Title

Millimeter Wave Sensing and Its Applications in Vehicular Communication

Author

Ku, Hansol

Publication Date

2024

Peer reviewed|Thesis/dissertation

Millimeter Wave Sensing and Its Applications in Vehicular Communication

By

Hansol Ku

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Prasant Mohapatra, Chair

Omeed Momeni

Anh-Vu Pham

Committee in Charge

2024

To my heavenly father, who gave me the strength to complete my PhD,

and to my parents, Bon Il Ku and Yoon Hee Lee.

Contents

List of Figures

List of Tables

Abstract

Millimeter Wave Sensing and Its Applications in Vehicular Communication

Millimeter wave (mmWave) sensing has emerged as a powerful technology with a wide range of applications in areas that require high-resolution detection and precise distance measurements. Among various mmWave sensing techniques, Frequency Modulated Continuous Wave (FMCW) radar stands out due to its ability to accurately measure a target's distance, velocity, and angle. FMCW radar is especially useful in environments where reliability, speed, and spatial awareness are critical, such as autonomous driving. However, given the high sensitivity of mmWave signals to environmental factors, careful optimization is essential to ensure mmWave sensing is effective for vehicular communication.

In this dissertation, we present solutions for enhancing mmWave sensing using FMCW radar in vehicular communication. First, our work aims to establish reliable vehicle-to-vehicle (V2V) links for 6G communication by utilizing FMCW radar to optimize a beamforming protocol for IEEE 802.11ad. Experimental results show that our proposed protocol significantly improves throughput performance compared to the default 802.11ad protocol in real-world scenarios. Secondly, we utilize multiple FMCW radars to improve vehicular sensing for cooperative perception. We introduce the first vehicular cooperative dataset based on radar point clouds, V2VRP, developed with a novel experimental setup to synchronize multiple sensors. This dataset fills a critical gap in the real-world V2V cooperative perception field. Finally, we demonstrate that V2VRP dataset can be effectively utilized by the existing deep learning model. Our preliminary result based on our real-world dataset shows the cooperative perception approach offers a 70% improvement in maximum 3D IoU over the single perception approach.

Acknowledgments

First and foremost, I would like to thank Prof. Prasant Mohapatra for allowing me to pursue the mmWave project and for his belief in my abilities and knowledge of hardware. His financial support and mentorship have enabled me to focus entirely on my research and continue growing in this field. I am grateful for his patience, expertise, and dedication, which have significantly enriched my research experience and contributed to the successful completion of the project. I also thank Prof. Parth Pathak for providing direction for our research through our weekly meetings and insightful discussions. His guidance has been instrumental in helping me navigate challenges and refine my work.

I appreciate my dissertation committee members, Prof. Omeed Momeni and Prof. Anh-Vu Pham, for taking the time to read my dissertation and for providing valuable feedback and insights. Their guidance has been instrumental in shaping this dissertation, and I am grateful for their dedication and support throughout this process.

I thank my lab mates and friends in Davis who accompanied me throughout my PhD journey. Their encouragement and insights made the challenges more manageable and the successes more meaningful. I am grateful for the shared experiences, fruitful discussions, and emotional support that kept me motivated and focused, even during the most demanding times. From late nights in the lab to moments of celebration, their companionship made this journey an enjoyable and memorable one. Thanks to their support, I was able to navigate the challenges of the pandemic era without difficulty, finding comfort and inspiration in Davis community. Their resilience and positivity not only helped me stay on track but also reminded me of the strength we draw from each other in times of adversity. I am incredibly thankful for the friendships I have formed along the way, knowing that these bonds will last well beyond the walls of academia.

Lastly, I am profoundly grateful to my parents, Dr. Bon Il Ku and Mrs. Yoon Hee Lee, for their enduring support throughout my academic journey. Their unwavering belief in me has been a constant source of strength, motivating me to persevere even in the face of challenges. I also appreciate my sister, Yeowool Ku, and my brother-in-law, Josh Kim, for their encouragement and kind words, which lifted my spirits during difficult times.

My lovely nieces, Chloe and Clara Kim, have brought so much joy and laughter into my life, reminding me of the importance of balance and perspective. I am deeply grateful to my grandparents for the unconditional love they have always shown me. I am incredibly fortunate to be surrounded by such a loving and supportive family. I dedicate this work to them as a testament to their influence and guidance in my life.

Chapter 1

Introduction

Vehicle-to-vehicle (V2V) communication is essential for improving road safety by allowing vehicles to share crucial information about their environment, speed, and driver's actions. Through real-time data exchange, V2V communication enables vehicles to better predict and react to potential hazards, helping to reduce collisions and improve traffic flow. According to the National Highway Traffic Safety Administration (NHTSA), annual motor vehicle traffic fatalities have risen from 32,893 in 2013 to 40,990 in 2023 [6]. By integrating V2V technology, vehicles can enhance situational awareness, contributing to safer highways and potentially reducing traffic accidents and injuries.

In V2V communication, Frequency Modulated Continuous Wave (FMCW) radar operating in the millimeter wave (mmWave) spectrum is a valuable tool for enhancing automobile safety, situational awareness, and communication efficiency [7–10]. At mmWave frequencies, FMCW radar enables precise measurements of distance, velocity, and angle, allowing vehicles to detect nearby vehicles, pedestrians, and infrastructure effectively. LiDAR (Light Detection and Ranging) provides high-resolution 3D point clouds by emitting laser pulses and measuring the reflected light, making it highly effective for detailed object detection and distance estimation [11–22]. However, LiDAR systems are expensive, require significant computational resources due to large data outputs, and are sensitive to adverse weather conditions like fog, rain, and snow, where light-based measurements can become inaccurate or even fail entirely. Cameras, on the other hand, capture detailed visual information in 2D, allowing for the identification of colors, textures, and

specific object types, which are useful for tasks like lane detection and traffic sign recognition [23–29]. While more affordable than LiDAR, cameras rely on ambient light and can struggle in low-light conditions, and their 2D nature makes it challenging to measure depth information accurately.

Unlike LiDAR and cameras, FMCW radar offers a cost-effective, compact, and weather-resilient alternative for mmWave sensing. Leveraging mmWave FMCW radar enables vehicular communication systems to achieve higher safety and performance, paving the way for scalable and reliable V2V communication solutions in autonomous driving. However, due to the short wavelength of mmWave signals, they experience high path loss and are easily blocked by obstacles such as trees and other vehicles. In V2V communication, the quality of mmWave signals degrades further in dynamic scenarios due to beam misalignment caused by their highly directional, narrow beam patterns. This narrow beam pattern also presents challenges for FMCW radars, as it can create sparse point clouds due to specular reflection. Additionally, radars are susceptible to noise from background clutter and multi-path effects, which can compromise the accuracy of vehicular perception.

This thesis is organized to address the challenges associated with highly sensitive mmWave signals and resolve them for optimized vehicular communication and environmental perception. Chapter 2 proposes real-time radar-assisted beamforming in mmWave technology for vehicular communication, emphasizing the advantages and challenges of using FMCW radar to enhance V2V connectivity. Chapter 3 introduces a real-world V2V dataset based on the FMCW radar point cloud, V2VRP, designed to facilitate cooperative environmental perception. Chapter 4 evaluates the V2VRP dataset using a 3D bounding box estimation model, analyzing performance metrics such as Intersection over Union (IoU) in real-world scenarios. Finally, Chapter 5 concludes the dissertation.

1.1 Objectives

This report addresses the challenges of using FMCW radar for vehicular communication. Our objectives are outlined below.

- **Propose a radar-assisted real-time beamforming protocol.** Current 6G com-

munication protocols, such as IEEE 802.11ad, rely on exhaustive beamforming, which leads to significant overhead. Our approach aims to reduce this overhead by integrating radar-assisted techniques to achieve more efficient and responsive beamforming, improving throughput and connection stability in dynamic environments.

- **Introduce a cooperative perception dataset using FMCW radar point clouds.** There are few datasets available for cooperative perception in autonomous driving, and those that do exist primarily rely on LiDAR, which is costly, has high transmission cost due to large file size, and is less reliable in adverse weather conditions like rain and snow. To address this gap, we aim to develop a dataset specifically designed for cooperative perception tasks using cost-effective, lightweight, and weather-resilient FMCW radar.

- **Demonstrate the feasibility of using the cooperative perception dataset with existing machine learning models for 3D perception.** Since V2VRP is the first dataset of its kind for cooperative radar point clouds, it has not yet been fully evaluated. We aim to validate V2VRP's effectiveness by applying it to 3D perception tasks with an established deep learning model.

1.2 Contributions

The contributions of each chapter are summarized below to achieve our objectives.

Chapter 2: Radar-assisted Beamforming Protocol

- We implement a radar-assisted beamforming protocol, which can be done in real-time with Commercial Off-The-Shelf (COTS) devices.

- We provide the preliminary results of real-world scenarios by comparing the default 802.11ad protocol and our proposed radar-assisted beamforming protocol. To the best of our knowledge, there are few experimental observations regarding real-time radar-assisted beamforming for V2V communication.

- We experimentally show that our protocol's throughput performance can be improved by 66% compared to the default 802.11ad protocol in highway scenario.

Chapter 3: V2VRP: Cooperative autonomous driving dataset

- We introduce the first large-scale radar-based cooperative perception (V2VRP) dataset tailored for V2V applications to the best of our knowledge.
- We present a novel, low-cost experimental setup using GPS-RTK, camera, and FMCW radar, demonstrating how to combine multiple sensors' data for cooperative perception.
- We provide a comprehensive radar point cloud dataset featuring over 39K annotated 3D bounding boxes for cars in the scene, along with synchronized radar point cloud frames, camera images, and GPS coordinates along the driving routes.
- By tailoring our dataset for V2V communication, it supports and enhances research in cooperative perception, a crucial aspect of autonomous driving that allows connected autonomous vehicles to share sensory data for improved environmental awareness and safety.

Chapter 4: 3D bounding box estimation with V2VRP

- We demonstrate that V2VRP can be effectively used with existing machine learning models for 3D bounding box estimation.
- We present preliminary results for 3D bounding box estimation using our cooperative dataset, achieving a maximum 3D IoU of 0.67 and an mAP score of 0.59 at an IoU threshold of 0.1.
- We show compared results for using single and combined view data samples and show combined view samples outperform single view by 70% in maximum 3D IoU.

1.3 Background: Frequency Modulated Continuous Wave Radar

Frequency Modulated Continuous Wave (FMCW) radars are widely used for accurately detecting a target's distance, velocity, and angle, making them invaluable in applications

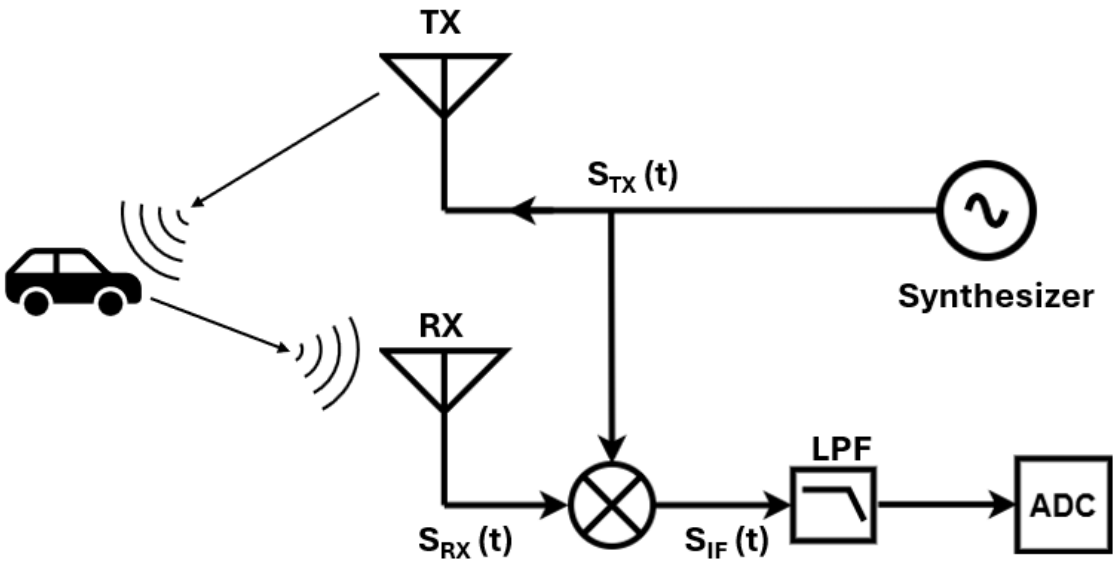
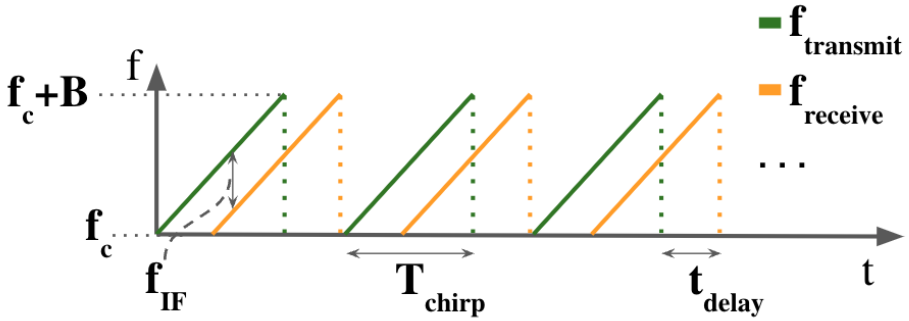


Figure 1.1: System block diagram of an FMCW radar. $S_{RX}(t)$, $S_{TX}(t)$, and $S_{IF}(t)$ are the received signal, transmitted signal, and IF signal, respectively. Also, LPF is a low pass filter.

