that is easily adapted to different notch frequencies $(\theta_i, \varphi_i, v_{r,i})$ by simple frequency translation. This eliminates the need to design a new filter for each specific notch frequency. Further, the frequency response of the notch filter can be limited to a narrow region around the desired notch frequency $(\theta_i, \varphi_i, v_{r,i})$. By focusing only on a limited area, the number of operations required for the element-wise multiplication between the filter and the radar image is significantly reduced.

Remark 2: If the image is desired only for a specific elevation angle (i.e. fixed φ), it is possible to greatly reduce the number of angle pairs that need to be considered during the iteration process. Consequently, instead of employing a 3D notch filter, a more efficient 2D notch filter for background removal may be employed.

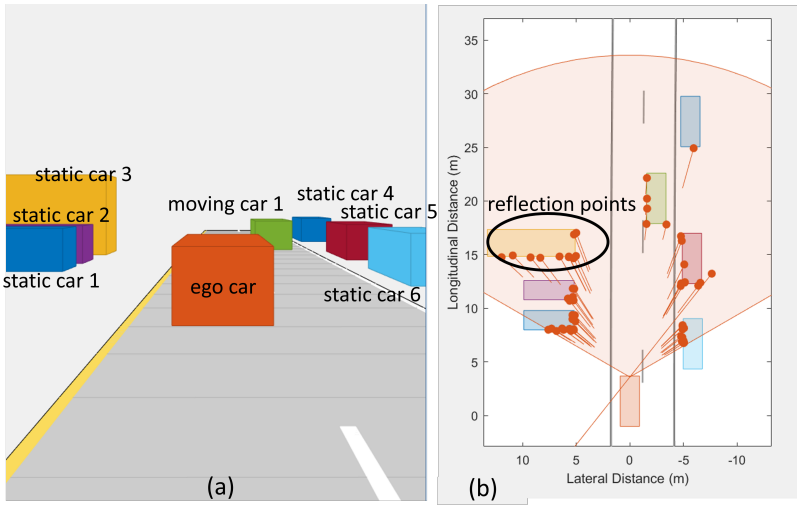

Figure 4.6: (a) Simulation scenario overview. (b) The simulated reflection points for targets are presented in the bird's-eye view.

4.4 Simulations

In this section, we performed simulations using the MATLAB Automated Driving Toolbox to simulate a scenario involving a moving vehicle equipped with an automotive radar, in a scene comprising both background (static) and moving objects. We applied both the 3D ego-motion estimation algorithm developed here and the background removal algorithm to the simulated raw radar data in order to explore system performance.

4.4.1 Simulated Scenario and Configuration

Scenario

We simulated a typical driving scenario involving an ego-moving vehicle along a linear road, three parked vehicles each on either side of the road, and one moving vehicle ahead of the ego-moving vehicle, as illustrated in Fig. 4.6(a). The ego-motion car has a forward velocity of $(8\,\mathrm{m/s}, 0, -0.5\,\mathrm{m/s})$ and an acceleration of $(2\,\mathrm{m/s^2}, 1\,\mathrm{m/s^2}, 0)$ and is equipped with a front-view FMCW radar. The other moving car had a constant velocity of $(5\,\mathrm{m/s}, 0, 0)$. The parked cars were represented by a collection of 3D reflection points within the radar's field of view, as indicated by the red markers in Fig. 4.6(b). Using the reflection points and the signal model described by Eq. (4.7), we generated post-demodulated ADC samples for each frame at the receiver, considering specific radar configuration described next.

Configuration

The radar configurations used for generating the ADC samples are as follows: $f_\mathrm{c} = 77\,\mathrm{GHz}$, $S_\mathrm{w} = 21.0017\,\mathrm{MHz/us}$, $A_\mathrm{T} = A_\mathrm{R} = 1$, $\phi_0 = 0$, $f_\mathrm{s} = 4\,\mathrm{Msps}$, the number of samples per chirp is 128, the number of chirps per frame sent by each TX is 255. We assume the formed virtual array is planar with 8×8 elements and antenna distance $h = \lambda/2$, where λ is the wavelength c_0/f_c. We consider the pulse repetition interval for TX to be 60 us, corresponding to a maximum unambiguous Doppler velocity of 16.5 m/s, which creates no Doppler ambiguity for most of the city/town street driving cases and is referred to as the '*w/o Doppler ambiguity*'. For evaluation purposes, we simulated a total of 40 frames with a frame rate of 20 fps.

4.4.2 3D Ego-motion Estimation

Baselines

The results from the ODR-based method in Sec. 4.3.2 are referred to as '*ODR*' in the subsequent evaluation. For an assessment of the estimation accuracy of the proposed method, we used the output of RANSAC and LSR $\widehat{\mathbf{v}}_\mathrm{c}$ (step 1 in Algorithm 1) as a baseline, referred to as '*LSR*' in the following. It is expected that the ODR performs better than the LSR

Table 4.2: Results for 3d ego-motion estimation.

Scenario	Method	RMSE (m/s)		
		v_x	v_y	v_z
w/o Doppler	LSR	0.0627	0.0489	0.4355
ambiguity	ODR	**0.0619**	**0.0389**	**0.3245**

since in Algorithm 1, the LSR estimation serves as the initialization for ODR optimization.

Implementation

We applied the ODR and LSR-based 3D radar odometry algorithm to the simulated radar ADC samples in order to estimate the ego-motion velocity for each frame, using the following parameter values: 128 range FFT points, 256 Doppler FFT points, 128 FFT points for azimuth angle estimation, 128 FFT points for elevation angle estimation, CFAR false alarm probability of 10^{-2}, and measurement error standard deviations of $\sigma_\theta = 0.25$, $\sigma_\phi = 0.25$, and $\sigma_v = 0.085$. The standard deviations σ_θ, σ_ϕ, and σ_v were determined based on the theoretical resolution of azimuth angle, elevation angle, and Doppler velocity, respectively [25]. The 3D ego-motion estimation algorithm utilized the RANSAC algorithm with a sample size of 4, a maximum distance threshold of 0.1 m/s for determining inliers, and a maximum number of trials of 2000. All methods were implemented with the same parameter values (where applicable) to ensure a fair comparison. To evaluate the accuracy of ego-motion velocity $\mathbf{v}_\mathrm{c}$ estimation, we compared the root mean square error (RMSE) relative to known ground truth values that were set for each frame. The results are presented in Table 4.2, where the 3 dimensions of $\mathbf{v}_\mathrm{c}$ - v_x, v_y, v_z - are listed separately.

Results and Analysis

For the results presented in Table 4.2, several observations can be made. First, the ODR predictions for v_x and v_y are accurate with low RMSE (0.0619 m/s and 0.0389 m/s, respectively),

Table 4.3: Results for 3d ego-motion estimation with Doppler ambiguity.

Scenario	Method	RMSE (m/s)		
		v_x	v_y	v_z
with Doppler ambiguity	LSR	0.0497	0.0361	0.3011
	LSR ($k = 0$)	9.5405	5.4323	2.3244
	ODR	**0.0411**	**0.0193**	**0.1536**

while the prediction for v_z has a larger error of 0.3245 m/s. A similar trend is observed for the LSR algorithm. This discrepancy is mainly due to the fact that the column $\sin\varphi_i$ in matrix $\mathbf{P}$ of Eq. (4.12) is close to zero, as most of the elevation angles φ_i are small and the estimation of v_z is more sensitive to errors. Secondly, compared to LSR, ODR achieves significant improvements in the v_y and v_z directions. This is attributed to the introduction of a model for measurement errors in the ODR model and the additional refinement process in step 2 of Algorithm 1.

4.4.3 3D Ego-motion Estimation with Doppler Ambiguity

While most operational cases fall into the 'w/o Doppler ambiguity' regime, we consider a scenario where the pulse interval for a TX is increased, leading to Doppler ambiguity. We consider a longer pulse repetition interval 180 us, which corresponds to the maximum unambiguous Doppler velocity 5.5 m/s. We refer to this scenario as '*with Doppler ambiguity*' in the subsequent evaluation. For a regular urban street driving with a speed limit of 11.2 m/s, we make the set for all possible k values to be $\mathcal{K} = \{-1, 0, 1\}$. Since the heuristic algorithm proposed in Section 4.3.3 only adds one step for estimating k to the algorithm in Section 4.3.2, we still use '*ODR*' to refer to and adopt the same baseline (i.e. '*LSR*') and metric in Section 4.4.2 for comparison and evaluation. To highlight its importance, we also consider the serious impact of *not* tackling the Doppler ambiguity (running algorithm assuming $k = 0$) and refer to this baseline as '*LSR ($k = 0$)*' within the description in

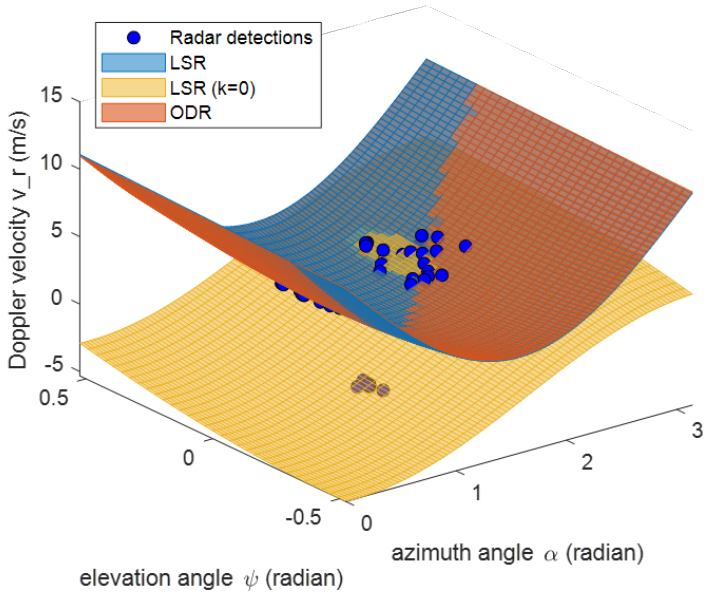

Figure 4.7: Regression planes of LSR, LSR ($k = 0$) estimate, and ODR methods for the scenario of 'with Doppler ambiguity'.