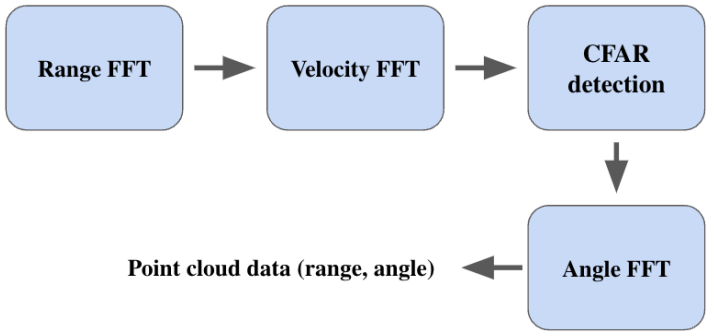
like automotive safety, industrial automation, and robotics [7, 30–39]. By transmitting a continuously modulated frequency, FMCW radars measure range based on the frequency difference between transmitted and received signals, known as the beat frequency or the intermediate (IF) frequency, which indicates the time delay and, thus, the distance to the target. FMCW radars are widely used to detect a target's distance, velocity, and angle. It transmits the signal whose frequency linearly increases with time. This signal is known as a chirp, and a series of chirps are transmitted over one frame. The frequency of a transmitted chirp is given by

$$\begin{aligned} f_{transmit} &= f_c + \frac{B}{T_{chirp}} t \\ &= f_c + St, \quad 0 \leq t \leq T_{chirp} \end{aligned} \tag{1.1}$$

where f_c is the starting frequency, B is the RF bandwidth of the FMCW radar, T_{chirp} is the time duration of a chirp, and S is the slope of a chirp with time t (see Fig. 1.2a). We explain the details for obtaining the range, velocity, and angle parameters below.



(a) FMCW signal of multiple chirps



(b) FMCW radar signal processing chain

Figure 1.2: The frequency vs. time illustration of the FMCW signal and basic radar signal processing flow chart.

1.3.1 Range Estimation

To estimate the range of an object detected by an FMCW radar, it is essential to analyze both the transmitted and received signals from the object. The transmitted signal can be modeled as [40, 41]:

$$S_{TX}(t) = \cos\left(2\pi(f_c t + \frac{St^2}{2})\right) \tag{1.2}$$

When the chirp is reflected back from an object, it is captured by the radar receivers. The received signal is a delayed version of the transmitted signal and can be written as:

$$S_{RX}(t) = \cos\left(2\pi(f_c(t - t_{delay}) + \frac{S(t - t_{delay})^2}{2})\right) \tag{1.3}$$

where t_{delay} is the round-trip delay from the radar to the target object.

The received signal is then combined with the transmitted signal in the mixer, resulting in an intermediate frequency (f_{IF}) signal.

$$S_{IF}(t) = \cos\left(2\pi(f_c t_{delay} + S t_{delay} t - \frac{S(t_{delay})^2}{2})\right) \tag{1.4}$$

This frequency is proportional to the round-trip delay from the radar to the target object so that it can be written as

$$f_{IF} = S t_{delay} = S\frac{2d}{c} \tag{1.5}$$

where d is the range between the radar and the target, and c is the speed of light. After the low pass filter, the beat signal becomes

$$\begin{aligned} S_{IF}(t) &= \frac{1}{2}\cos\left(2\pi(f_c t_{delay})\right) \\ &= \frac{1}{2}\cos\left(2\pi(f_c\frac{2d}{c})\right) \end{aligned} \tag{1.6}$$

To convert the intermediate frequency from the time domain to the frequency domain, a fast Fourier transform (FFT) is applied to the signal (see Fig. 1.2b). Then, we can estimate the range profile from the detected peaks from the FFT process. The range resolution is related to the RF bandwidth (B) and is given by

$$d_{resolution} = \frac{c}{2B} \tag{1.7}$$

1.3.2 Velocity Estimation

The velocity of an object is calculated by examining the phase shift ($\Delta\phi$) between consecutive chirps. By looking at the phase difference between each pair of consecutive chirps, the radar can determine how much the target moved during the interval between these chirps. If there is a change of distance (Δd) due to the movement of an object, the resulting phase shift is

$$\begin{aligned} \Delta\phi &= 2\pi f_c\frac{2\Delta d}{c} \\ &= \frac{4\pi v T_{chirp}}{\lambda} \end{aligned} \tag{1.8}$$

where v is the velocity of the object, and λ is the wavelength. Thus, velocity can be rewritten as

$$v = \frac{\lambda \Delta \phi}{4\pi T_{chirp}} \tag{1.9}$$

The velocity resolution is given by

$$v_{resolution} = \frac{\lambda}{2N_{chirp}T_{chirp}} \tag{1.10}$$

where N_{chirp} is the number of chirps in one frame. After acquiring range-velocity data from range and velocity FFT, the Constant False Alarm Rate (CFAR) threshold [42–44] is applied to obtain robust detection in the presence of noise and clutter.

1.3.3 Angle Estimation

The azimuth angle is determined by performing an FFT across multiple receiver antennas in the radar array. When a reflected signal returns from a target, it reaches each receiver antenna at slightly different times due to the physical spacing between the antennas. This spacing, or differential distance, between neighboring antennas, causes a phase difference in the received signals. That small distance change results in a phase change ($\Delta\Phi$) in the FFT peak for each antenna and can be written as

$$\Delta\Phi = \frac{2\pi l \sin\theta}{\lambda} \tag{1.11}$$

where l is the distance between receive antennas, and θ is the azimuth angle of the object.

For example, in Fig. 1.3, we use one TX and two RX antennas. The signal from the TX1 antenna is reflected from a target at an angle θ with regard to the radar and is received at both RX antennas. The signal from the target must travel an additional distance, $l \sin\theta$, to reach the RX2 antenna. This corresponds to a phase difference ($\Delta\Phi$) shown in Eq. 1.11. Then, the angle of arrival can be calculated as

$$\theta = \sin^{-1} \frac{\Delta\Phi\lambda}{2\pi l} \tag{1.12}$$

Since $\Delta\Phi$ can be estimated only in the range $(-\pi, \pi)$ for the unambiguous field of view (FOV), we design the spacing between RX antennas with $\frac{\lambda}{2}$ to give the maximum FOV (see Fig. 1.4a). The FOV with the $l = \frac{\lambda}{2}$ can be written as

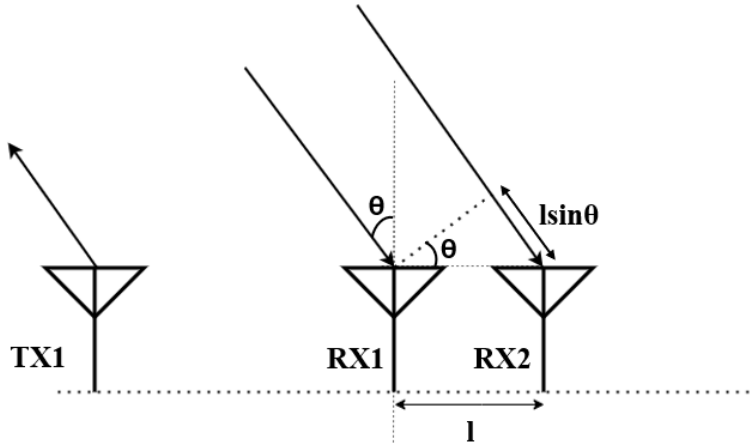


Figure 1.3: Angle of arrival estimation using one TX and two RX antennas

$$\begin{aligned} \theta_{FOV} &= \pm \sin^{-1}(\frac{\lambda}{2l}) \\ &= \pm 90^{\circ} \end{aligned} \tag{1.13}$$

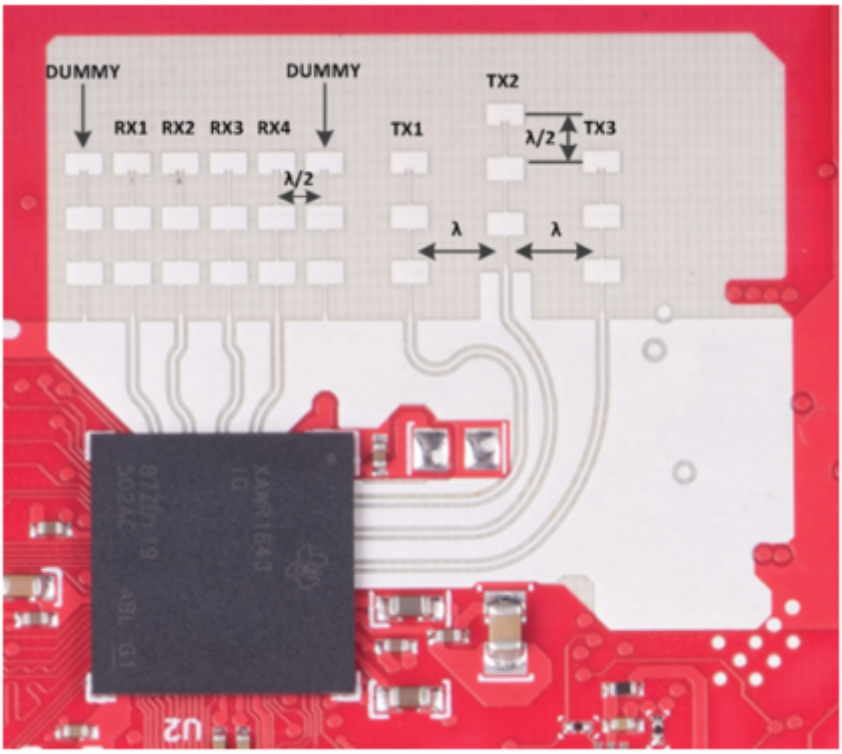
The angle resolution of an FMCW radar refers to its ability to distinguish between two targets that are close together in angle (i.e., direction relative to the radar). It determines how precisely the radar can measure the direction of incoming signals from multiple objects. Angle resolution depends primarily on two factors: the antenna array's size and the wavelength of the radar signal (see Fig. 1.4a). The angle resolution is given by

$$\theta_{resolution} = \frac{\lambda}{N_{Rx} l} \tag{1.14}$$

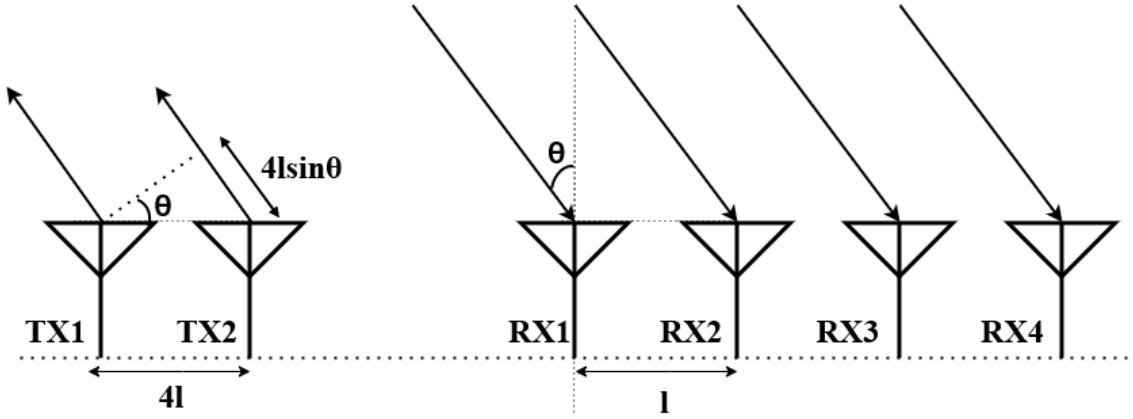
for a bore sight view where N_{Rx} is the number of receive antennas. Higher-frequency radar signals (which have shorter wavelengths) and wider antenna arrays both improve the radar's angle resolution, enabling it to better distinguish between objects that are close together in angle.

1.3.4 Multiple-Input Multiple-Output Radar

Multiple-input multiple-output (MIMO) radar is a radar technology that utilizes multiple transmit and receive antennas to improve detection, resolution, and accuracy in various applications. If there is a single transmitter antenna, it is referred to as a single-input



(a) Example of antenna layout for MIMO configuration [45].



(b) Angle of arrival estimation using two TX and four RX antennas resulting in eight virtual antennas. Due to the phase differences between multiple TX and multiple RX antennas, the angle resolution of the radar can be improved.

Figure 1.4: Antenna layout for MIMO radar and the principle of MIMO radar

multiple-output (SIMO) radar. The angle resolution of this type of radar depends on the number of RX (receive) antennas (see Eqn. 1.14). For instance, a device with four RX antennas achieves an angle resolution of approximately 30°, while a device with eight RX antennas improves the angle resolution to around 15°. Thus, a straightforward way to enhance angle resolution is by increasing the number of RX antennas. However, this approach has limitations, as each additional RX antenna necessitates a separate RX processing chain within the device, adding complexity and resource requirements.

Unlike traditional radar systems, which often use a single antenna for transmitting and receiving signals or a small number of elements in phased array configurations, MIMO radar transmits multiple independent waveforms from multiple antennas. This setup allows for advanced signal processing, forming virtual antenna arrays that significantly enhance spatial resolution. For instance, as shown in Fig. 1.4b, when examining phase differences across four RX antennas, signals transmitted from TX1 have phase differences of 0, $\Delta\Phi$, $2\Delta\Phi$, and $3\Delta\Phi$ relative to RX1. In contrast, signals from TX2—positioned at two wavelengths ($2\lambda = 4l$) from TX1—produce phase differences of $4\Delta\Phi$, $5\Delta\Phi$, $6\Delta\Phi$, and $7\Delta\Phi$ at each RX antenna. This additional phase shift results from the extra distance traveled by TX2's signal, equivalent to $4l \sin\theta$, providing MIMO radar with finer angular resolution through precise phase measurements. This configuration effectively creates the equivalent of a single TX with eight virtual RX antennas.