Section 4.4.2.

From the results shown in Table 4.3, several observations can be made. First, among all three methods, ODR continues to achieve the best ego-motion estimation performance in all three dimensions: v_x, v_y, and v_z. This outcome confirms the effectiveness of the ODR approach in handling errors in the variables model. Secondly, the LSR ($k = 0$) method performs poorly compared to LSR and ODR as it fails to account for the Doppler velocity measurements that lie outside $[-v_{\max}, v_{\max}]$, resulting in a poor fit to the data in Eq. (4.14). To illustrate it, we present the regression fitting results of the three methods in Fig. 4.7. The input radar detections are shown as blue circles, the regression plane for ODR is represented in orange, the regression plane for LSR is shown in blue, and the regression plane for LSR ($k = 0$) is depicted in yellow. From Fig. 4.7, it is evident that the LSR ($k = 0$) method only fits a portion of the radar input, and the overall regression plane deviates significantly from the correct regression plane (e.g., the one obtained by ODR) when Doppler ambiguity exists.

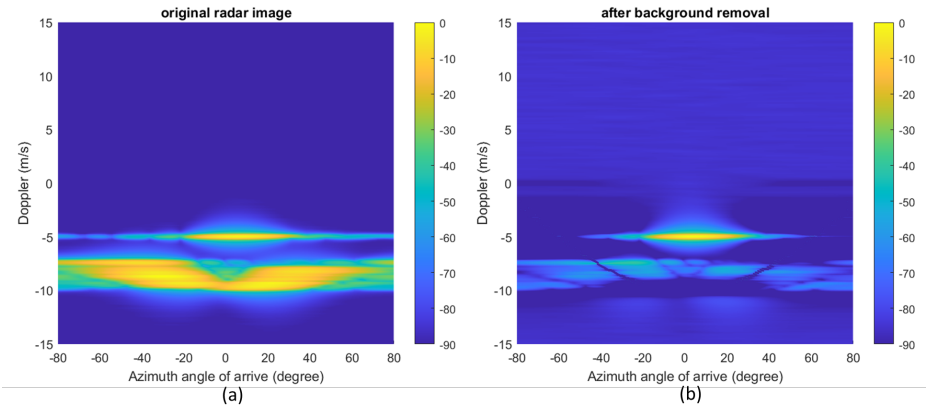

Figure 4.8: Doppler-azimuth angle images (for 0° elevation angle) before and after the background removal.

4.4.4 Background Removal Algorithm

Implementation

We applied the proposed static background removal algorithm to the simulated data under the 'w/o Doppler ambiguity' scenario. The required radar ego velocity was estimated through the ODR method in Section 4.4.1. The results from post-background filtering are shown via two 2-D Doppler-azimuth and range-azimuth angle images to avoid plotting sophisticated 3D images. As discussed in Section 4.3.4, these images are obtained by performing FFT operations on the raw radar ADC data [25].

Results and Analysis

The background removal results for the first frame of the simulated data are illustrated in Fig. 4.8 and Fig. 4.9. Specifically, we focused on 0° elevation plane for both the Doppler-azimuth and the range-azimuth angle image. In Fig. 4.8(a), which represents the Doppler-azimuth angle image before background removal, we observe that the main Doppler components of the static background are limited to the range $[-7\,\text{m/s}, -10\,\text{m/s}]$. We also observe that the relationship between the Doppler velocities of the static background and its azimuth angles follows a U-shaped pattern. Mathematically, this can be explained by

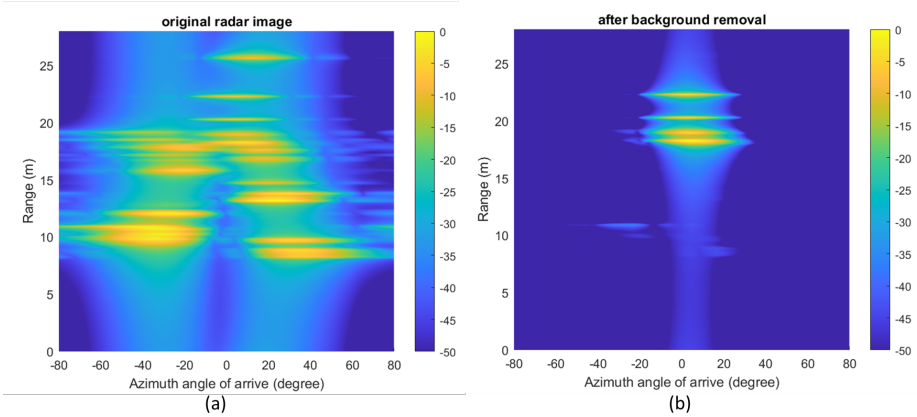

Figure 4.9: Range-azimuth angle images (for 0° elevation angle) before and after the background removal.

Eq. (4.3) when $\varphi = 0$. After substituting $\varphi = 0$ into Eq. (4.3), we get $v_r = v_y \cos\theta - v_x \sin\theta$ represents a U-shaped curve when plotted for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. After applying the background removal algorithm, the resulting Doppler-azimuth angle image is shown in Fig. 4.8(b). In this image, the mainlobe of the Doppler signal for each azimuth angle has been filtered out, leaving only a few remaining sidelobe components. Fig. 4.9(a) displays the original range-azimuth angle image, where the reflections from the moving car are mixed with clutter from stationary backgrounds (static cars). Separating these components through spatial filtering alone can be challenging. However, by utilizing the proposed background filtering method in the azimuth-elevation-Doppler domain, we obtain a significantly cleaner spatial image, as depicted in Fig. 4.9(b). This processed image effectively distinguishes the extended moving car and removes a significant portion of the clutter present in the original image.

To quantitatively assess the performance of the background removal algorithm, we analyzed the range profiles before and after the background removal process. In Fig. 4.10, the range profiles are plotted along with the average amplitudes (in dB) of the target (the moving car) and the background (the static cars). These amplitudes are labeled as the '*signal*' and '*interference*', respectively, and the ratio between them (i.e., SIR) is computed as the evaluation metric. Before background removal, the signal amplitude was measured

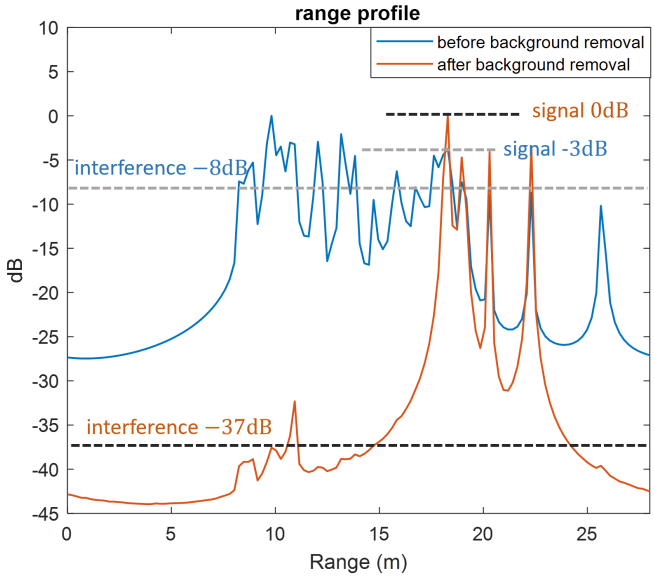

Figure 4.10: Range profile before and after the background removal.

at -3 dB, while the interference amplitude was -8 dB leading to SIR of $-3-(-8)=5$ dB. After the background removal, the signal amplitude increased to 0 dB, while the interference amplitude significantly decreased to -37 dB. This results in a significantly improved SIR of $0-(-37)=37$ dB. The significant improvement in the SIR, from 5 dB to 37 dB, demonstrates the remarkable clutter suppression capability of the proposed algorithm.

4.5 Experiments

Besides simulations, we also gathered significant measurement data using an off-the-shelf automotive radar testbed that allows us to assess the performance of the proposed static background removal algorithm in a real-world setting.

4.5.1 Experimental Setup and Configuration

Setup

A testbed was constructed using a TIDEP-01012 77 GHz mmWave radar [128] and two cameras, as depicted in Fig. 4.11(a). The Texas Instruments radar evaluation module employs

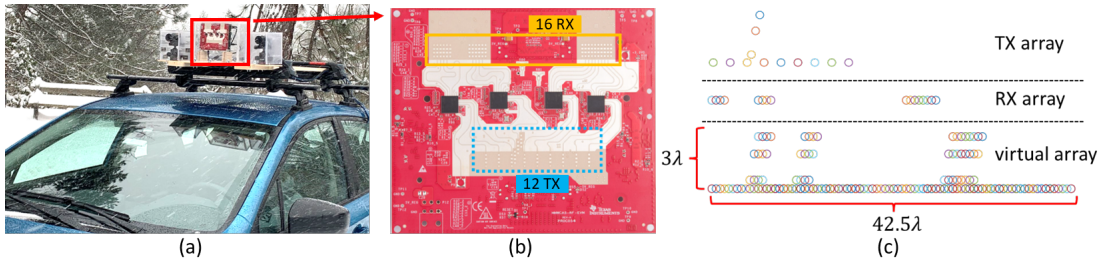

Figure 4.11: (a) Experimental radar testbed mounted on a vehicle. (b) The front view of the cascaded-chip radar board with the RX and TX highlighted. (c) The arrangement for the TX array, RX array, and the formed virtual array using TDM-MIMO.

a cascade of four radar chips, enabling greater MIMO dimension[128]. For data capture, the testbed was positioned at the front of the vehicle, enabling simultaneous collection of camera images and corresponding radar raw data (post-demodulated I-Q samples) [36]. The front view of the radar setup is illustrated in Fig. 4.11(b), featuring a 2D arrangement with 12 TXs and 16 RXs. By employing TDM MIMO techniques [25,27], the orthogonal signals from different TXs result in the formation of a 2D virtual receiver array, as shown in Fig. 4.11(c), obtained through the spatial convolution of all physical TX and RX pairs [25, 27]. The virtual array exhibits a sparse configuration (with 4 non-uniformly spaced elements spanning a 3λ aperture) in the vertical direction and a large uniform array (consisting of 86 uniformly spaced elements spanning a 42.5λ aperture) in the horizontal direction.