MIMO radar has numerous advantages, including improved target detection, enhanced angle and range resolution, and increased robustness to interference and signal fading. These attributes make MIMO radar especially suitable for applications in automotive safety, where accurate detection of multiple objects is essential, as well as in defense, surveillance, and weather monitoring, where precise resolution and tracking of targets are critical. With the growing demand for high-resolution radar systems in autonomous vehicles, industrial automation, and security, MIMO radar is becoming a key technology in advanced sensing systems, offering reliable performance in diverse and challenging environments.

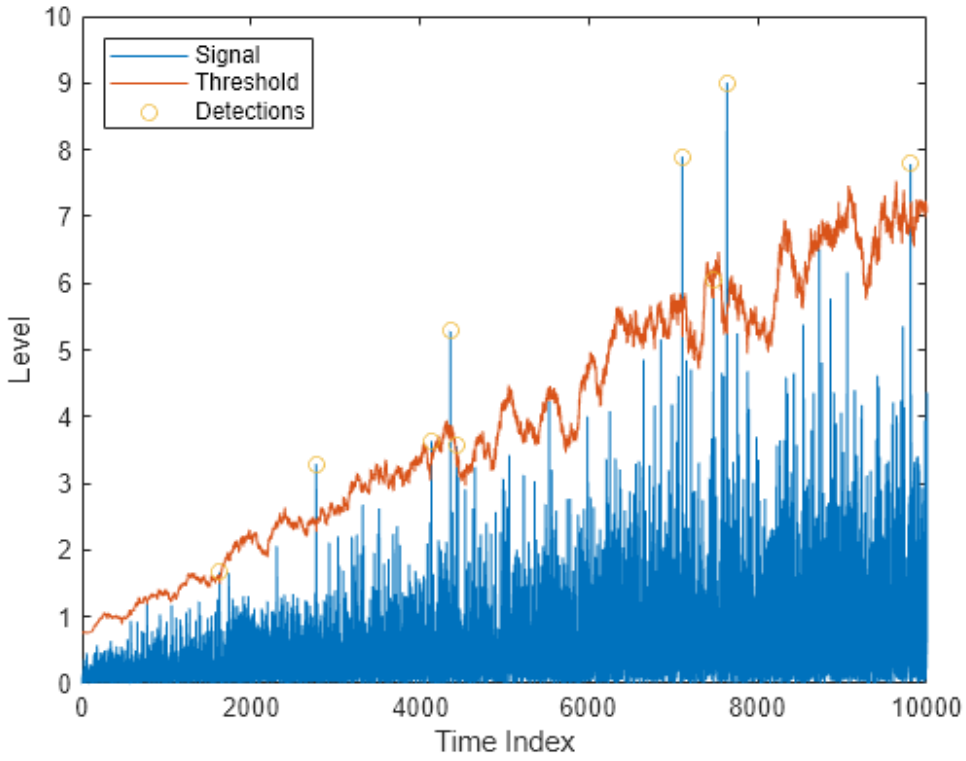


Figure 1.5: Example of CFAR detection [1]. The threshold rises with signal noise to keep a constant false alarm rate, and detections happen when the signal level goes above this threshold.

1.3.5 Constant False Alarm Rate

Constant False Alarm Rate detection is a fundamental technique used in radar and signal processing to detect the presence of objects within a noisy background [42]. The primary objective of CFAR detection is to maintain a constant probability of false alarms in varying noise conditions. This is crucial in radar systems, where environmental noise levels can fluctuate due to factors like weather, clutter, or interference. By setting a threshold that adapts based on the surrounding noise, CFAR detection ensures that false detections (i.e., alarms triggered by noise rather than an actual target) remain consistent, allowing operators to detect true signals more reliably.

One of the common CFAR techniques is CA-CFAR (Cell-Averaging Constant False Alarm Rate), which is a widely used adaptive thresholding technique in radar signal processing designed to maintain a constant false alarm rate despite variations in noise

and interference levels. CA-CFAR works by dynamically adjusting the detection threshold based on the average noise power of cells surrounding a target cell in the radar's range-Doppler map (see Fig. 1.3.5). In this approach, the radar first selects a group of reference cells (guard cells) around the cell under test (the target cell) to estimate the noise level. The average noise power in these reference cells is then used to calculate a threshold. If the power in the target cell exceeds this adaptive threshold, it is classified as a detection, indicating the presence of a potential target. CA-CFAR is effective in environments with uniform or Gaussian noise backgrounds, as it adapts to local noise variations while avoiding excessive false alarms.

Applying CFAR (Constant False Alarm Rate) processing to radar point clouds is essential for accurately detecting and filtering out noise, enabling reliable identification of objects within the radar's field of view. Radar point clouds are collections of data points that represent detected objects and their attributes, such as range, velocity, and angle. However, these point clouds often include noise and clutter due to environmental factors, reflections, and multipath effects. CFAR algorithms, such as CA-CFAR (Cell-Averaging CFAR) and OS-CFAR (Ordered Statistic CFAR), are used to set adaptive detection thresholds across the radar's point cloud data by evaluating each data point in relation to its surrounding cells. By dynamically adjusting the threshold based on the noise level estimated from neighboring points, CFAR helps discriminate between true target detections and noise or clutter, maintaining a constant false alarm rate regardless of environmental conditions. This approach is particularly useful in automotive radar, where the detection of multiple moving objects in complex scenarios, such as urban settings with pedestrians and other vehicles, requires robust and adaptive thresholding. Applying CFAR to radar point clouds enhances object detection accuracy and contributes to better situational awareness in systems like autonomous vehicles and advanced driver-assistance systems (ADAS).

Chapter 2

Characterizing Real-time Radar-assisted Beamforming in Millimeter Wave Vehicle-to-Vehicle Links

Due to the increasing demand for high data rates for applications such as virtual reality, 8K video streaming, and autonomous driving, the mmWave 6G wireless communication is considered a promising solution to support those applications since it can provide multi-gigabyte per second data rate [46–55]. However, mmWave wireless communication suffers from high propagation loss due to its directional link towards the receiver using the 60 GHz band for a carrier frequency. One of the key applications of 6G wireless communication will be in the realm of V2V communication [56]. Vehicle-to-vehicle (V2V) communication enables vehicles to share information with each other, such as speed, location, and road conditions, in real-time. With 6G wireless communication, V2V technology is expected to become even more advanced and reliable. It will enable faster and more efficient data transfer between vehicles, allowing for better coordination and decision-making on the road.

Mobility is one of the significant challenges of mmWave networks in V2V [57, 58]. When mmWave devices in high-speed movement perform data transmission, the narrow and directional beam pairs will be misaligned due to changes in relative positions. The

802.11ad protocol passively adjusts the beam pairing, so when the current beam can no longer meet the communication needs, it will sweep sectors and match new pairs, which is a very high time cost alignment [59–61]. Such a setting cannot satisfy high-speed movement while maintaining high-quality communication quality in vehicular networks. Thus, the mmWave communication needs to be highly reliable and efficient under mobile scenarios to support the wireless communication between connected autonomous vehicles [62].

Various sensors, such as light detection and ranging (LIDAR), cameras, and mmWave radar, have been integrated into vehicles to support advanced driver-assistance systems (ADAS) and autonomous driving systems [63, 64]. These sensors provide useful sensory data that can be used to locate the transmitter/receiver of the vehicles/infrastructure, as well as to support the efficient beamforming process in mmWave communication.

In [65], LIDAR attached to the stationary base station with a mmWave receiver collected the LIDAR image data of a moving vehicle with a mmWave transmitter. Those LIDAR images were fed into the deep learning model to predict current and future beam pairs for mmWave communication. Similarly, an RGB camera was installed on the base station containing a receiver to capture the moving vehicle equipped with a transmitter. The captured RGB images were utilized by the recurrent neural network (RNN) to predict the candidates for optimal beam pairs with high receive beam power [66]. In [67], frequency-modulated continuous wave (FMCW) radar mounted on the stationary base station was used to collect sensory data containing the range, velocity, and angle of arrival (AoA) of detected objects. This work used a machine learning approach using radar data as input to estimate the top candidates for optimal beam pairs.

However, LIDAR and cameras can be highly ineffective under bad weather conditions such as heavy rain, snow, fog, or obstacles masking the sensors. Furthermore, [65], [66], and [67] are all based on the offline beam prediction, which may not be suitable for dynamic scenarios such as two moving vehicles to establish a mmWave link with the best beam pair in real-time. Also, they are limited to vehicle-to-infrastructure (V2I) scenarios where the receiver is fixed at the base station.

In this work [68], we study the performance of the real-time FMCW radar-assisted

beamforming with Commercial Off-The-Shelf (COTS) radar and 802.11ad devices for V2V applications. The main contributions are listed below:

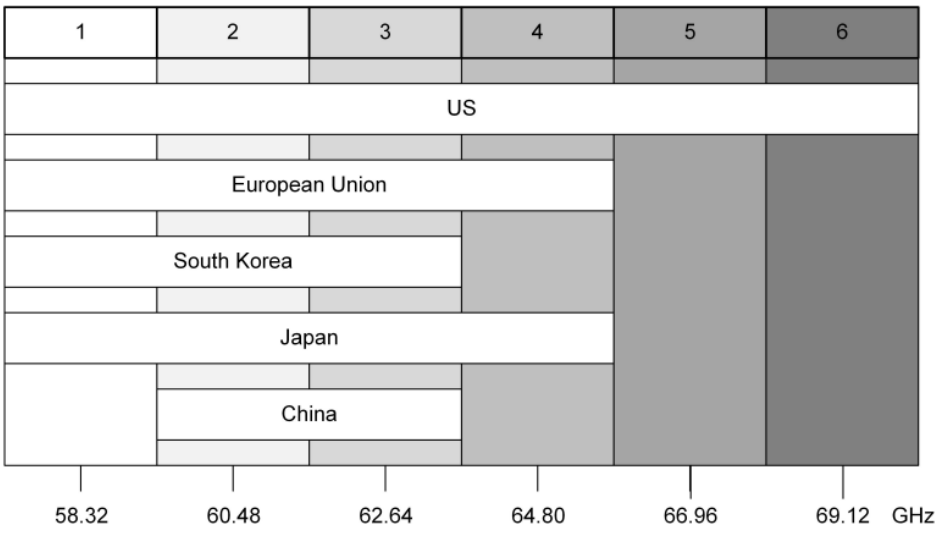
- We implement a radar-assisted beamforming protocol, which can be done in real-time with COTS devices.
- We provide the preliminary results of real-world scenarios by comparing the default 802.11ad protocol and our proposed radar-assisted beamforming protocol. To the best of our knowledge, there are few experimental observations regarding real-time radar-assisted beamforming for V2V communication.
- We experimentally show that the throughput performance can be improved with our protocol compared to the default 802.11ad protocol.

2.1 Preliminaries

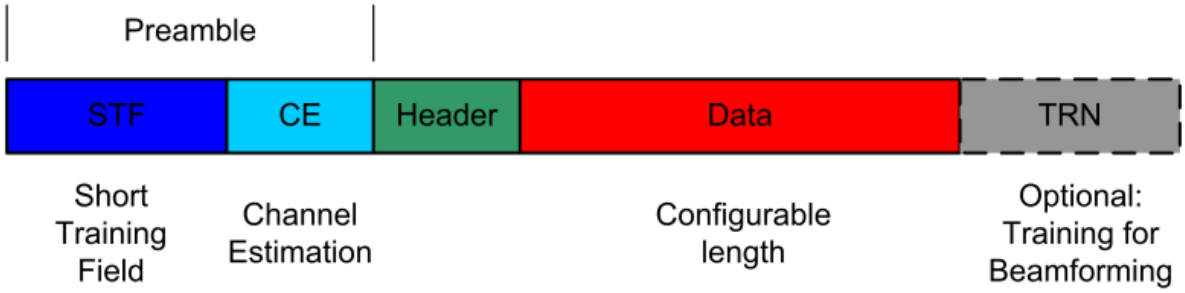
2.1.1 IEEE 802.11ad Protocol

The 802.11ad protocol is a wireless communication standard developed for the 60 GHz frequency band with a bandwidth of 2.16 GHz (59.40-61.56 GHz), as shown in Fig. 2.1a. This protocol is designed to support high-speed data transfer rates of up to 7 Gbps over short distances, typically within a room or a short-range line-of-sight (LOS) connection. Based on the design of phased array antennas, the 802.11ad can focus wireless radio signals in a specific direction, increasing transmission range and reducing interference from other devices.