Configuration

The specific configuration of the cascaded-chip radar used in the data capture is as follows: $f_c = 77\,\text{GHz}$, $S_w = 45\,\text{MHz/µs}$, $f_s = 15\,\text{Msps}$, $N_T = 12$, $N_R = 16$, $T_c = 20\,\text{µs}$, sweep bandwidth 384 MHz, the number of chirps per TX per frame is 128, the number of samples per chirp is 128, and the frame rate is 30 fps. Based on Eq. (4.13), the maximum unambiguous Doppler velocity is 4 m/s.

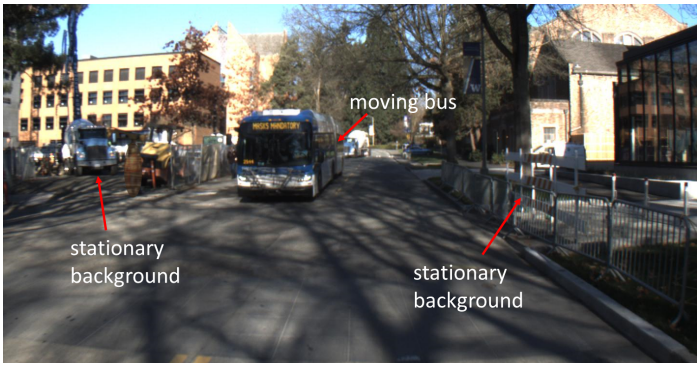

Figure 4.12: Typical Experiment scenario.

4.5.2 Background Removal Algorithm

Scenario

The experimental data was collected in a practical scenario with an ego car in motion in a complex environment, as shown in Fig. 4.12. The scene consisted of various stationary objects such as fences, trees, buildings, and parked cars on the roadsides. Additionally, there was a bus moving towards the ego vehicle. The testing phase involved the analysis of 15 frames of recorded radar I-Q samples, capturing the dynamics of the scene and the interactions between the ego car and its surroundings.

Implementation

We implemented the proposed 3D radar odometry algorithm and static background removal algorithm using the following parameter settings: Range, Doppler, Azimuth Angle, Elevant Angle FFTs with 128 points each, CFAR false alarm probability of 10^{-2}, ambiguity set $\mathcal{K} = \{-1, 0, 1\}$, and measurement error standard deviations $\sigma_\theta = 0.023$, $\sigma_\phi = 0.5$, $\sigma_v = 0.0634$. For the 3D ego-motion estimation, we utilized the RANSAC algorithm with a sample size of 4, a maximum distance threshold of 0.2 m/s for determining inliers, and a maximum number of trials set to 2000. The implementation was carried out using MATLAB R2019b on a computer equipped with an Intel i7-9750H CPU. Due to the limited availability of accurate radar ego-motion ground truth in real-world experiments, we were unable to perform a

comprehensive evaluation of the 3D ego estimation results.

Results and Analysis

We present the qualitative results of the static background removal algorithm in Fig. 4.13 and Fig. 4.14. Our focus is on the plane of 0° elevation for the Doppler-azimuth angle image and range-azimuth angle image. Fig. 4.13 shows the Doppler-azimuth angle image before and after the background removal. In Fig. 4.13(a), we observe that the Doppler components of the stationary background reflections span a range $[1.5\,\mathrm{m/s}, 3\,\mathrm{m/s}]$, forming a U-shaped relationship with the azimuth angle. Considering the vehicle's average speed (which exceeds $4\,\mathrm{m/s}$), the actual Doppler velocity for the stationary background should be a negative value calculated based on the azimuth angle and radar ego velocity. The measured Doppler velocity within the range of $1.5\,\mathrm{m/s}$ to $3\,\mathrm{m/s}$ indicates the presence of Doppler ambiguity (i.e $k \neq 0$). By utilizing our proposed 3D ego-motion estimation algorithm, the best k value is estimated as 1, confirming our earlier assumption. In Fig. 4.13(b), we observe that after the background removal, the strong Doppler components for each azimuth angle are mostly filtered out, while the main lobe from the moving bus and some sidelobe components from the background are still present. Furthermore, we assess the performance of the background removal algorithm on the range-azimuth angle image, as depicted in Fig. 4.14. In Fig. 4.14(a), we observe that the background components are efficiently removed, with the post-removal clutter amplitude below $-10\,\mathrm{dB}$, while the signal reflections from moving objects exhibit minimal change. This demonstrates the effectiveness of the static background removal algorithm in suppressing clutter and preserving the desired moving targets.

4.6 Discussion

4.6.1 Sensitivity Analysis for Ego-motion Estimation

The performance of the proposed end-to-end background removal algorithm relies on the accuracy of 3D ego-estimation results. It is important to note that the RANSAC algorithm is not guaranteed to find the correct inliers in all cases. It is possible for the 3D ego-estimation algorithm to converge to a model that contains false inliers (i.e., data points that do not

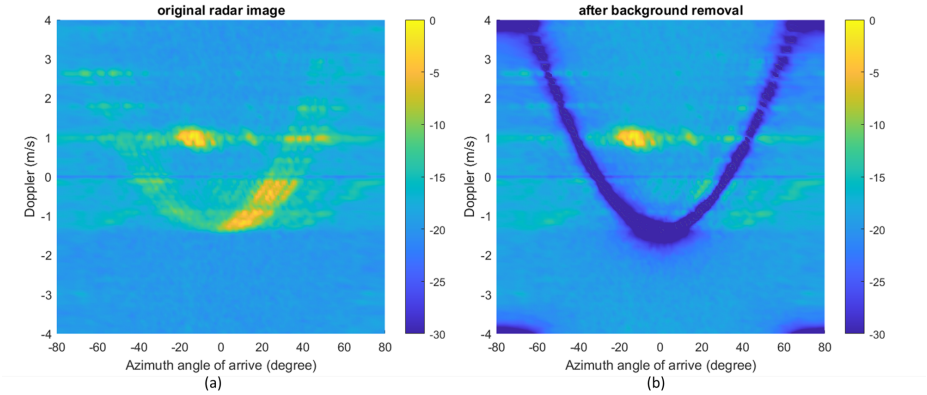

Figure 4.13: Doppler-azimuth angle images (for 0° elevation angle) before and after the background removal.

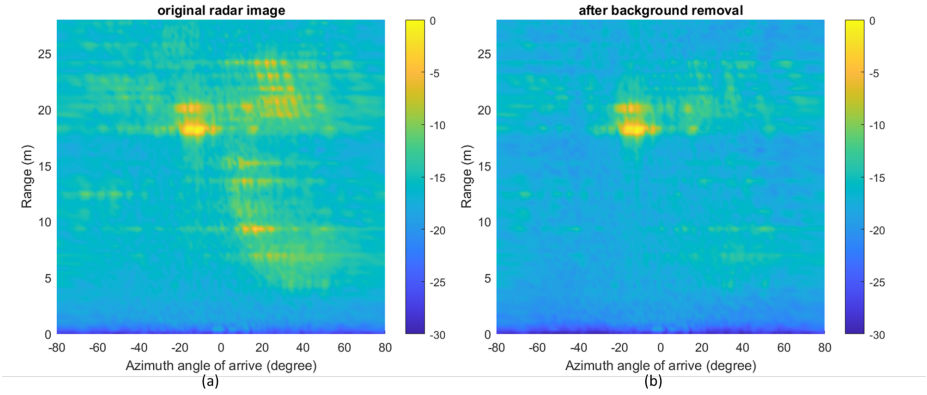

Figure 4.14: Range-azimuth angle images (for 0° elevation angle) before and after the background removal.

belong to the static background) or to miss some of the static elements, especially in challenging scenarios with a few static objects and a high density of moving objects. In the former cases, the presence of false inliers in the subsequent processing steps will lead to inaccurate background removal. To showcase the sensitivity of the ego-motion estimation algorithm in such instances, we extend the simulation to different scenarios. The original scenario in the simulation of Section 4.4.1 has 6 static cars off the road and 1 moving car on the road (except the ego car). We increased the number of moving cars to 3, 6, and 10 respectively, and tested the performance of ego-motion estimation. The results in Fig. 4.15 show that the ego-motion estimation errors for the scenarios of 1, 3, 6, and 10 moving cars are on a similar level, which indicates that with a sufficient amount of static cars, the ego-motion estimation is robust to the number of moving objects. Next, we kept 10 moving cars and reduced the number of static cars to 3 and 1. The results in Fig. 4.15 show that the ego-motion estimation works well for having 3 static cars but does not work for only 1 static car. Hence without a sufficient amount of static objects relative to the number of moving objects, the ego-motion estimation algorithm cannot identify the static set correctly using the RANSAC algorithm, which impacts the subsequent steps. In summary, the above shows that ego-motion estimation can work robustly unless there are very few stationary objects (relative to moving targets) in the scene. In such cases, combining other methodologies such as sensor fusion, to improve the robustness of 3D ego-motion estimation may be a potential solution.