A packet consists of several common parts: the Preamble, Header, Data, and an optional TRN field (see Fig. 2.1b). The preamble includes the short training field (STF) and channel estimation (CE) field, both essential in every packet. It assists the receiver with automatic gain control (AGC), packet recognition, and frequency offset estimation, and it indicates the PHY type (SC or OFDM) being used. The known CE field also enables the receiver to estimate the channel. The header, specific to each PHY, provides additional critical information for the receiver, including modulation mode (MCS), data field length, and a checksum. The data section transmits the actual payload with varying



(a) Channels in the 60 GHz band

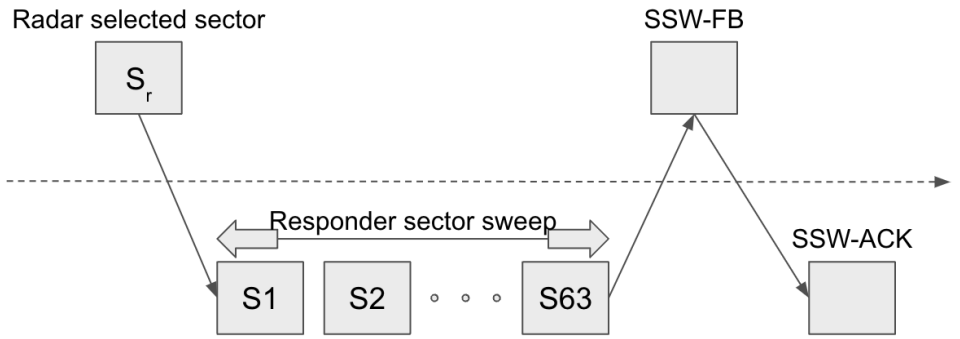


(b) A general packet structure in IEEE 802.11ad

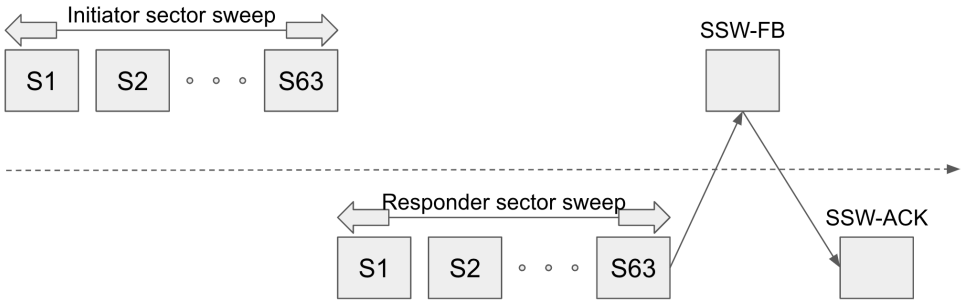
Figure 2.1: The 60 GHz channel allocations and the packet structure in IEEE 802.11ad [2].

lengths and modulation schemes (MCS). Finally, the optional TRN field can be appended to packets, allowing for beamforming optimization.

In the 802.11ad standard, beamforming is essential to ensure that both the transmitter and receiver antennas are optimally aligned for maximum signal strength and transmission efficiency. Beamforming allows the system to direct the signal narrowly toward the receiver, compensating for the significant path loss typically experienced at 60 GHz. To achieve this in Commercial Off-The-Shelf (COTS) devices, three protocols based on 802.11ad are commonly implemented: MID (Multiple Sector ID capture), MID-BC (Beam Combining), and no-RxBF (no Receiver Beamforming) [69]. The MID and MID-BC protocols implement receiver-side beamforming (RxBF) to optimize directional alignment and improve signal quality. The MID protocol uses an exhaustive search to find the optimal



(a) Radar selects the best sector for the Tx in SLS



(b) Default SLS phase: Two sector sweeps, SSW-FB, SSW-ACK

Figure 2.2: Radar aided SLS vs. default SLS.

sector pair for data transmission, where the transmitter scans multiple sectors until the strongest signal is identified. This approach ensures the highest link quality. MID-BC, on the other hand, enhances this process with a beam-combining phase, allowing up to seven sectors to be selected and tested in 49 combinations to find the best pair, maximizing signal alignment for greater reliability. The no-RxBF option skips receiver beamforming, relying solely on the transmitter for directionality. While this may result in lower link performance, it simplifies the setup for applications where ultra-high performance isn't essential. MID and MID-BC provide higher quality and reliability in mmWave communications, particularly suited for applications that demand robust, high-throughput data links.

As shown in Fig. 2.2a, at the beginning of sector-level sweep (SLS), the side that raises this training is called the initiator, and the accepter is called the responder. Therefore, the SLS phase includes the initiator's sector sweep, the responder's sector sweep, sector sweep feedback (SSW-FB), and sector sweep acknowledgment (SSW-ACK), where the

sector selection of both sides in this training is exchanged. Then, after the MID and BC phases, the best sector selections for the initiator and responder can be determined.

2.1.2 Density-based Spatial Clustering of Applications with Noise

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a popular unsupervised machine learning algorithm used for clustering spatial data based on density [70]. Unlike other clustering algorithms such as k-means [71, 72], DBSCAN does not require specifying the number of clusters in advance. Instead, it identifies clusters by finding densely packed groups of data points, where points within a specified distance, known as *epsilon*, have a minimum number of neighbors defined by another parameter, *minPts*. The algorithm for DBSCAN is shown in Algorithm 1.

This density-based approach allows DBSCAN to efficiently identify clusters of arbitrary shape, making it especially useful for applications like FMCW radar point clouds, where clusters may not be spherical or evenly distributed [74]. One of DBSCAN's main advantages is its ability to recognize and label noisy data points, which is crucial for radar point cloud data. Points that do not belong to any dense region (i.e., lack enough neighbors within the specified epsilon radius) are marked as outliers. This noise tolerance enables DBSCAN to handle real-world scenarios with irregular data distributions or background noise, as is often the case with radar data in complex environments. For instance, in FMCW radar applications, DBSCAN can help detect objects by forming clusters of reflective points, while isolated or sparse reflections, such as background noise, are ignored as outliers. This makes DBSCAN a robust tool for distinguishing significant objects from noise in radar-based perception systems.

2.2 Related Work

2.2.1 Millimeter Wave for Wireless Communication

In recent years, mmWave communication research has focused on directional beamforming, beam alignments [69,75], and susceptibility to blockage [76,77], which comes with developing new techniques and protocols including MIMO beamforming algorithms [78,79], and interference [77,80]. Additionally, there has been a growing interest in using machine

Algorithm 1 DBSCAN Clustering [70, 73]

Input: D: a set of data points, $distFunc$: a distance function, ϵ: radius of a neighborhood, $minPts$: the minimum number of points required to form a cluster

$C \leftarrow 0$ ▷ Initialize cluster counter
for each point P in D **do**
  if $label(P) \neq undefined$ **then**
    continue ▷ Skip if already processed
  end if
  $N \leftarrow RangeQuery(D, distFunc, P, \epsilon)$ ▷ Find neighbors of P
  if $|N| < minPts$ **then** ▷ Density check
    $label(P) \leftarrow Noise$ ▷ Label P as noise
    continue
  end if
  $C \leftarrow C + 1$ ▷ Increment cluster label
  $label(P) \leftarrow C$ ▷ Label initial point
  $S \leftarrow N \setminus \{P\}$ ▷ Neighbors to expand
  for each point Q in S **do**
    if $label(Q) = Noise$ **then**
      $label(Q) \leftarrow C$ ▷ Change noise to border point
    end if
    if $label(Q) \neq undefined$ **then**
      continue ▷ Skip if previously processed
    end if
    $label(Q) \leftarrow C$ ▷ Label neighbor
    $N \leftarrow RangeQuery(D, distFunc, Q, \epsilon)$ ▷ Find neighbors of Q
    if $|N| \geq minPts$ **then** ▷ Density check (if Q is a core point)
      $S \leftarrow S \cup N$ ▷ Add new neighbors to seed set
    end if
  end for
end for

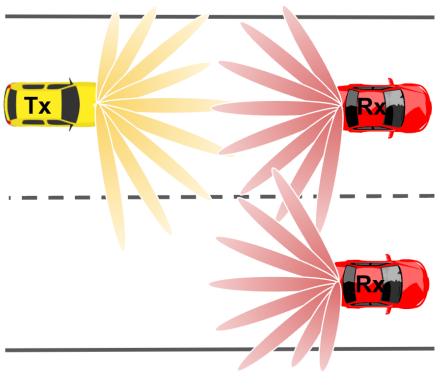
Algorithm 2 Function: *RangeQuery* [70, 73]

```
1: function RangeQuery(D, distFunc, Q, ϵ)
2:     N ← empty list
3:     for each point P in D do                    ▷ Scan all points in the database
4:         if distFunc(Q, P) ≤ ϵ then              ▷ Compute distance and check ϵ
5:             N ← N ∪ {P}                         ▷ Add P to result
6:         end if
7:     end for
8:     return N
9: end function
```

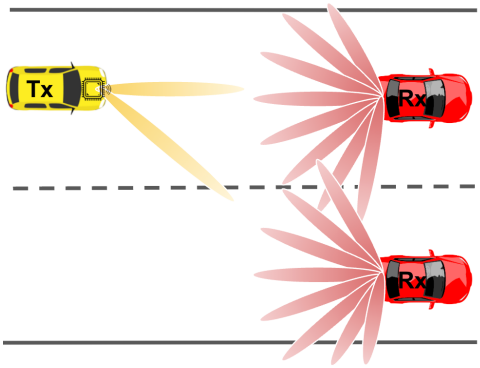
learning techniques for mmWave communication, such as channel estimation and beam prediction [81, 82]. However, these works mostly happen indoors and take advantage of the signal reflection. Our work practices the outdoor measurements and evaluation in the real world and precisely implements the beam patterns. Furthermore, most papers may not modify the beamforming training significantly at the protocol design level. Our proposal changes the workflow of the SLS to reduce the training overhead from the low-level design perspective.

2.2.2 Vehicular Millimeter Wave Beam Alignment

Specific to vehicular environments, mmWave communication research has gained diverse solutions to provide reliable wireless communication in high-speed mobility based on different assistance, such as machine learning [83, 84], roadside units (RSUs) [85, 86], camera [87,88], or radar [89–91]. Among the given solutions, the design of RSUs is impractical, which may require high construction costs and extra software implementation. Compared with these solutions using cameras or radars, our solution is more affordable and practiced nowadays. At the same time, our radar accurately detects the target vehicle, which can provide sufficient detection information for a reliable beam alignment. Besides that, our design works as a real-time processing system, and it supports instant beam alignment with real-world data input, which is more robust and practical than previous work. Also,



(a) Exhaustive beam searching of 802.11ad protocol for V2V scenario



(b) Radar-aided beam searching for V2V scenario

Figure 2.3: Comparison between beam training of IEEE 802.11ad protocol and radar-aided mmWave communication protocol. The latter approach avoids the exhaustive search of every possible beam direction.

it does not load heavy-weight machine learning models into communication devices, which is more lightweight and energy efficient.

2.3 System Overview

In this work, we attach COTS devices on two vehicles to test mmWave communication. One vehicle was set up as a transmitter (Tx), and the other was used as a receiver (Rx). The V2V scenarios based on the IEEE 802.11ad protocol and our proposed radar-aided protocol are illustrated in Fig. 2.3. For our protocol, the yellow vehicle (carrying Tx) is configured as a client mode with an FMCW radar attached, and the red vehicle (carrying Rx) is set as an Access Point (AP) (see Fig. 2.3b). Fig. 2.4 shows that our protocol can avoid exhaustive searching for all the possible mmWave beam directions by exploiting

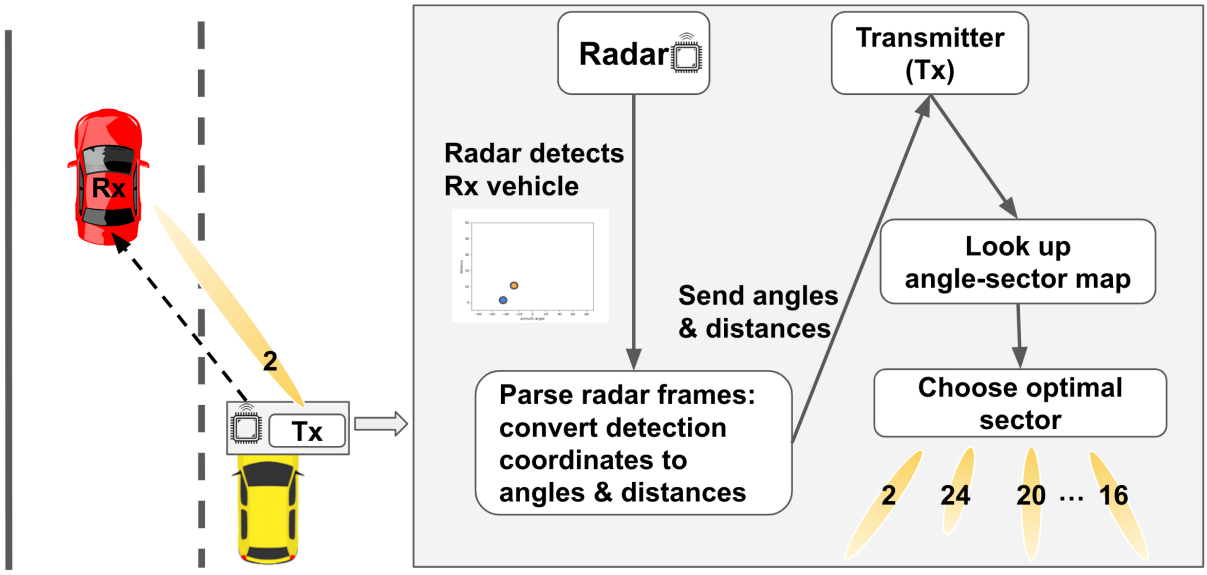
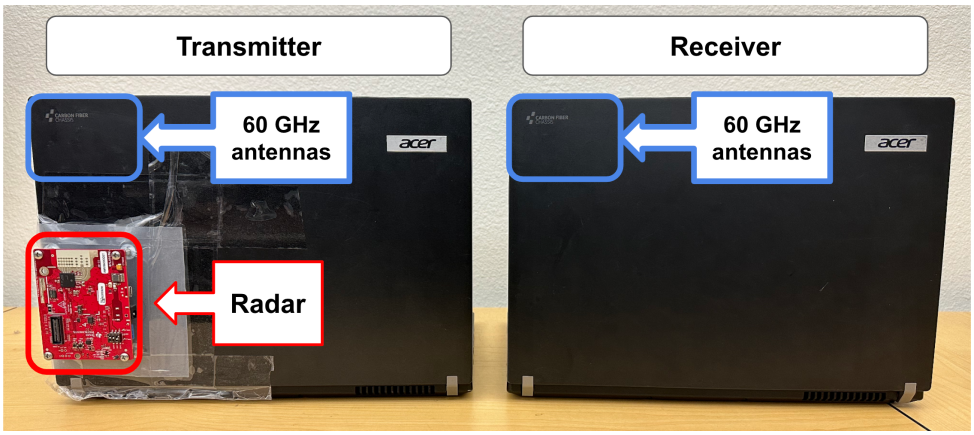


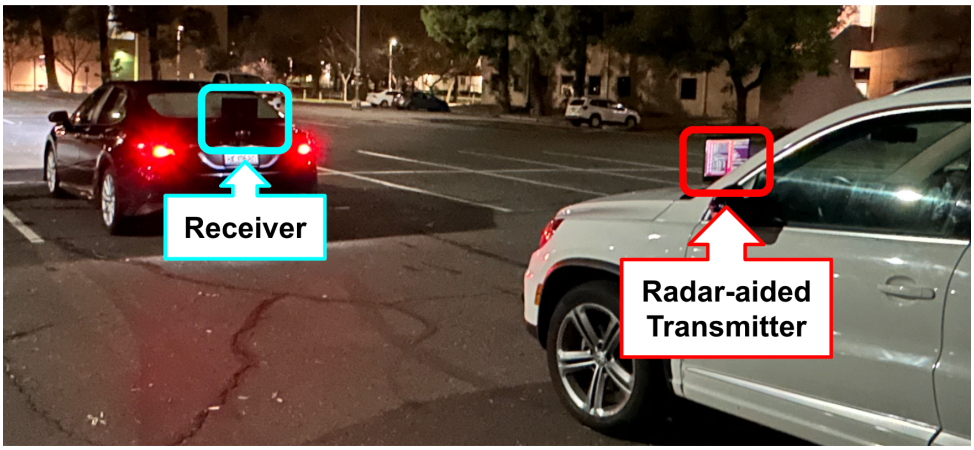
Figure 2.4: Overall radar and Tx workflow.

the sensory data of the radar to directly set the best beam sector in the Tx. Herein, the receiver in AP mode is always set with an omnidirectional beam pattern.

Fig. 2.5 shows two identical COTS laptops and an FMCW radar. The laptop model is Acer Travelmate TMP648-MG, which contains an 802.11ad wireless module. Qualcomm QCA9500 chipset is used for the module, which is equipped with a phased antenna array with 32 elements operating at 60 GHz. This chipset was widely used by researchers to understand and optimize beamforming protocol for IEEE 802.11ad standard [69, 92, 93]. Blue squares in Fig. 2.5a are the locations of the QCA 9500 chipset. The wil6210 driver for QCA 9500 wireless chipset was used for our laptops installed with the Ubuntu operating system. For the COTS FMCW radar, AWR1843 BOOST by Texas Instrument was used, whose radar chip works on the 77 GHz mmWave frequency band. It has three transmitter antennas and four receiver antennas. The FMCW radar parameters and values used for our experiment are provided in Table 2.1. The radar is attached to the transmitter laptop, and its antennas are aligned with the laptop's phased antenna array so that the radar and transmitter laptop can share the detected angles at the same coordinates. Fig. 2.5b shows that the receiver laptop in AP mode is attached to the back of the dark vehicle; the transmitter laptop equipped with radar is set in client mode and attached to the front



(a) COTS devices used for the experiment. Two Acer 802.11ad laptops equipped with phased antenna array on 60GHz and one TI FMCW radar on 77GHz attached to the laptop.



(b) Vehicles equipped with the Radar-assisted Tx and Rx for V2V.

Figure 2.5: Experimental setup for V2V communication.

hood of the light color vehicle.

In order to obtain the real-time position data of the target vehicle (attached with the Rx), such as angle and distance, we modify the real-time radar detection code [94] and then add the sector selection function in beamforming training to implement our proposed protocol. As shown in Fig. 2.4, the Rx vehicle reflects the chirps of the radar, so the radar frame contains the orientation data of the Rx vehicle; then the orientation data is converted into the angle and distance of the Rx vehicle relative to the Tx vehicle. Based on our pre-calculated angle-optimal sector map (see Table 2.2 for details), the transmitter can set the optimal sector for the detected angle of the current Rx vehicle. The threshold for Constant false alarm rate (CFAR) [42] was set to 15 dB in the radar configuration to

Table 2.1: FMCW Radar Configuration

Parameter	**Value**
Starting Frequency, f_c	77 GHz
Sweep Slope, S	30 MHz/μs
RF Sweep Bandwidth, B	1.8 GHz
Number of chirps per frame, N_{chirp}	36
Number of samples per chirp	128
Number of Tx, Rx	2, 4
Frame Duration	10 ms
Sampling Frequency	10,000 ksps

detect the Rx vehicle with high reflectivity while filtering out environmental noise. The DBSCAN algorithm was also adapted to our detection results. Although unlimited frames can be sent from the radar, we applied a limited number (i.e., 20) of frames needed for the experiment to reduce the signal processing delay. The modified wil6210 driver enables the utilization of defining the wanted sectors [69] based on the range-angle data measured by the FMCW radar on the transmitter laptop. With the modified driver, we can directly choose the transmit beam sector with the predefined beam pattern in real-time without operating a Tx's full sector sweep. The codebook for the driver uses 36 sectors ranging from 1-31 and 59-63 [69, 92], which cover all the angles. Thus, the specific sector number can be set based on the measured angle from the FMCW radar. For example, if the radar on the transmission side detects the receiver vehicle at an angle of 20 degrees, then the transmitter can set the sector 2, providing a high data throughput rate and low latency for that angle. Optimizing the beam sector number with high performance for each angle range for radar detection is described in Section 2.4.

For our experiments, the transmitter laptop was operated as the client, and the receiver laptop was set as the AP. We conducted the measurement at least five times, with each lasting 30 seconds. Throughput was measured by running the iperf3 to test the downlink TCP flow, and the latency was estimated by ping.

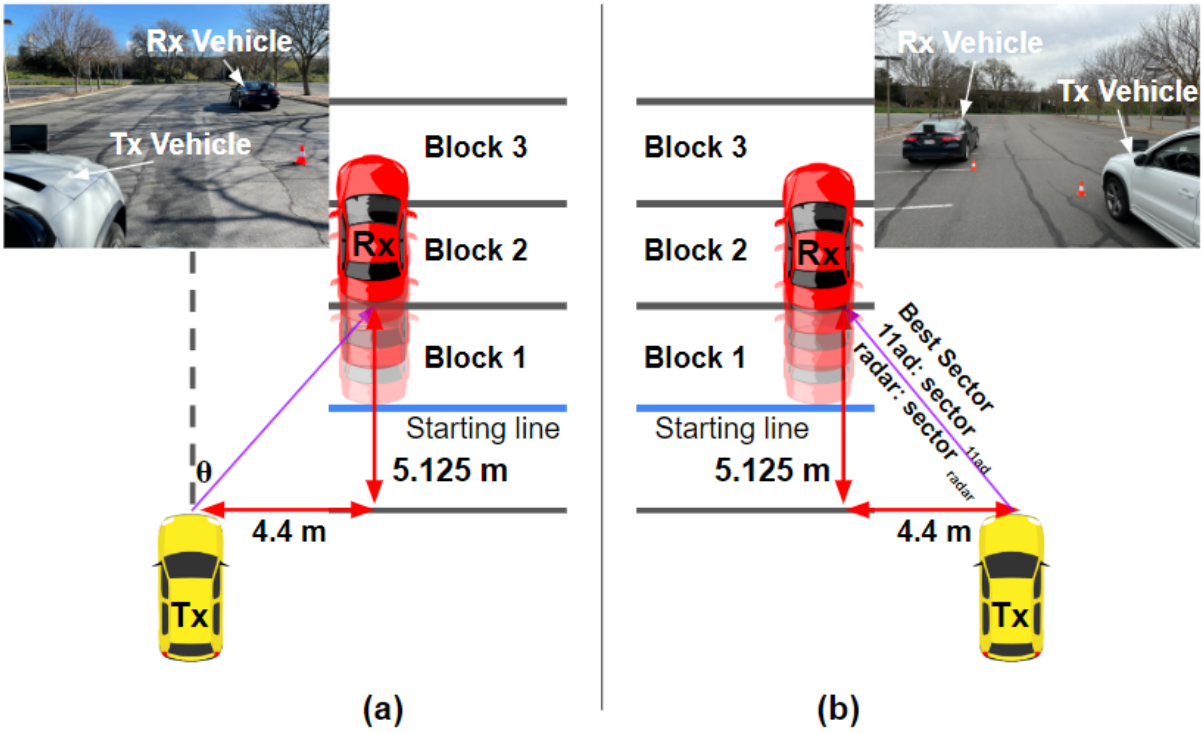


Figure 2.6: Experimental setup for optimizing the best beamforming sector for each block and scenarios in sections II and III when (a) the Rx car is upper-right to the Tx car, and (b) the Rx car is upper-left to the Tx car. θ is the azimuth angle between the transmitter antenna and the receiver antenna with respect to the surface of the transmitter laptop.

2.4 Static Transmitter and Dynamic Receiver

2.4.1 Calibration of Beam Sectors

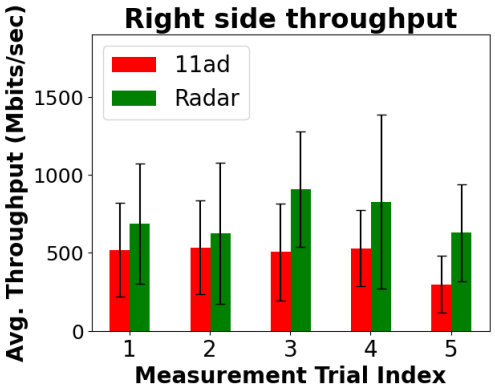
In order to determine the most effective beam pattern for a detected vehicle's location, it is necessary to identify the optimal beam sector based on the detected angle. Although the ground truth regarding the angle between the Tx and Rx laptop antennas remains fixed for each block, variations may occur in relation to angles detected by radar. This is due to radar's ability to detect both a vehicle's head and tail, resulting in multiple detected points. To address this issue, we implemented a clustering method that consolidates vehicle detections into singular data points containing both distance and angle information. The method employs threshold values for distances and angles; if a given set of detected point values falls below these thresholds, they are averaged together, resulting in one consolidated data point.

Fig. 2.6 illustrates the experimental setup to determine the optimal sectors for stationary transmitter and receiver vehicles. After each measurement, the receiver vehicle advanced one block. The experiment was conducted in a parking lot where equidistant parking lines were spaced 2.5 meters apart with a line width of 0.125 meters. Initially, the distance between the transmitter vehicle and the receiver vehicle is approximately 5 m when the back of the receiver vehicle is located at the starting line in the first block. Measurements were taken up to Block 12, located 31.375 meters from Block 1's starting line (refer to Fig. 2.6).

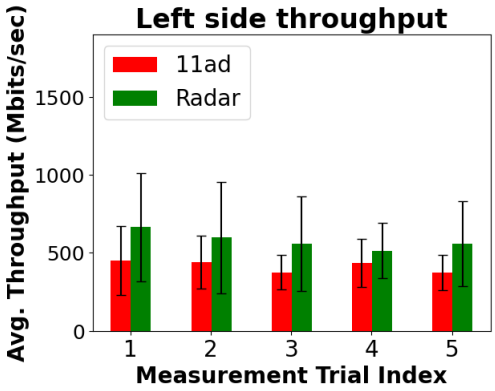
The experiment site was suitable for a line of sight (LOS) scenario, with minimal reflectors or noise, including trees and lighting poles. We use the throughput value to evaluate each sector's performance. At each block, we ran iperf3 for 60 seconds and verified the sectors selected with log data generated by the wil6210 driver. For the default 802.11ad protocol, we utilized the original wil6210 driver built-in Ubuntu, which identifies qualified sectors for communication and provides usable throughput rates. For the radar-assisted protocol, we manually set each sector in the codebook (i.e., 36 sectors) and selected the one with the highest average throughput based on the iperf3 measurements. Although the default protocol ought to ideally decide the sector with the highest quality for each block, some manually chosen sectors yielded higher data rates than the default protocol's choice (see Table 2.2). Also, the default protocol used multiple sectors for certain blocks, which may be attributed to the beam patterns with similar SNR values. We confirmed that most blocks used the optimal beam patterns with strong SNR values reported in [92]. Additionally, we tested radar detection of the Rx vehicle for each block without noise interference and carefully considered the detection angles when assigning the corresponding beam sectors for our radar-assisted beamforming protocol.