4.6.2 Time Complexity Analysis

To analyze the time complexity of the algorithms, let's denote the dimensions of the radar data cube as N_s (samples), N_c (chirps), N_h (horizontal antennas), and N_e (vertical antennas). The overall time complexity of the proposed algorithm can be divided into three parts: 3D ego-motion estimation, radar imaging, and background filtering. For the 3D ego-motion estimation, the approximate complexity is $\mathcal{O}(N_s N_c N_h N_e \log N_s N_c) + \mathcal{O}(N_s N_c)$, which is determined by the range-Doppler FFT and CFAR detection operations. Here, we do not include the complexity of the azimuth angle FFT, elevation angle FFT, and ODR estimator

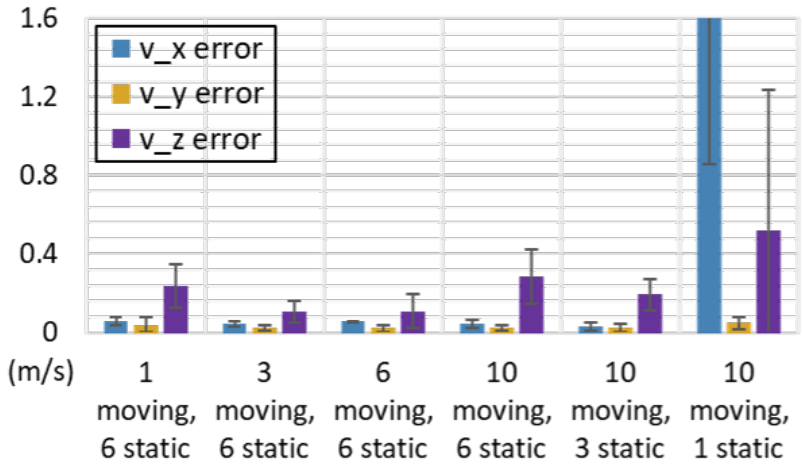

Figure 4.15: Performance of 3D ego-motion estimation for the cases of different numbers of moving cars and static cars.

since the number of CFAR detections in these steps is relatively small. Regarding radar imaging, the approximate complexity is $\mathcal{O}(N_s N_c N_h N_e \log N_h N_e)$, assuming that we start from the range-Doppler intermediate results obtained in the previous step. For background filtering, the approximate complexity is $\mathcal{O}(N_h N_e C)$, where C is the computation required for every notch filtering, which can be very small based on the analysis presented in the remark. Therefore, the overall time complexity of the algorithm is $\mathcal{O}(N_s N_c N_h N_e \log N_s N_c N_h N_e)$, after summing up the complexities mentioned above and ignoring the smaller terms.

We compare the time complexity of the proposed algorithm with two best-known reduced-rank STAP baselines: adjacent-bin post-Doppler STAP [110] and joint domain localized STAP [111]. Note that two baselines do not consider the elevation domain N_e in its degree of freedom for filtering. For a fair comparison to our algorithm, we assume the two baselines repeat the filtering operation at each elevation sample and incoherently sum the results. For adjacent-bin post-Doppler STAP, the overall time complexity is determined by the range-Doppler estimation, which is $\mathcal{O}(N_s N_c N_h N_e \log N_s N_c)$, and the space-time adaptive processing with reduced DOF $N_h L$, where L is the number of adjacent Doppler bins. For space-time adaptive processing, the complexity of processing one Doppler-angle position includes

Table 4.4: Time Complexity of Algorithms

Methods	Time Complexity
proposed method	$\mathcal{O}(N_s N_c N_h N_e \log N_s N_c N_h N_e)$
adjacent-bin post-Doppler STAP [110]	$\mathcal{O}(N_s N_c N_h N_e(\log N_s N_c + (N_h L)^3))$
joint domain localized STAP [111]	$\mathcal{O}(N_s N_c N_h N_e(\log N_s N_c N_h + (HL)^3))$

covariance matrix calculation $\mathcal{O}(2(N_h L)^3)$, covariance matrix inverse $\mathcal{O}((N_h L)^3)$, and weight computation and application $\mathcal{O}((N_h L)^2 + (N_h L))$ [131]. By summing up these complexities for all range bins, all Doppler-angle positions, and all elevation samples, the overall time complexity of space-time adaptive processing is $\mathcal{O}(N_s N_c N_h N_e (N_h L)^3)$. Therefore, the overall time complexity for adjacent-bin post-Doppler STAP is $\mathcal{O}(N_s N_c N_h N_e(\log N_s N_c + (N_h L)^3))$.

Compared to adjacent-bin post-Doppler STAP, the joint domain localized STAP further reduces the DOF to HL through beamforming on the antenna (space) domain, where H is the number of adjacent angle bins. Similar to the previous analysis, we can break down the time complexity of joint domain localized STAP into two parts: range-Doppler-angle estimation, which has a complexity of $\mathcal{O}(N_s N_c N_h N_e \log N_s N_c N_h)$, and space-time adaptive processing, which has a complexity of $\mathcal{O}(N_s N_c N_h N_e (HL)^3)$. Therefore, the overall time complexity of joint domain localized STAP is $\mathcal{O}(N_s N_c N_h N_e(\log N_s N_c N_h + (HL)^3))$. Based on the above analysis and the results summarized in Table. 4.4, it is evident that the proposed algorithm has a significantly lower time complexity compared to the two state-of-the-art reduced-rank STAP methods. This is particularly advantageous when the N_e dimension is typically very small in practical scenarios, resulting in $\log N_h N_e \ll (N_h L)^3$ and $\log N_e \ll (HL)^3$.

4.7 Summary

This chapter presents an efficient algorithm for background removal in automotive radar applications using FMCW radar. The algorithm encompasses radar signal preprocessing, 3D ego-motion estimation, and clutter removal in the azimuth-elevation-Doppler domain using a notch filter approach. Extensive evaluations and analyses have demonstrated the algorithm's effectiveness in suppressing background clutter and reducing computation. However, the proposed background removal algorithm highly relies on the accuracy of 3D ego-estimation results, which may lack performance when applied to real-world conditions where there are few stationary objects and the 3D ego-estimation works poorly. Besides, the current 3D ego-motion estimation does not consider the temporal consistency across frames, which leads to a suboptimal solution. As a future direction, we plan to explore and develop robust techniques for enhancing 3D ego-motion estimation accuracy (e.g., multi-sensor fusion, incorporating multiple frames), which has the potential to address this issue and further improve the overall performance of the algorithm.

Chapter 5

CONCLUSIONS AND FUTURE DIRECTIONS

Throughout this thesis, several research works and preliminary results have highlighted the potential of mmWave radars in various applications for autonomous driving and security check. In the context of mmWave radars, we have explored feasible avenues to leverage available sensing and data sources, enabling the design of learning-based algorithms for these applications, and ensuring performance, reliability, and robustness under diverse scenarios. Specifically, in Chapter 2, we presented a deep learning-based system for recognizing road users in challenging conditions, a crucial aspect of autonomous driving. In Chapter 3, we introduced the novel MIMO-SAR algorithm, enabling high-resolution imaging in vehicle-mounted side-view radars while reducing computational load. Additionally, in Chapter 4, we proposed a stationary background removal algorithm catering to the complexities of vehicular radar. These contributions collectively underscore the potential of mmWave radars in advancing the field of autonomous driving applications.

Looking ahead, there are numerous opportunities for further research and development in this field. As radar-based applications continue to gain traction due to advancements in hardware, sensing, and computing technologies, designing intelligent systems with performance and reliability guarantees becomes essential. Future research directions may include exploring novel algorithms and models to further enhance the capabilities of radar-based systems. Incorporating enhanced-angular-resolution SAR imaging into the current framework of radar object detection is a promising direction to push the limit of the system's performance. Additionally, investigating the fusion of radar data with other sensor modalities, such as cameras and LiDAR, could lead to more robust and comprehensive perception systems. Furthermore, the application of machine learning and artificial intelligence techniques in radar signal processing holds the potential for optimizing radar performance and expanding its use in various domains.

In conclusion, this thesis sets the stage for further exploration and innovation in the field of mmWave radar applications for autonomous driving and security check. By leveraging the power of data-driven technologies and intelligent signal processing algorithms, we can unlock new possibilities and advancements in radar-based systems, ultimately contributing to safer, more efficient, and reliable autonomous driving and security solutions.

BIBLIOGRAPHY

[1] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Instantaneous ego-motion estimation using doppler radar," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2013, pp. 869–874.

[2] S. Singh, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," Tech. Rep., 2015.

[3] T. J. Crayton and B. M. Meier, "Autonomous vehicles: Developing a public health research agenda to frame the future of transportation policy," *Journal of Transport & Health*, vol. 6, pp. 245–252, 2017.

[4] W. D. Montgomery, R. Mudge, E. L. Groshen, S. Helper, J. P. MacDuffie, and C. Carson, "America's workforce and the self-driving future: Realizing productivity gains and spurring economic growth," 2018.

[5] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE access*, vol. 8, pp. 58 443–58 469, 2020.

[6] Synopsys. (2021) The 6 levels of vehicle autonomy explained. [Online]. Available: https://www.synopsys.com/automotive/autonomous-driving-levels.html

[7] NHTSA. (2021) Automated vehicles for safety. [Online]. Available: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

[8] S. International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," *SAE Int.*, vol. 4970, no. 724, pp. 1–5, 2018.

[9] F. Lambert. (2020) Tesla is adding a new '4d' radar with twice the range for self-driving. [Online]. Available: https://electrek.co/2020/10/22/tesla-4d-radar-twice-range-self-driving/

[10] L. Cao and H. Yin, "A blockchain-empowered platoon communication scheme for vehicular safety applications," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–6.

[11] L. Cao, H. Yin, R. Wei, and L. Zhang, "Optimize semi-persistent scheduling in nr-v2x: An age-of-information perspective," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2053–2058.

[12] L. Cao, H. Yin, J. Hu, and L. Zhang, "Performance analysis and improvement on dsrc application for v2v communication," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–6.