2.4.2 Measurement Results for Throughput and Latency

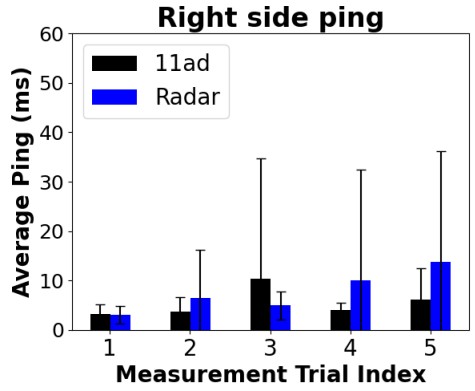
We have used the same experimental setup as shown in Fig. 2.6 for the scenario where the transmitter vehicle is stationary, but the receiver vehicle is moving forward until the twelfth block. Initially, the connection was established by the default 11ad protocol. When the receiver vehicle started to move forward at 3 mph, the radar frames were



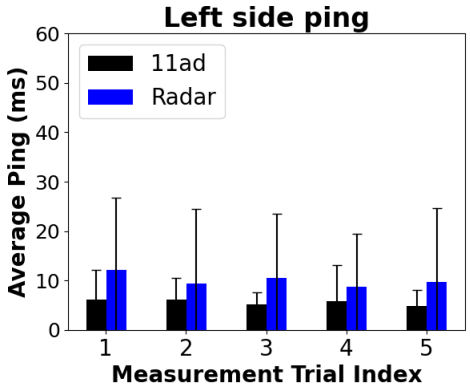
(a) Average throughput when the Rx vehicle is located at the upper-right to the Tx vehicle.



(b) Average throughput when the Rx vehicle is located at the upper-left to the Tx vehicle.



(c) Average latency when the Rx vehicle is located at the upper-right to the Tx vehicle.

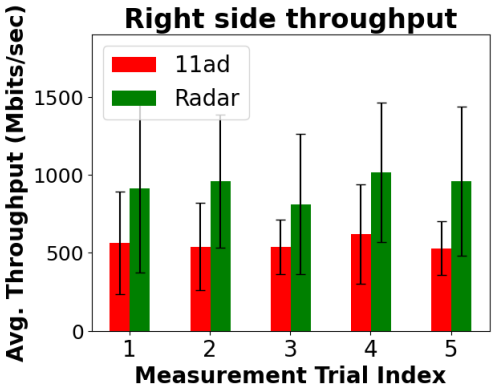


(d) Average latency when the Rx vehicle is located at the upper-left to the Tx vehicle.

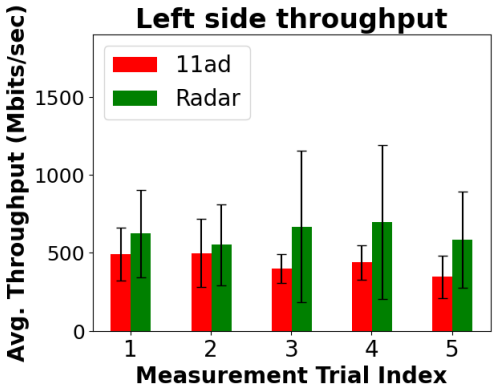
Figure 2.7: Throughput and latency measurements for five trials when the transmitter vehicle is stationary but the receiver vehicle moves forward with 3 mph.

sent periodically from the transmitter vehicle to detect the receiver vehicle's angle and distance. Each radar frame selects the optimized beam sector based on the detected angle and automatically changes the beam sector from the transmitter laptop in real-time. The beam sector is not changed if the sequential radar frame selects the same beam sector to prevent repeated sector change. For example, the radar detection will not trigger the beam sector change if the detection angle lies within the range of blocks 2 to 4 for 'Rx right' since the radar's sector choice is the same for those blocks (i.e., sector 16).

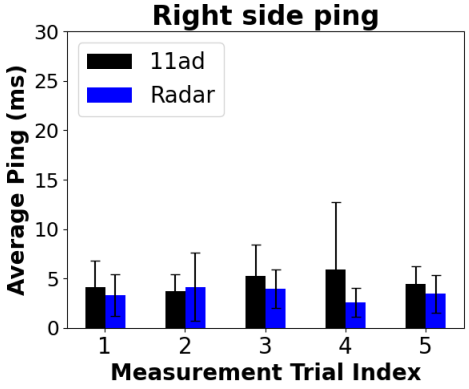
The throughput and latency were measured for the above scenario from the transmitter laptop. As a baseline, we also tested the default 11ad protocol. Fig. 2.7a and Fig. 2.7b



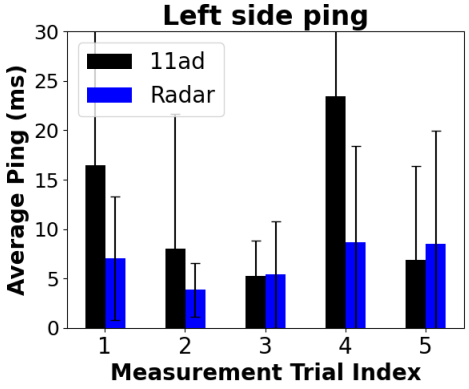
(a) Average throughput when the Rx vehicle moves from the upper-right to the Tx vehicle.



(b) Average throughput when the Rx vehicle moves from the upper-left to the Tx vehicle.



(c) Average latency when the Rx vehicle moves from the upper-right to the Tx vehicle.



(d) Average latency when the Rx vehicle moves from the upper-left to the Tx vehicle.

Figure 2.8: Throughput and latency measurements for five trials when both the transmitter and the receiver vehicles are in motion.

show the average throughput for the receiver vehicle positioning at upper-right/upper-left to the transmitter vehicle. Twelve continuous timestamps were chosen for data analysis. For the right side throughput, the average values of five trials for the radar-assisted and the default protocols were 735.98 MB/s and 477.88 MB/s, respectively. The overall average throughput was improved in the radar-assisted protocol by 54% compared to the default protocol (see Fig. 2.7a). On the other hand, the average values for the radar-assisted and the default protocols were 578.96 MB/s and 415.58 MB/s for the left side, respectively. There was a 39% increase in our protocol compared to the baseline. The right side improvement is higher than the left side value since the right side contains the robust

Table 2.2: Comparison of the optimal sector chosen in 802.11ad and radar-assisted protocol. Both Tx and Rx vehicles were stationary during the measurement. 'Rx right' indicates the Rx vehicle is located in the front-right position relative to the Tx vehicle while 'Rx left' means that the Rx vehicle is located in the front-left position. θ is the calculated angle between the Tx antenna and Rx antenna (see Fig. 2.6). 'TP' is the average throughput (Mbits/sec) for 60 seconds.

	Rx right		**Rx left**	
Block # / θ	**11ad Sector / TP**	**Radar Sector / TP**	**11ad Sector / TP**	**Radar Sector / TP**
1 / 60.3°	multiple sectors[a] / 526	12 / 468	26 / 483	2 / 581
2 / 40.6°	61, 16 / 947	16 / 1010	multiple sectors[b] / 389	2 / 399
3 / 29.6°	16 / 1140	same sector (16) / 1140	25 / 401	same sector (25) / 401
4 / 23.0°	16 / 473	same sector (16) / 473	multiple sectors[c] / 392	2 / 374
5 / 18.7°	24 / 439	same sector (24) / 439	multiple sectors[d] / 405	10 / 440
6 / 15.7°	24 / 407	61 / 448	20 / 400	same sector (20) / 400
7 / 13.6°	24 / 456	same sector (24) / 456	20 / 464	same sector (20) / 464
8 / 11.9°	24 / 519	same sector (24) / 519	20 / 407	2 / 473
9 / 10.6°	24, 20 / 372	20 / 392	20 / 381	2 / 433
10 / 9.7°	24, 20 / 482	61 / 503	20 / 386	63 / 435
11 / 8.7°	24, 20 / 423	24 / 392	20 / 437	2 / 422
12 / 8.0°	multiple sectors[e] / 330	24 / 546	20 / 450	61 / 551

[a]Sectors 30, 21, and 12 were used.
[b]Sectors 30, 26, 25, 12, and 2 were used.
[c]Sectors 22, 20, 16, and 10 were used
[d]Sectors 61, 22, 20 were used
[e]Sectors 63, 24, 20, and 16 were used.

beam sector (i.e., sector 16) with more gigabyte-level throughput values. Furthermore, our protocol only chooses beam patterns with the highest gain for each angle range so the protocol can outperform the baseline, which uses a variety of beam sectors with weak beam patterns. The large standard deviations for the throughput could have originated from the abrupt change of beam patterns during the measurement. Also, the difference between the detected angle by the radar and the angle between the Tx and Rx can give the sub-optimal beam sector to the transmitter, resulting in low throughput.

The latency measurements are illustrated in Fig. 2.7c and Fig. 2.7d. Thirteen continuous timestamps were sampled for latency analysis. The average latency values for the default and the radar-assisted protocols for the right side were 5.47 ms and 7.64 ms. The

average ping for the baseline and our protocol for the left side were 5.58 ms and 10.10 ms. The default protocol's latency was more reliable than our protocol's latency. Nonetheless, the latency differences were less than 5 ms between the baseline and our protocol. The overall latency for the left side was higher than the right side for our protocol. This is likely because the right side beam patterns are more robust than the left side beam patterns. To be specific, the left side beam patterns are too narrow to make beam patterns difficult to align with each other. The latency for our protocol can be improved if we can optimize the transmit beam patterns in the codebook without using the predefined beam patterns.

2.5 Dynamic Transmitter and Receiver

We have used the same location and setup shown in Fig. 2.6 for the dynamic scenario, except that the transmitter car was also driven forward. The transmitter car was moving at 3 mph starting from 0 mph, and the receiver car was speeding up to 5 mph from 0 mph. The maximum relative speed between the vehicles can be considered as 2 mph. The high mobility scenario was not tested due to the short range of the parking lot.

Fig. 2.8a and Fig. 2.8b show the measured average throughput values for the 'Rx right' and 'Rx left' cases when both vehicles were driving straight forward. Twelve sequential timestamps were used for data analysis. The average throughput values for the 'Rx right' were 932.45 MB/s and 558.13 MB/s for the radar-assisted and the default protocols, respectively. The radar-assisted protocol improved the throughput by 67% compared to the default protocol. Each throughput values were higher than those of the static scenario in section 2.4. This is because the distance between moving vehicles was less than the static scenario, so it can exploit the high throughput sectors within blocks 1 to 4 (see Table II). The radar detection was also more accurate for this case than the static scenario, resulting in less difference between the detected angle and the ground truth angle for Tx and Rx antennas. The average throughput values for the 'Rx left' were 624.58 MB/s and 435.38 MB/s for the radar-assisted and the default protocols, respectively. There was a 43% improvement in our protocol compared to the baseline

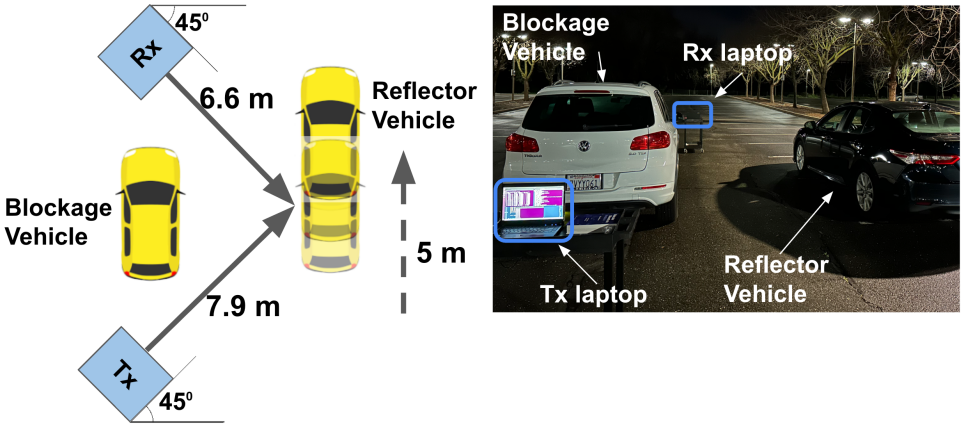
protocol, which was slightly improved than the static case. Similar to the static case, the overall throughput was higher with the 'Rx right' than the 'Rx left'. We conclude the right side beam patterns are more substantial than the left side beam patterns.

Fig. 2.8c and Fig. 2.8d illustrate the latency performance for the radar-assisted and the default protocols for the dynamic scenario. Thirteen sequential timestamps were used for latency analysis. The average latency values for the default and the radar-assisted protocols for the 'Rx right' were 6.68 ms and 3.48 ms. Our protocol's latency was 90% lower than the default protocol's latency. The default The average latency values for the default and the radar-assisted protocols for the 'Rx left' were 12.02 ms and 4.70 ms. In contrast to the static case, the latency performance in the radar-assisted protocol was better than that in the default protocol. This is because our protocol uses the high throughput beam sectors, which results in low latency when there is an optimal matching of beam pairs when the distance between the cars is close. Also, our protocol could derive benefits from less frequent beamforming compared to the default protocol. For comparison for the right and the left sides, the right side ping was better than the left side, similar to the static case.

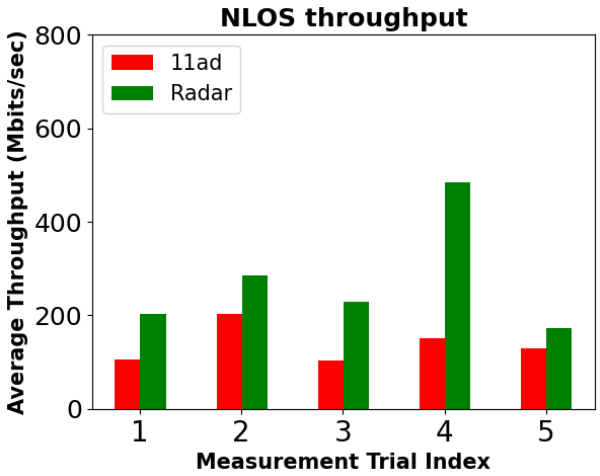
2.6 Non-Line-of-Sight Scenario

We tested a simple None-line-of-sight (NLOS) scenario, as shown in Fig. 2.9a, where vehicles are used as a blockage and a reflector for a real-world scenario. The NLOS condition was verified by removing the reflector vehicle to confirm the disconnection between Tx and Rx laptops. Tx and Rx laptops were located on identical moving carts. The Tx and Rx antennas were tilted towards the reflector for 45 degrees. During the throughput measurement, the reflector moved forward at 3 mph for 5 meters. We only moved this distance since the mmWave link was disconnected shortly after the movement of the reflector car.