[13] L. Cao and H. Yin, "Resource allocation for vehicle platooning in 5g nr-v2x via deep reinforcement learning," in *2021 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2021, pp. 1–7.

[14] H. Yin, L. Zhang, and S. Roy, "Multiplexing urllc traffic within embb services in 5g nr: Fair scheduling," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1080–1093, 2021.

[15] H. Yin, S. Roy, and S. Jin, "Ieee wlans in 5 vs 6 ghz: A comparative study," in *Proceedings of the 2022 Workshop on Ns-3*, ser. WNS3 '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 25–32. [Online]. Available: https://doi.org/10.1145/3532577.3532580

[16] L. Cao, S. Roy, and H. Yin, "Resource allocation in 5g platoon communication: Modeling, analysis and optimization," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 5035–5048, 2023.

[17] L. Zhang, H. Yin, S. Roy, and L. Cao, "Multiaccess point coordination for next-gen wi-fi networks aided by deep reinforcement learning," *IEEE Systems Journal*, vol. 17, no. 1, pp. 904–915, 2023.

[18] H. Yin, S. Roy, and L. Cao, "Routing and resource allocation for iab multi-hop network in 5g advanced," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6704–6717, 2022.

[19] H. Yin, L. Cao, and X. Deng, "Scheduling and resource allocation for multi - hop urllc network in 5g sidelink," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–7.

[20] L. Zhang, H. Yin, Z. Zhou, S. Roy, and Y. Sun, "Enhancing wifi multiple access performance with federated deep reinforcement learning," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–6.

[21] H. Yin, P. Liu, K. Liu, L. Cao, L. Zhang, Y. Gao, and X. Hei, "Ns3-ai: Fostering artificial intelligence algorithms for networking research," in *Proceedings of the 2020 Workshop on Ns-3*, ser. WNS3 '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 57–64. [Online]. Available: https://doi.org/10.1145/3389400.3389404

[22] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, p. 2140, 2021.

[23] S. Campbell, N. O'Mahony, L. Krpalcova, D. Riordan, J. Walsh, A. Murphy, and C. Ryan, "Sensor technology in autonomous vehicles: A review," in *2018 29th Irish Signals and Systems Conference (ISSC)*. IEEE, 2018, pp. 1–4.

[24] X. Gao, S. Roy, and L. Zhang, "Static background removal in vehicular radar: Filtering in azimuth-elevation-doppler domain," *arXiv preprint arXiv:2307.01444*, 2023.

[25] X. Gao, G. Xing, S. Roy, and H. Liu, "Experiments with mmwave automotive radar test-bed," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 1–6.

[26] X. Gao, "Towards millimeter-wave radar signal processing and learning-based applications," 2021.

[27] X. Gao, S. Roy, G. Xing, and S. Jin, "Perception through 2d-mimo fmcw automotive radar under adverse weather," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, 2021, pp. 1–5.

[28] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, "Learning to detect open carry and concealed object with 77 ghz radar," pp. 791–803, 2022.

[29] X. Gao, S. Ding, K. Vanäs, H. R. Dasari, H. Söderlund, and Yuan, "Deformable radar polygon: A lightweight and predictable occupancy representation for short-range collision avoidance," 2022, available at https://arxiv.org/pdf/2203.01442.pdf.

[30] D. Sihao, X. Gao, and H. R. Dasari, "Deformable radar polygon systems and methods for a virtual bumper," May 18 2023, uS Patent App. 17/980,618.

[31] S. Jin, X. Gao, and S. Roy, "Cognition in automotive radars," in *Next-Generation Cognitive Radar Systems (Radar, Sonar and Navigation)*, K. V. Mishra, B. S. M.R., and M. Rangaswamy, Eds. IET, 2023.

[32] X. Gao, Y. Luo, G. Xing, S. Roy, and H. Liu, "Raw adc data of 77ghz mmwave radar for automotive object detection," 2022. [Online]. Available: https://dx.doi.org/10.21227/xm40-jx59

[33] X. Gao, S. Roy, H. Liu, Y. Luo, and G. Xing, "Raw adc data of 2d-mimo mmwave radar for carry object detection," 2022. [Online]. Available: https://dx.doi.org/10.21227/begn-ye78

[34] X. Gao, Y. Sun, H. Chen, X. Xu, and S. Cui, "Joint computing, pushing and caching optimization for mec networks via soft actor-critic learning," 2023.

[35] X. Gao, Y. Sun, H. Chen, S. Cui, and X. Xu, "Soft actor-critic learning-based joint computing, pushing, and caching framework in mec networks," *arXiv preprint arXiv:2305.12099*, 2023.

[36] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-cnn: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5119–5132, 2021.

[37] X. Gao, S. Roy, and G. Xing, "Mimo-sar: A hierarchical high-resolution imaging algorithm for mmwave fmcw radar in autonomous driving," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7322–7334, 2021.

[38] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M. Kaafar, "The impact of adverse weather conditions on autonomous vehicles: how rain, snow, fog, and hail affect the performance of a self-driving car," *IEEE Vehicular Technology Magazine*, vol. 14, no. 2, pp. 103–111, 6 2019.

[39] A. Angelov, A. Robertson, R. Murray-Smith, and F. Fioranelli, "Practical classification of different moving targets using automotive radar and deep neural networks," *IET Radar, Sonar Navigation*, vol. 12, no. 10, pp. 1082–1089, 2018.

[40] B. Major, D. Fontijne, A. Ansari, R. Teja Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. Subramanian, "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[41] R. Pérez, F. Schubert, R. Rasshofer, and E. Biebl, "Deep learning radar object detection and classification for urban automotive scenarios," in *2019 Kleinheubach Conference*, Sep. 2019, pp. 1–4.

[42] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "Rodnet: Object detection under severe conditions using vision-radio cross-modal supervision," 2020.

[43] Liang Yi, Wang Hongxian, Xing Mengdao, and Bao Zheng, "Imaging study of high squint sar based on fmcw," in *1st Asian and Pacific Conference on Synthetic Aperture Radar*, 2007, pp. 6–9.

[44] G. Brooker, "Understanding millimetre wave fmcw radars," *1st International Conference on Sensing Technology*, 01 2005.

[45] S. Suleymanov, "Design and implementation of an fmcw radar signal processing module for automotive applications," August 2016, available at http://essay.utwente.nl/70986/.

[46] C. D. Monaco and S. N. Brennan, "Radarodo: Ego-motion estimation from doppler and spatial data in radar images," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 475–484, 2020.

[47] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017.

[48] M. A. Richards, Ed., *Principles of Modern Radar: Basic principles*, ser. Radar, Sonar and amp; Navigation. Institution of Engineering and Technology, 2010, available at https://digital-library.theiet.org/content/books/ra/sbra021e.

[49] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[50] C. Iovescu and S. Rao, *White paper: The Fundamentals of Millimeter Wave Sensors*. Texas Instrument, 2017, no. SPYY005.

[51] S. Rao, *White paper: MIMO Radar*. Texas Instrument, 2017, no. SWRA554A. [Online]. Available: https://www.ti.com/lit/an/swra554a

[52] J. Bechter, F. Roos, and C. Waldschmidt, "Compensation of motion-induced phase errors in tdm mimo radars," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 12, pp. 1164–1166, 2017.

[53] W. Wang, "Virtual antenna array analysis for mimo synthetic aperture radars," *International Journal of Antennas and Propagation*, vol. 2012, pp. 1–10, 2012.

[54] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on Antennas and Propagation*, vol. 16, no. 2, pp. 172–175, 1968.

[55] Z. Shou, J. Chan, A. Zareian, K. Miyazawa, and S.-F. Chang, "Cdc: Convolutional-deconvolutional networks for precise temporal action localization in untrimmed videos," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[56] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," 06 2011, pp. 52–59.

[57] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 7356–7365.

[58] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[59] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *CVPR*, 2017.

[60] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 734–750.

[61] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv preprint arXiv:1904.07850*, 2019.

[62] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[63] M. A. Richards, *Fundamentals of Radar Signal Processing.* US: McGraw-Hill Professional, 2005. [Online]. Available: https://mhebooklibrary.com/doi/book/10.1036/0071444742

[64] *xWR1843 Evaluation Module (xWR1843BOOST) Single-Chip mmWave Sensing Solution.* US: Texas Instrument, 2018. [Online]. Available: https://www.ti.com/tool/AWR1843BOOST#technicaldocuments

[65] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017, pp. 464–472.

[66] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

[67] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

[68] S. Devi, P. Malarvezhi, R. Dayana, and K. Vadivukkarasi, "A comprehensive survey on autonomous driving cars: A perspective view," *Wireless Personal Communications*, 05 2020.

[69] K. Yoneda, N. Suganuma, R. Yanase, and M. Aldibaja, "Automated driving recognition technologies for adverse weather conditions," *IATSS Research*, vol. 43, no. 4, pp. 253 – 262, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0386111219301463

[70] Y. Golovachev, A. Etinger, G. Pinhasi, and Y. Pinhasi, "Millimeter wave high resolution radar accuracy in fog conditions-theory and experimental verification," *Sensors*, vol. 18, 07 2018.

[71] R. Feng, F. Uysal, and A. Yarovoy, "Target localization using mimo-monopulse: Application on 79 ghz fmcw automotive radar," in *15th European Radar Conference (EuRAD)*, 2018, pp. 59–62.

[72] T. Kan, G. xin, L. xiaowei, and L. zhongshan, "Implementation of real-time automotive sar imaging," in *IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2020, pp. 1–4.

[73] F. Harrer, F. Pfeiffer, A. Löffler, T. Gisder, and E. Biebl, "Synthetic aperture radar algorithm for a global amplitude map," in *14th Workshop on Positioning, Navigation and Communications (WPNC)*, 2017, pp. 1–6.

[74] A. Laribi, M. Hahn, J. Dickmann, and C. Waldschmidt, "Performance investigation of automotive sar imaging," in *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, 2018, pp. 1–4.