Fig. 2.9b shows the throughput measurement result for our and the 11ad protocols. Three continuous timestamps were taken for Fig. 2.9b. The average throughput values were 138.13 MB/s and 274.47 MB/s for the 11ad and the radar-assisted protocols. This



(a) Experimental setup for the NLOS scenario



(b) Average throughput when the reflector vehicle is moving forward

Figure 2.9: Throughput measurements for the NLOS scenario when the blockage car is present, and the reflector car is moving forward.

is approximately 100% improvement with our protocol compared to the baseline. This is because our protocol only used sector 63 with the highest gain for zero degrees, but the default protocol used other sectors along with sector 63. Also, the average time duration values of connection until disconnection, where the data rate goes to 0 MB/s, were 4.2 seconds for the default protocol and 6 seconds for our protocol.

2.7 Highway Scenario

For the preliminary result for the highway, we have characterized our protocol on the county highway. Fig. 2.10a shows that the transmitter and receiver vehicles are in motion

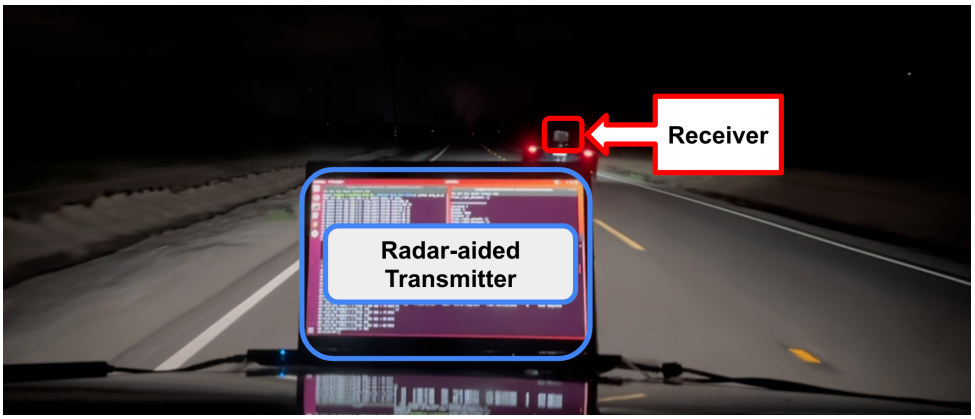
on the local road. The road was free of other cars or highly reflective objects during the measurement. We have used the same experimental setup as shown in Fig. 2.5b. For the highway scenario, the receiver vehicle was initially in front of the transmitter vehicle at the same speed at 20 mph. Then, the receiver vehicle was driven to the right side of the road while accelerating to 25 mph. The initial distance between the vehicles was 5 m, which was calculated from the GPS application of each car. Twelve sequential timestamps were used for the throughput analysis.

Fig. 2.10b shows the throughput measurement result for the default and our protocols. The average throughput values were 666.67 MB/s and 401.07 MB/s for our protocol and the default protocol, respectively. There was a 66% improvement in our protocol compared to the default protocol. Although the improvement is similar to the dynamic case for the 'Rx right,' the average throughput was less than that case. This is because the relative speed was higher in the highway case, resulting in the sub-optimal alignment of beam patterns.

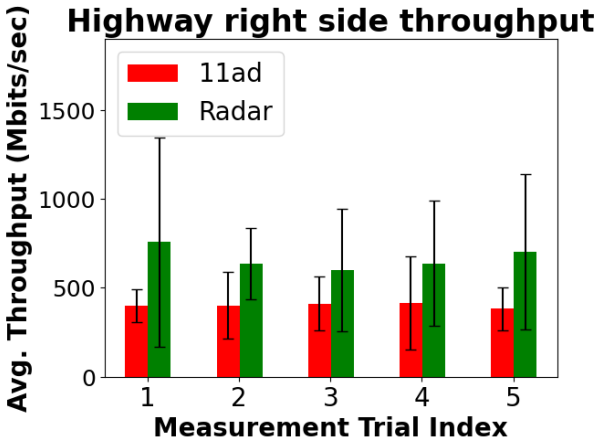
2.8 Summary

Despite the fact that our protocol enhanced the throughput in comparison to the baseline protocol, we encountered several challenges during the measurement. Our findings especially indicate a considerable degree of variability. This variability can be attributed to the beam misalignment for the specific angle ranges. The misalignment can occur since our radar cannot detect the exact location of the antenna in the Rx laptop on the vehicle. Also, due to the limited number of receiving antennas for the radar (i.e., four antennas), our angular resolution is limited to 15 degrees. Thus, the misalignment problem can be eased with more receiving antenna configuration by improving the angular resolution. By enhancing the angular resolution through the incorporation of additional receiving antenna configurations, it is possible to address the misalignment issue.

Additionally, based on our experimental results, measurements on the left side typically exhibit reduced throughput and increased latency. This is due to the asymmetric beam patterns between the negative and positive angle ranges. Also, the beam patterns



(a) Experimental setup for the highway scenario



(b) Average throughput for the right side highway scenario

Figure 2.10: Throughput measurements for the highway scenario when the receiver vehicle is moving to right side from in front of the transmitter vehicle.

in the negative angle ranges could be weaker than the positive ones. If the beam patterns for the negative angle ranges can be redesigned, we can acquire symmetric results with robust throughput and improved latency.

To summarize, we conducted experiments regarding the real-time radar-assisted beamforming protocol for four scenarios. The first scenario where the Tx vehicle was fixed, but the Rx vehicle was moving showed that the throughput with our protocol could be improved by 1.5× when the Rx vehicle was positioned to the upper-right of the Tx vehicle. Nonetheless, there was a slight delay in latency for the static scenario due to the beam misalignment. The delayed latency was less than 5 ms. Then, the dynamic scenario where both Tx and Rx vehicles were moving forward was studied. This scenario outperformed

the performance compared to the static scenario. This is because the range between vehicles was shorter than in the first scenario, allowing the radar to choose more robust sectors with a high gain. The throughput was improved by approximately 70% with our protocol, and the latency was improved by 90% to the default 11ad protocol when the Rx vehicle was located on the upper-right side. Furthermore, the NLOS scenario with a simple experimental setup was shown. Selectively choosing the most robust sector by the radar's decision enabled our protocol to gain higher throughput than the baseline. Lastly, our protocol was tested on the county highway. Our protocol showed 66% improvement compared to the baseline for the 'Rx right' scenario on the highway. Our work can be a preliminary result for investigating the high mobility scenario, such as driving on a highway with a radar-assisted beamforming protocol.

Chapter 3

V2VRP: Cooperative Perception Dataset for Vehicle-to-Vehicle Communication using FMCW Radars

Automotive technologies have grown at a rapid pace over the past decade, incorporating features like adaptive cruise control, lane-keeping assistance, and autonomous driving capabilities [95–97]. Especially, autonomous driving has emerged as one of the most transformative technologies, promising to revolutionize transportation by enhancing safety, efficiency, and accessibility. Companies like Waymo and Zoox have been at the forefront of this movement, pioneering autonomous vehicle technology and launching fully autonomous taxi services in select cities [98, 99]. Their efforts have not only advanced the technology but have also helped build public trust in self-driving vehicles as a viable mode of transport. Tesla has also unveiled the Cybercap robotaxi for commercial autonomous driving, signaling its commitment to revolutionizing the transportation industry [100]. These developments are reshaping urban mobility, promising greater accessibility and efficiency. These autonomous vehicles (AVs) are designed to perceive and navigate complex environments with minimal human intervention [101–107] .

Central to these advancements is the vehicle's ability to perceive its environment using sensors such as cameras, LiDAR, and radar, which generate 3D representations

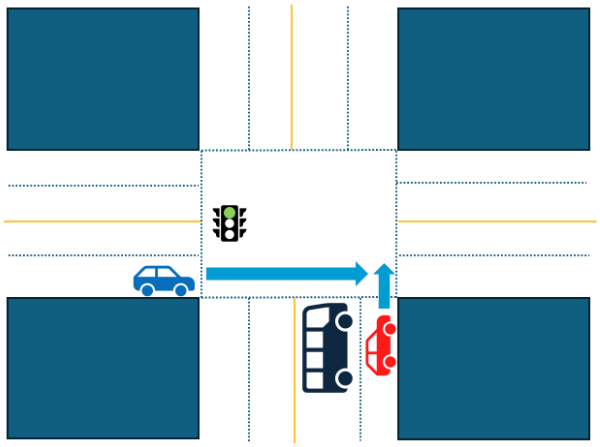


Figure 3.1: A scenario involving a limited field of view can lead to traffic accidents. The black bus is blocking the red car's sight so that it can crash into the oncoming blue car.

of the surroundings. However, these sensors have limitations, including occlusions and a restricted field of view of a single vehicle, which can result in an incomplete understanding of the environment.

These limitations pose significant challenges for autonomous driving, as vehicles must comprehensively understand their environment to make safe and accurate decisions. Occlusions, for instance, can obscure obstacles or other vehicles, creating blind spots that compromise the vehicle's ability to anticipate and react to potential hazards. Similarly, the restricted field of view means that conventional sensors may not detect objects outside a certain range, particularly in complex urban or crowded environments. One scenario involving occlusion that leads to traffic accidents, as reported by the National Highway Traffic Safety Administration (NHTSA), is depicted in Fig. 3.1 [108]. For example, as shown in Fig. 3.1, the view of the red vehicle is obstructed by the black bus. It is possible that the red vehicle could collide with the blue vehicle when making a right turn at the intersection. However, if the black bus shares its perception data with the red vehicle, the latter could avoid a collision with the blue vehicle.

To overcome these challenges, researchers are exploring cooperative perception methods, where vehicles communicate with infrastructure (V2I) or other vehicles (V2V). For example, OPV2V [109], V2X-Sim [110], and V2XSet [111] are simulated datasets designed to facilitate research in cooperative perception for autonomous vehicles. These datasets

use open-source simulators like SUMO [112] and CARLA [113] to create simulated road scenes and traffic scenarios. However, these datasets have significant limitations, such as a lack of real-world variability, as simulated environments often struggle to capture the full diversity and unpredictability of real-world settings. Additionally, there is the simplified modeling of sensor noise and synchronization; in reality, sensors are affected by varying environmental noise and GPS errors that are challenging to replicate accurately in simulations.

There is currently a lack of real-world datasets for cooperative perception with V2X technology. At present, the primary datasets available for this purpose are DAIR-V2X [114] and V2V4Real [115], which are measured in real-world V2X scenarios. DAIR-V2X [114] is notable for being the first real-world cooperative V2I dataset, but its implementation is costly and depends on stable, reliable connections between vehicles and infrastructure, which can be challenging to maintain consistently. V2V4Real [115], on the other hand, is the first large-scale real-world V2V dataset. While it offers valuable data for V2V scenarios, its sensors, such as LiDAR and GPS, are expensive, which can make it less accessible for widespread use.

As shown in Table 3.1, all the recent cooperative datasets use LiDAR point cloud data as a primary input for training their machine learning models. LiDAR point clouds provide high-resolution, dense data, capturing detailed information about objects and surroundings. However, this density comes at a cost: the resulting files are often very large, sometimes reaching gigabytes in size. Such large file sizes can lead to high latency when transmitting data, especially over networks with limited bandwidth, which can slow down real-time applications or require significant storage resources.

In contrast, frequency-modulated continuous wave (FMCW) radar point clouds produce significantly smaller file sizes, often in the kilobyte range. While radar point clouds are less detailed and have a lower resolution than LiDAR, they are more efficient for real-time communication due to their compact size. This smaller data footprint allows FMCW radar to operate with minimal latency and is more suitable for applications where rapid data transfer is essential, such as in automotive safety systems and early obstacle

detection.

In this work, we present a radar-based cooperative perception dataset (V2VRP) specifically designed for V2V scenarios. Our main contributions are as follows:

- We present the first large-scale radar-based cooperative perception (V2VRP) dataset tailored for V2V applications to the best of our knowledge.
- We provide a comprehensive radar point cloud dataset featuring over 39K annotated 3D bounding boxes for cars in the scene, along with synchronized radar point cloud frames, camera images, and GPS coordinates along the driving routes.
- Our dataset is specifically designed for V2V communication, supporting and advancing research in cooperative perception, which is a vital component of autonomous driving that enables connected vehicles to share sensory data, enhancing environmental awareness and safety.

Table 3.1: Comparison of the existing cooperative perception datasets.