[75] X. Hu, Y. Li, M. Lu, Y. Wang, and X. Yang, "A multi-carrier-frequency random-transmission chirp sequence for tdm mimo automotive radar," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3672–3685, 2019.

[76] X. Dong and Y. Zhang, "Sar/isar imaing using commercial millimeter-wave radar," in *6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, 2019, pp. 1–4.

[77] J. Kulpa, M. Malanowski, D. Gromek, P. Samczyńsk, K. Kulpa, and A. Gromek, "Experimental results of high-resolution isar imaging of ground-moving vehicles with a stationary fmcw radar," *International Journal of Electronics and Telecommunications*, vol. 59, pp. 293–299, 2013.

[78] F. Fembacher, F. B. Khalid, G. Balazs, D. T. Nugraha, and A. Roger, "Real-time synthetic aperture radar for automotive embedded systems," in *15th European Radar Conference (EuRAD)*, 2018, pp. 517–520.

[79] Z. Li, J. Wang, and Q. H. Liu, "Frequency-domain backprojection algorithm for synthetic aperture radar imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 4, pp. 905–909, 2015.

[80] K. Hu, X. Zhang, S. He, H. Zhao, and J. Shi, "A less-memory and high-efficiency autofocus back projection algorithm for sar imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 4, pp. 890–894, 2015.

[81] L. A. Gorham and L. J. Moore, "SAR image formation toolbox for MATLAB," in *Algorithms for Synthetic Aperture Radar Imagery XVII*, vol. 7699. SPIE, 2010, pp. 46 – 58, available at https://doi.org/10.1117/12.855375.

[82] S. Jun, M. Long, and Z. Xiaoling, "Streaming bp for non-linear motion compensation sar imaging based on gpu," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 4, pp. 2035–2050, 2013.

[83] X. Kong, *Inertial navigation system algorithms for low cost IMU*. Department of Mechanical and Mechatronic Engineering, Graduate School, 2000.

[84] A. Pishehvari, M. Stefer, and B. Tibken, "Robust range-doppler registration with hd maps," in *5th International Conference on Control and Robotics Engineering (ICCRE)*, 2020, pp. 140–147.

[85] S. Jin and S. Roy, "Fmcw radar network: Multiple access and interference mitigation," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2021.

[86] W. Zhang, P. Wang, N. He, and Z. He, "Super resolution doa based on relative motion for fmcw automotive radar," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8698–8709, 2020.

[87] Z. Xu, C. J. Baker, and S. Pooni, "Range and doppler cell migration in wideband automotive radar," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5527–5536, 2019.

[88] *Minimum mean square error*. US: Wikipedia, 2021, available at https://en.wikipedia.org/wiki/Minimum_mean_square_error.

[89] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, p. 381–395, Jun. 1981, available at https://doi.org/10.1145/358669.358692.

[90] I. Cumming and F. Wong, "Digital processing of synthetic aperture radar data: Algorithms and implementation. norwood, ma: Artech house, 2005. 2, 20," 1999.

[91] K. D. Rolt and H. Schmidt, "Azimuthal ambiguities in synthetic aperture sonar and synthetic aperture radar imagery," *IEEE Journal of Oceanic Engineering*, vol. 17, no. 1, pp. 73–79, 1992.

[92] L. Li and J. L. Krolik, "Vehicular mimo sar imaging in multipath environments," in *IEEE RadarCon (RADAR)*, 2011, pp. 989–994.

[93] *AWR1843 Single-Chip 77- and 79-GHz FMCW Radar Sensor datasheet.* US: Texas Instrument, 2020, available at https://www.ti.com/lit/ds/symlink/awr1843.pdf.

[94] *Folded normal distribution.* US: Wikipedia, 2020, available at https://en.wikipedia.org/wiki/Folded_normal_distribution.

[95] *Automated Driving Toolbox.* US: MATLAB, 2020, available at https://www.mathworks.com/help/driving/index.html.

[96] T. Gisder, F. Harrer, and E. Biebl, "Application of a stream-based sar-backprojection approach for automotive environment perception," in *19th International Radar Symposium (IRS)*, 2018, pp. 1–10.

[97] Y. Jin, A. Deligiannis, J.-C. Fuentes-Michel, and M. Vossiek, "Cross-modal supervision-based multitask learning with automotive radar raw data," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 3012–3025, 2023.

[98] X. Cai, M. Giallorenzo, and K. Sarabandi, "Machine learning-based target classification for mmw radar in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 4, pp. 678–689, 2021.

[99] J. Xu, S. Zhu, and G. Liao, "Space-time-range adaptive processing for airborne radar systems," *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1602–1610, 2014.

[100] W. L. Melvin, "A stap overview," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 19–35, 2004.

[101] J. Ward, "Space-time adaptive processing for airborne radar," 1998.

[102] W. L. Melvin and M. C. Wicks, "Improving practical space-time adaptive radar," in *Proceedings of the 1997 IEEE National Radar Conference.* IEEE, 1997, pp. 48–53.

[103] Y. Gu, J. Wu, Y. Fang, L. Zhang, and Q. Zhang, "End-to-end moving target indication for airborne radar using deep learning," *Remote Sensing*, vol. 14, no. 21, p. 5354, 2022.

[104] B. Himed, Y. Zhang, and A. Hajjari, "Stap with angle-doppler compensation for bistatic airborne radars," in *Proceedings of the 2002 IEEE Radar Conference (IEEE Cat. No. 02CH37322).* IEEE, 2002, pp. 311–317.

[105] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, no. 6, pp. 853–863, 1974.

[106] P. Berens, "Very high resolution sar and multichannel sar/mti," FGAN-FHR Research Institution for High Frequency Physics and Radar Technique, Tech. Rep., 2006.

[107] C. E. Muehe and M. Labitt, "Displaced-phase-center antenna technique," *Lincoln Laboratory Journal*, vol. 12, no. 2, pp. 281–296, 2000.

[108] W. Bürger, "Space-time adaptive processing: Algorithms," *Advanced Radar Signal and Data Processing. Educational Notes RTO-EN-SET-086, Paper*, vol. 7, 2006.

[109] R. Klemm, "Introduction to space-time adaptive processing," 1998.

[110] R. C. DiPietro, "Extended factored space-time processing for airborne radar systems," in *Conference Record of the Twenty-Sixth Asilomar Conference on Signals, Systems & Computers*. IEEE Computer Society, 1992, pp. 425–426.

[111] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 660–670, 1994.

[112] R. Solimene, A. Cuccaro, A. Dell'Aversano, I. Catapano, and F. Soldovieri, "Background removal methods in gpr prospecting," in *2013 European Radar Conference*, 2013, pp. 85–88.

[113] A. Bati, L. To, and D. Hilliard, "Advanced radar cross section clutter removal algorithms," in *Proceedings of the Fourth European Conference on Antennas and Propagation*, 2010, pp. 1–5.

[114] Y.-S. Yoon and M. G. Amin, "Spatial filtering for wall-clutter mitigation in through-the-wall radar imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 9, pp. 3192–3208, 2009.

[115] M. G. Ehrnsperger, M. Noll, U. Siart, and T. F. Eibert, "Background and clutter removal techniques for ultra short range radar," in *2020 17th European Radar Conference (EuRAD)*, 2021, pp. 78–81.

[116] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "An svd-based approach for mitigating wall reflections in through-the-wall radar imaging," in *2011 IEEE RadarCon (RADAR)*. IEEE, 2011, pp. 519–524.

[117] J. Liu, W. Xiong, L. Bai, Y. Xia, T. Huang, W. Ouyang, and B. Zhu, "Deep instance segmentation with automotive radar detection points," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 84–94, 2023.

[118] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, "Learning to detect open carry and concealed object with 77 ghz radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 791–803, 2022.

[119] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "Rodnet: Radar object detection using cross-modal supervision," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 504–513.

[120] N. Pandey and S. S. Ram, "Classification of automotive targets using inverse synthetic aperture radar images," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 675–689, 2022.

[121] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[122] X. Gao, S. Ding, S. Vanas, R. Dasari Harshavardhan, and H. Soderlund, "Deformable radar polygon: A lightweight and predictable occupancy representation for short-range collision avoidance," *arXiv preprint arXiv:2203.01442*, 2022.

[123] P. T. Boggs and J. E. Rogers, "Orthogonal distance regression," *Contemporary Mathematics*, vol. 112, pp. 183–194, 1990.

[124] C. C. Stahoviak, "An instantaneous 3d ego-velocity measurement algorithm for frequency modulated continuous wave (fmcw) doppler radar data," Ph.D. dissertation, University of Colorado at Boulder, 2019.

[125] "Moore–penrose inverse," https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse, accessed: 2023-07-18.

[126] "Linear systems & pseudo-inverse," http://websites.uwlax.edu/twill/svd/systems/index.html, accessed: 2023-07-25.

[127] P. T. Boggs, J. R. Donaldson, R. h. Byrd, and R. B. Schnabel, "Algorithm 676: Odrpack: software for weighted orthogonal distance regression," *ACM Transactions on Mathematical Software (TOMS)*, vol. 15, no. 4, pp. 348–364, 1989.

[128] T. Instrument, *White paper: Imaging Radar Using Cascaded mmWave Sensor Reference Design*. Texas Instrument, 2019, no. TIDUEN5A, available at https://www.ti.com/lit/ug/tiduen5a/tiduen5a.pdf.

[129] "Speed limits and traffic calming - mrsc," https://mrsc.org/explore-topics/transportation/traffic-codes,-regulation-and-enforcement/speed-limits-traffic-calming, accessed: 2023-07-18.

[130] C. M. Wang and W. C. Xiao, "Second-order iir notch filter design and implementation of digital signal processing system," in *Applied Mechanics and Materials*, vol. 347. Trans Tech Publ, 2013, pp. 729–732.

[131] K. C. Cain, J. A. Torres, and R. T. Williams, *RT-STAP, Real-time Space-time Adaptive Processing Benchmark*. Rome Laboratory, Air Force Materiel Command, 1997.

[132] G. M. Brooker, "Understanding millimetre wave fmcw radars," in *1 st International Conference on Sensing Technology, IEEE, New Zealand, 2005*, pp. 152–157.

Appendix A

RADAR DATA AUGMENTATION

A.1 Translating in Range

Suppose there is a target located at a certain range r_1, azimuth angle θ_1, the corresponding FMCW de-chirped signal (single chirp) at time t and receiver q can be expressed as [132]:

$$f(t,q) = A_1 \exp\left(j2\pi\left(f_c\tau_1 + St\tau_1 - \frac{1}{2}S\tau_1^2 + \frac{qd\sin\theta_1}{\lambda}\right)\right) \tag{A.1}$$

where $\tau_1 = \frac{2r_1}{c_0}$ is the round-trip delay between radar and target, A_1 is the signal amplitude determined by target's radar cross section (RCS), range, and antenna gain.

With the Quadrature analog-to-digital conversion (ADC) sampling, the de-chirped signal can be expressed as $f(\frac{i}{f_s}, q)$, i.e., t is replaced by the sampling time $\frac{i}{f_s}$, where i is the sample index, and f_s is the sampling frequency.

Implementing the Range FFT on the digitized de-chirped signal, we obtain the range profile $F_R(m_r, q)$ below:

$$\begin{aligned} F_R(m_r,q) =& \sum_{i=0}^{N_s-1} f(\frac{i}{f_s},q)\exp(-j\frac{2\pi m_r i}{M_r}) \\ =& A_1\exp\left(j2\pi\left(f_c\tau_1 - \frac{1}{2}S\tau_1^2\right)\right)\exp\left(j2\pi\frac{qd\sin\theta_1}{\lambda}\right) \\ &\times\left[\sum_{i=0}^{N_s-1}\exp\left(j2\pi S\frac{i}{f_s}\tau_1\right)\exp(-j\frac{2\pi m_r i}{M_r})\right] \\ =& A_1\exp\left(j2\pi\left(f_c\tau_1 - \frac{1}{2}S\tau_1^2\right)\right)\exp\left(j2\pi\frac{qd\sin\theta_1}{\lambda}\right) \\ &\times\frac{1-\exp\left(j2\pi N_s\left(\frac{S\tau_1}{f_s}-\frac{m_r}{M_r}\right)\right)}{1-\exp\left(j2\pi\left(\frac{S\tau_1}{f_s}-\frac{m_r}{M_r}\right)\right)} \end{aligned} \tag{A.2}$$

where M_r are the number of points for Range FFT, $m_r \in \{0, 1, ..., M_r - 1\}$ is the range bin

index. The mapping relationship between any range r and bin index is shown below:

$$m_r = \lfloor \frac{2M_r Sr}{c_0 f_{\mathrm{s}}} \rfloor \tag{A.3}$$

If the target at (r_1, θ_1) with corresponding Cartesian-coordinates (x_1, y_1) has a translation in range, Δr, as we defined in Section 2.4, i.e. $x_1' = x_1,\ y_1' = \sqrt{y_1^2 + \Delta r^2 + 2\Delta r r_1}$, then its new location (r_1', θ_1') would be $r_1' = r_1 + \Delta r$, $\theta_1' = \arctan(\frac{x_1'}{y_1'})$. According to the radar range equation, the energy loss ratio due to range shift is $(r_1/\left(r_1 + \Delta r\right))^2$. Thus, the corresponding FMCW de-chirped signal $f_{\mathrm{new}}(t, q)$ for the target at new location (r_1', θ_1') is expressed as:

$$f_{\mathrm{new}}(t, q) = A_1 \left(\frac{r_1}{r_1 + \Delta r}\right)^2 \exp(j2\pi(f_{\mathrm{c}}\tau_1' + St\tau_1' - \frac{1}{2}S{\tau_1'}^2 + \frac{qd\sin\theta_1'}{\lambda})) \tag{A.4}$$

where $\tau_1' = \frac{2r_1'}{c_0}$ is the new round-trip delay between radar and target.

The range profile for the digitized signal $f_{\mathrm{new}}(\frac{i}{f_{\mathrm{s}}}, q)$ can be obtained with similar derivations in Eq. (A.2) and Eq. (A.3). That is,

$$\begin{aligned} F_{\mathrm{R,\ new}}(m_r, q) =& A_1 \left(\frac{r_1}{r_1 + \Delta r}\right)^2 \exp\left(j2\pi\left(f_{\mathrm{c}}\tau_1' - \frac{1}{2}S{\tau_1'}^2\right)\right) \\ & \times \exp\left(j2\pi\frac{qd\sin\theta_1'}{\lambda}\right) \frac{1 - \exp\left(j2\pi N_{\mathrm{s}}\left(\frac{S\tau_1'}{f_{\mathrm{s}}} - \frac{m_r}{M_r}\right)\right)}{1 - \exp\left(j2\pi\left(\frac{S\tau_1'}{f_{\mathrm{s}}} - \frac{m_r}{M_r}\right)\right)} \end{aligned} \tag{A.5}$$

Based on the triangular geometry, the $\sin\theta_1'$ term in Eq. (A.5) can be simplified by $\sin\theta_1' = \frac{x_1'}{r_1'} = \frac{x_1}{r_1'}$. Ignoring the change of Doppler phase term $\exp\left(j2\pi\left(f_{\mathrm{c}}\tau_1' - \frac{1}{2}S{\tau_1'}^2\right)\right)$, we would approximate the remaining terms of Eq. (A.5) with the original profile Eq. (A.2) by following relation:

$$\begin{aligned} F_{\mathrm{R,\ new}}(m_r, q) \approx & \left(\frac{r_1}{r_1 + \Delta r}\right)^2 \left|F_{\mathrm{R}}(m_r + \Delta m_{r_1}, q)\right| \\ & \times \exp\left(j\frac{r_1}{r_1 + \Delta r}\phi\left(F_{\mathrm{R}}(m_r + \Delta m_{r_1}, q)\right)\right) \end{aligned} \tag{A.6}$$

where $\Delta m_{r_1} = \lfloor -\frac{2M_r S\Delta r}{c_0 f_{\mathrm{s}}} \rceil$, $\phi\left(F_{\mathrm{R}}(m_r + \Delta m_{r_1}, q)\right)$ is the phase of $F_{\mathrm{R}}(m_r + \Delta m_{r_1}, q)$.

Eq. (A.6) implies that we would obtain the new range spectrum $F_{\mathrm{R,\ new}}(m_r, q)$ by shifting Δm_{r_1} cells in the original range spectrum $F_{\mathrm{R}}(m_r, q)$, and then scale up or down its amplitude and phase with coefficient $\left(\frac{r_1}{r_1+\Delta r}\right)^2$, $\frac{r_1}{r_1+\Delta r}$ respectively. When there are multiple targets located at different range bins, we can filter out the spectrum of each target, and then implement the above transformations for each target.

A.2 Translating in Angle

Implementing the Range FFT and Angle FFT on the digitized de-chirped signal represented by Eq. (A.1), we obtain the 2D range-angle spectrum below:

$$\begin{aligned} F_{\mathrm{RA}}(m_r, m_\theta) &= \sum_{i=0}^{N_{\mathrm{s}}-1} \sum_{q=0}^{N_{\mathrm{Rx}}-1} f(\frac{i}{f_{\mathrm{s}}}, q) \exp(-j\frac{2\pi m_r i}{M_r}) \exp(-j\frac{2\pi m_\theta q}{M_\theta}) \\ &= F_{\mathrm{R}}(m_r) \sum_{q=0}^{N_{\mathrm{Rx}}-1} \exp(j2\pi\frac{qd\sin\theta_1}{\lambda}) \exp(-j\frac{2\pi m_\theta q}{M_\theta}) \\ &= F_{\mathrm{R}}(m_r) \frac{1-\exp\left(j2\pi N_{\mathrm{Rx}}\left(\frac{d\sin\theta_1}{\lambda}-\frac{m_\theta}{M_\theta}\right)\right)}{1-\exp\left(j2\pi\left(\frac{d\sin\theta_1}{\lambda}-\frac{m_\theta}{M_\theta}\right)\right)} \end{aligned} \tag{A.7}$$

where $F_{\mathrm{R}}(m_r) = \sum_{i=0}^{N_{\mathrm{s}}-1} A_1 \exp\left(j2\pi\left(f_{\mathrm{c}}\tau_1 + S\frac{i}{f_{\mathrm{s}}}\tau_1 - \frac{1}{2}S\tau_1^2\right)\right) \exp(-j\frac{2\pi m_r i}{M_r})$ is the Range FFT output for range bin m_r; M_θ is the number of points for Angle FFT; $m_\theta \in \{0, 1, ..., M_\theta - 1\}$ is the angle bin index. The mapping relationship between any azimuth angle θ and bin index is:

$$m_\theta = \lfloor \frac{M_\theta d\sin\theta}{\lambda} \rfloor \tag{A.8}$$

If the target at (r_1, θ_1) has a translation in angle, $\Delta\theta$, as we defined in Section 2.4, i.e. $r_1' = r_1, \theta_1' = \theta_1 + \Delta\theta$, and the antenna gain ratio due to angle shift is $\frac{G_{\theta_1'}}{G_\theta}$, then the corresponding FMCW de-chirped signal $f_{\mathrm{new}}(t, q)$ for the target at new location (r_1, θ_1') is expressed as:

$$f_{\mathrm{new}}(t, q) = A_1 \frac{G_{\theta_1'}}{G_\theta} \exp(j2\pi(f_{\mathrm{c}}\tau_1 + St\tau_1 - \frac{1}{2}S\tau_1^2 + \frac{qd\sin\theta_1'}{\lambda})) \tag{A.9}$$

Again, the range-angle spectrum for the digitized signal $f_{\mathrm{new}}(\frac{i}{f_s}, q)$ can be obtained with similar derivations in Eq. (A.7) and Eq. (A.8). That is,

$$F_{\mathrm{RA,\ new}}(m_r, m_\theta) = \frac{G_{\theta_1'}}{G_\theta} F_{\mathrm{R}}(m_r) \frac{1-\exp\left(j2\pi N_{\mathrm{Rx}}\left(\frac{d\sin\theta_1'}{\lambda}-\frac{m_\theta}{M_\theta}\right)\right)}{1-\exp\left(j2\pi\left(\frac{d\sin\theta_1'}{\lambda}-\frac{m_\theta}{M_\theta}\right)\right)} \tag{A.10}$$

With $\Delta m_{\theta_1} = \lfloor \frac{M_\theta d\left(\sin\theta_1 - \sin\theta_1'\right)}{\lambda} \rceil$, we can approximate new range-angle spectrum (A.10) with the old one (A.7) by following relation:

$$F_{\mathrm{RA,\ new}}(m_r, m_\theta) \approx \frac{G_{\theta_1'}}{G_\theta} F_{\mathrm{RA}}(m_r, m_\theta + \Delta m_{\theta_1}) \tag{A.11}$$

Eq. (A.11) implies that we would obtain the new range-angle spectrum $F_{\mathrm{RA,\ new}}(m_r, m_\theta)$ by shifting Δm_{θ_1} cells along the angular direction in the original spectrum $F_{\mathrm{RA}}(m_r, m_\theta)$. Similarly, when there are multiple targets located at different angle bins, we can also filter out the spectrum of each target, and then implement the above transformations for each target.

VITA

EDUCATION

- **Ph.D. in Electrical and Computer Engineering** 3/2020 - 9/2023
  University of Washington, Seattle, WA, USA

- **M.S. in Applied Mathematics** 3/2020 - 9/2023
  University of Washington, Seattle, WA, USA

- **M.S. in Electrical and Computer Engineering** 9/2018 - 3/2020
  University of Washington, Seattle, WA, USA

- **B.S. in Telecommunication Engineering** 8/2014 - 6/2018
  Xidian University, Xi'an, Shaanxi, China,

EMPLOYMENT

- **Research/Teaching Assistant** 9/2018 - 8/2023
  University of Washington, Seattle, WA, USA

- **Research & Development Internship** 6/2022 - 9/2022
  Intel, Santa Clara, CA, USA

- **AI & AD Research Internship** 6/2021 - 9/2021
  Volvo Cars, Sunnyvale, CA, USA

PUBLICATIONS

Journal papers

[J1] **X. Gao**, H. Liu, S. Roy, G. Xing, A. Alansari and Y. Luo. Learning to Detect Open Carry and Concealed Object with 77 GHz Radar. *IEEE Journal of Selected Topics in*

Signal Processing, 2022.

[J2] **X Gao**, S Roy, and G Xing. MIMO-SAR: A Hierarchical High-resolution Imaging Algorithm for mmWave FMCW Radar in Autonomous Driving. *IEEE Transactions on Vehicular Technology, 70(8):7322-7334*, 2021.

[J3] **X. Gao**, G. Xing, S. Roy, and H. Liu. RAMP-CNN: A Novel Neural Network for Enhanced Automotive Radar Object Recognition. *IEEE Sensors Journal, 21(4):5119-5132*, 2021.

Book chapters

[B1] S. Jin, **X. Gao**, and S Roy. Cognition in Automotive Radars, Next Generation Cognitive Radar Systems. *The Institution of Engineering and Technology (IET)*, 2021.

Conference papers

[C1] **X. Gao**, S. Roy, G. Xing, and S. Jin. Perception Through 2D-MIMO FMCW Automotive Radar Under Adverse Weather, *2021 IEEE International Conference on Autonomous Systems (ICAS), 1-5*, 2021.

[C2] Y. Wang, Z. Jiang, **X. Gao**, J. -N. Hwang, G. Xing, and H. Liu. RODNet: Radar Object Detection using Cross-Modal Supervision, *2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 504-513*, 2021.

[C3] **X. Gao**, G. Xing, S. Roy, and H. Liu. Experiments with mmWave Automotive Radar Test-bed, *2019 53rd Asilomar Conference on Signals, Systems, and Computers, 1-6*, 2019.

Patents

[P1] S. Ding, **X. Gao**, and HR. Dasari. Deformable Radar Polygon Systems and Methods for A Virtual Bumper, 2021.

INVITED PRESENTATIONS

- Mathworks, *Invited Virtual Talk*: PMCW Waveform and Its Use Cases, May, 2023.
- Mathworks, *Invited Virtual Talk*: Towards Millimeter-wave Radar Perception and Imaging, March 2023.
- Metawave, *Invited Virtual Talk*: Towards Millimeter-wave Radar Perception and Imaging, December, 2022.
- Intel, *Invited Virtual Talk*: A Radar Pipeline Design for Indoor DL Applications, September, 2022.
- NXP Semiconductors, *Invited Virtual Talk*: Research Proposal for Automotive Radars, June, 2022.
- Intel Labs, *Invited Virtual Talk*: Doppler-aided Radar Perception, March, 2022.
- Microsoft Research, *Invited Virtual Talk*: MIMO-SAR Hierarchical High-resolution Imaging Algorithm, February, 2022.
- Mitsubishi Electric Research Laboratories, *Invited Virtual Talk*: Radar-based Perception, December, 2021.
- UW, Ph.D. General Exam, *Invited Virtual Talk*: Towards Millimeter-wave Radar Signal Processing and Learning-based Applications, December, 2021.
- Volvo Cars, *Invited Virtual Talk*: Deformable Radar Polygon: A Lightweight Occupancy Representation for Collision Avoidance, September, 2021.
- Volvo Cars, *Tech. Sharing Seminar*: Deformable Radar Polygon for Virtual Bumper, September, 2021.

- IEEE International Conference on Autonomous Systems, *Invited Virtual Talk*: Perception Through 2D-MIMO FMCW Automotive Radar Under Adverse Weather, July, 2021.

- NeurIPS, ML4AD Workshop, *Invited Virtual Talk*: RAMP-CNN: A Novel Neural Network for Enhanced Automotive Radar Object Recognition, December, 2020.

- UW, Ph.D. Qualify Exam, *Invited Virtual Talk*: Radar Net for Object Recognition, May, 2020.

- Asilomar Conference on Signals, Systems, and Computers, *Invited Talk*: Experiments with Automotive Radar Test-bed, November, 2019.

- UW ECE Research Showcase, *Invited Poster*: Millimeter Wave FMCW Vehicle Radar: Testbed and Experiments, March, 2019.

PROFESSIONAL ACTIVITIES

- **Editorial Board of International Journals:**

  - Editorial Board Member for the Journal of Electrical and Electronic Engineering (JEEE), 2022-2024

- **Committee and Board Membership:**

  - Scientific Review Committee, Washington State Science & Engineering Fair, 2022
  - Program Committee for ML4AD Workshop, IJCAI, 2022
  - Program Committee for ML4AD Workshop, ICML, 2022
  - MS Admission Triage Committee, UW, 2021

- **Reviewer for International Journals and Books:**
  IEEE Open Journal of the Computer Society, IEEE Transactions on Intelligent Vehicles, American Association for the Advancement of Science, Information & Communications

Technology Express, IEEE Transactions on Vehicular Technology, IET Signal Processing, IEEE Access, IEEE Transactions on Intelligent Vehicles, International Journal of Communication Systems

- **Reviewer for International Conferences:**
  IEEE Conference on Vehicular Technology (VTC), IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), IEEE/CIC International Conference on Communications in China (ICCC), IEEE International Conference on Communications (ICC), International Joint Conference on Artificial Intelligence (IJCAI), International Conference on Machine Learning (ICML), International Symposium on Personal, Indoor and Mobile Radio Communications

- **Director for Graduate/Undergraduate Entrepreneurial Capstone Research:**
  - PACCAR, Camera and Radar Fusion for Object Detection, 2020
  - PACCAR, LiDAR and Radar Interference and Synchronization using Short Range Communication, 2020
  - PACCAR, Automotive Radar Imaging, 2019

RESEARCH MENTORING

- **Youchen Luo** (Undergraduate, CSE, UW) 3/2021 - 12/2022
  Project: Radar Concealed Object Detection, Automotive Radar Moving Object Recognition.

- **Ali Alansari** (MS, ECE, UW) 9/2020 - 4/2022
  Project: Radar Signal Processing, Radar Concealed Object Detection.

- **Zhichao Lei** (MS, ECE, UW) 1/2020 - 6/2020
  Project: Radar Camera Fusion for Object Tracking.

- **Yuchong Liu** (MS, ECE, UW) 1/2020 - 6/2020
  Project: Radar Camera Fusion for Object Tracking.

- **Tekhong Meach** (Undergraduate, ECE, UW) 2/2020 - 6/2020
  Project: Flash Lidar/Radar and Short Range Communication.

- **Mukund Bharadwaj** (MS, ECE, UW) 2/2020 - 6/2020
  Project: Flash Lidar/Radar and Short Range Communication.

- **Justin Tran** (Undergraduate, ECE, UW) 5/2019 - 4/2020
  Project: Radar Signal Processing, Data Annotation.

- **Matthew Sissel** (Undergraduate, ECE, UW) 1/2019 - 6/2019
  Project: Sensor Fusion with Automotive Radar.

- **Kat Mikhaylyuta** (Undergraduate, ECE, UW) 1/2019 - 6/2019
  Project: Sensor Fusion with Automotive Radar.

TEACHING

- **EE 505: Probability and Random Processes** 9/2022 - 1/2023