Dataset	**Year**	**Category**	**Scenario**	**Sensors**	**Point cloud frame #**
OPV2V [109]	2022	V2V	Simulation	RGB, LiDAR, GPS	11k
V2X-Sim [110]	2022	V2X[a]	Simulation	Cameras[b], LiDAR	10k
V2XSet [111]	2022	V2X	Simulation	LiDAR	11k
DAIR-V2X [114]	2021	V2I	Real	RGB, LiDAR, GPS	39k
V2V4Real [115]	2022	V2V	Real	RGB, LiDAR, GPS	20k
V2VRP (Our work)	2024	V2V	Real	RGB, Radar, GPS	39k

[a]V2X includes both V2V and V2I
[b]RGB, Depth, Semantic Camera

3.1 Preliminaries

3.1.1 GPS Real Time Kinematics

For cooperative perception between vehicles, GPS location data is used to merge views from each vehicle. However, conventional GPS receivers provide meter-level accuracy, and the cost increases significantly for sub-meter accuracy. GPS RTK technology significantly

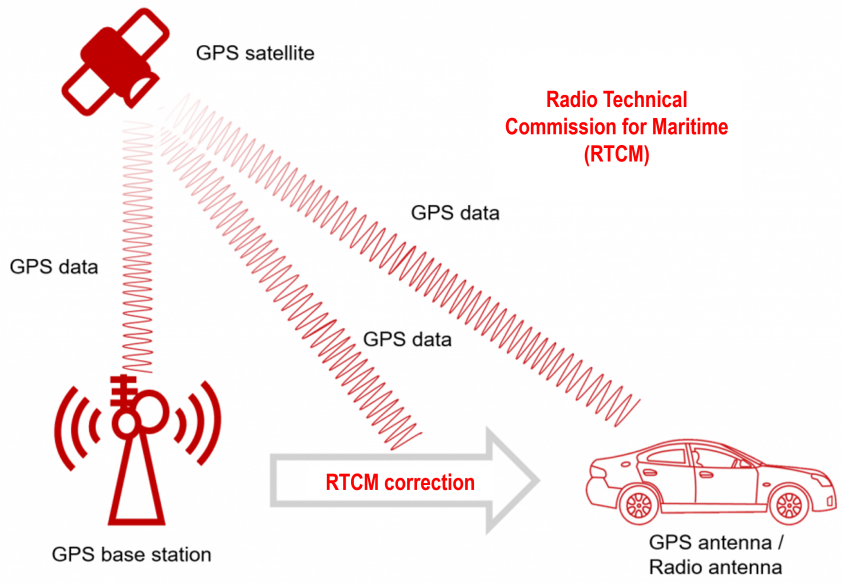


Figure 3.2: GPS RTK System. It can achieve centimeter-level positioning accuracy by combining satellite signals with real-time correction data from a nearby base station. Adapted from reference [3]

advances over traditional GPS, offering centimeter-level accuracy through real-time correction data [116]. While conventional GPS typically provides location accuracy within a few meters, GPS RTK combines signals from multiple global satellite systems—such as GPS, GLONASS, Beidou, and Galileo—with a correction stream, usually in the form of RTCM (Radio Technical Commission for Maritime Services) data [117].

Unlike traditional GPS, which relies solely on satellite data for positioning, GPS RTK systems enhance this data with correction information from a nearby base station, as illustrated in Fig. 3.2. This correction stream enables RTK receivers to adjust for common GPS errors, achieving positioning accuracy within 1 cm. Transitioning from traditional GPS to GPS RTK offers a new level of reliability and precision, which is especially valuable in fields like autonomous driving. Recent advancements have brought RTK receiver costs down to around $300, making high-precision GPS accessible to a wider range of applications that require real-time accuracy [118].

RTCM correction data can be supplied through reliable base stations if the GPS RTK device is within 10 km of the broadcasting station. UNAVCO, a U.S.-based nonprofit organization, provides high-precision geodetic data for research and operates a network of

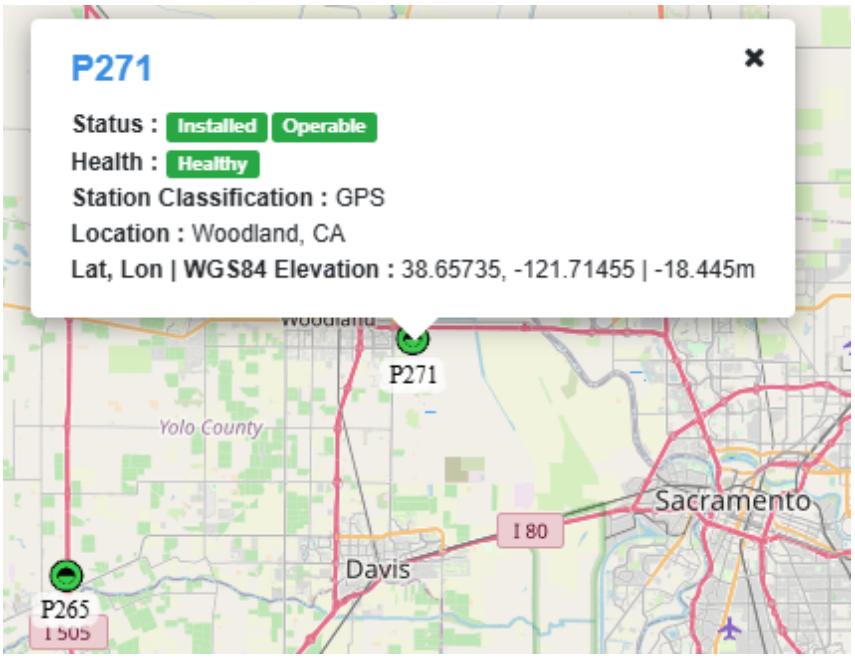


Figure 3.3: One of the base stations is located within 10 km of Davis to receive correction signals [4].

continuously functioning GPS and GNSS stations that broadcast real-time RTCM correction data that is accessible online [119]. For example, as shown in Fig. 3.3, we accessed the base station operated by UNAVCO in Woodland, located within 10 km of our measurement location in Davis. By connecting to UNAVCO's correction streams, researchers and professionals can seamlessly integrate these enhancements into their location solutions, improving RTK positioning accuracy for various applications.

3.1.2 3D Radar Point Cloud

3D point cloud data can be generated from radar IQ (In-phase and Quadrature) signals to perceive environmental information with FMCW radar. This process begins by capturing raw IQ radar signals, which contain complex data on the detected objects in the radar's field of view. To process this data, a series of Fast Fourier Transforms (3D FFT) is applied along three dimensions (i.e., range, velocity, and angle) to extract critical properties. After mixing the transmitted and received signals, the first step is to perform Range FFT for the IF (Intermediate Frequency) signal. From Eq. 1.5, range (d) can be calculated as

$$d = \frac{cf_{IF}T_{chirp}}{2B} \tag{3.1}$$

We can perform Doppler FFT with range-FFT data to calculate the velocity from Eq. 1.9. Finally, an angle FFT is performed on the 2D Doppler FFT data to calculate the angle of arrival. From Eq. 1.11, the azimuth angle (θ) can be written as

$$\theta = \sin^{-1}(\frac{\lambda\Delta\phi}{2\pi l}) = \sin^{-1}(\frac{\Delta\phi}{\pi}) \tag{3.2}$$

where l is $\frac{\lambda}{2}$ for the largest azimuth angular field of view. The elevation angle (φ) is calculated similarly and written as

$$\varphi = \sin^{-1}(\frac{\lambda\Delta\phi_z}{2\pi l}) = \sin^{-1}(\frac{\Delta\phi_z}{\pi}) \tag{3.3}$$

where $\Delta\phi_z$ is the phase difference between elevation antennas. By using these values, we can calculate (x,y,z) coordinates of the detected target with the following equations:

$$\begin{aligned} x &= d\cos\varphi\sin\theta \\ y &= \sqrt{R^2 - x^2 - z^2} \\ z &= d\sin(\varphi) \end{aligned} \tag{3.4}$$

These coordinates are crucial for constructing a 3D representation of the target's position in space.

3.2 Related Work

3.2.1 Autonomous Driving Datasets

There are widely used autonomous driving datasets for V2X research areas, such as KITTI [120], nuScenes [121], and Waymo Open [122]. KITTI [120] is one of the foundational autonomous driving datasets, providing labeled data for 3D object detection, tracking, and depth estimation. Collected in urban areas with cameras, lidar, and GPS, it has advanced early computer vision in autonomous driving but is limited in size and environmental diversity. nuScenes [121] offers 1,000 driving scenes with 360-degree sensor coverage in

diverse conditions, including varied weather and lighting. With extensive 3D annotations, it supports multi-sensor fusion, making it ideal for robust object detection and scene understanding. Waymo Open [122] is a large dataset that includes diverse urban scenes collected across the U.S., with high-quality 3D and 2D annotations from lidar and high-resolution cameras. Its scale and richness in scenarios make it valuable for training models in perception, planning, and motion prediction for autonomous driving. However, these datasets focus solely on developing single-vehicle driving capabilities, which have shown limited effectiveness in handling severe occlusions and long-range perception [123]. Other datasets such as SYNTHIA [124], Cityscapes [125], and BBD100k [126] focus on 2D perception for RGB camera images, but they restrict their effectiveness in autonomous driving applications that require accurate depth and spatial awareness.

Compared to these datasets, our work offers a cooperative perception to understand the environment in real-world scenarios more robustly. By enabling vehicles to share perception data, our approach overcomes limitations in single-vehicle perception, such as handling occlusions and expanding long-range visibility. It allows for enhanced spatial awareness and situational understanding, as vehicles can collectively interpret the environment with a more comprehensive view. Our work aims to improve the reliability and accuracy of autonomous systems by leveraging multi-vehicle collaboration.

3.2.2 Multi-Radar Fusion

Single-radar perception has notable limitations compared to multi-radar fusion. It has a restricted field of view and creates a sparse point cloud due to specular reflection, resulting in less detailed object detection. It is also more vulnerable to noise, leading to inaccurate perception. Multi-radar fusion [5, 127–129] addresses these challenges by expanding coverage, increasing point cloud density, and providing more robust detection through overlapping data from multiple radars. For example, Pointillism [5] combines data from multiple spatially separated radars with optimal placement to maximize spatial diversity and scene information. However, since the radars are mounted on the same vehicle, their field of view is limited. Our work leverages multiple radars mounted on separate vehicles, allowing for a wider range of view and enhanced environmental perception. Also, multi-

Ego Vehicle **Assistant Vehicle**

Figure 3.4: Sensors are attached to both the ego vehicle and the assistant vehicle for cooperative perception.

radar fusion has been used for advanced surveillance systems, but they are not suitable for V2V scenarios [130–132].

3.3 V2VRP Dataset

3.3.1 Experimental Setup

As shown in Fig. 3.4, the V2VRP dataset was collected using two cooperative vehicles equipped with identical sensors: a RGB camera, GPS-RTK, and FMCW radar. Each vehicle had these sensors connected to a dedicated laptop (see Fig. 3.5b. One vehicle serves as the ego vehicle, while the other acts as an assistant vehicle to share perception data. For GPS system in Fig. 3.5a, we used GPS-RTK (Sparkfun NEO-M8P-2) for centimeter-level accuracy. A USB-to-Serial adapter (SparkFun Serial Basic Breakout CH340G) transferred the GPS correction signal from the GPS-RTK board to the laptop. The laptop was connected to a mobile hotspot to receive correction signals in outdoor, real-world scenarios. A GNSS multiband antenna was used to receive both the traditional L1 GPS band and the newer L2 GPS band. The total cost of this GPS system is less than $400, providing a cost-effective alternative to traditional GPS sensors with comparable accuracy. The P271 base station in Woodland was used to receive the GPS correction signal (see Fig. 3.3). A smartphone (Google Pixel 6a) is used as an RGB camera to capture timestamped image frames of the environment, providing ground truth labels

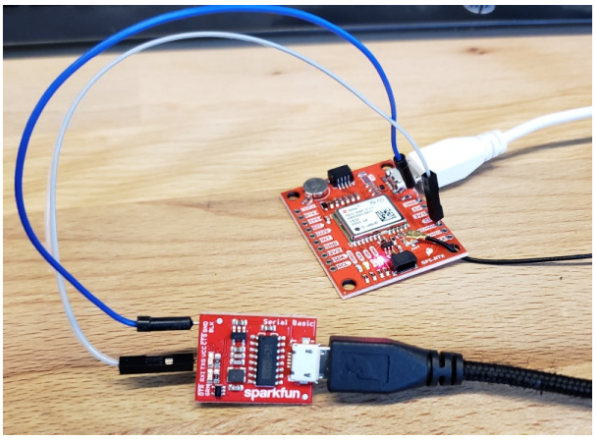
for 3D bounding box estimation. For a COTS FMCW radar, AWR1843 BOOST by Texas Instrument was used. The radar parameters used are shown in Table 3.3. The DCA1000 EVM capture board was also used to stream the ADC data from AWR1843 radar to a laptop over Ethernet. The camera lens was aligned with the radar antennas for calibration, and the GPS was positioned close to the radar. Each set of sensors was mounted at the same height on the vehicles. The sensor specifications for each vehicle are outlined in Table 3.2. We collected 100 scenarios, each containing 600 frames. Since each frame is 100 ms in duration, each scenario spans 60 seconds. Eighty scenarios were static cases, measured in a parking lot, while twenty scenarios were dynamic cases, measured on the UC Davis campus.

Due to the different capture frequencies of the sensors, we sampled key frames from the radar at 10 Hz to synchronize with the camera images and GPS coordinates. Since the GPS capture rate was the lowest, we interpolated the GPS data from 1 Hz to 1000 Hz to align with the radar timestamps. The GPS and camera timestamps closest to each radar timestamp were selected for synchronization for each vehicle. To synchronize radar frames between the ego vehicle and the assistant vehicle, the starting timestamps of the radar sensors were compared. For static scenarios, precise synchronization is not required due to the static point cloud data from the radar. However, there is an unavoidable asynchronous delay for dynamic scenarios due to the different system times between different laptops. We referred to the benchmark dataset V2V4Real [115] and selected measurement samples with less than a 50 ms delay.

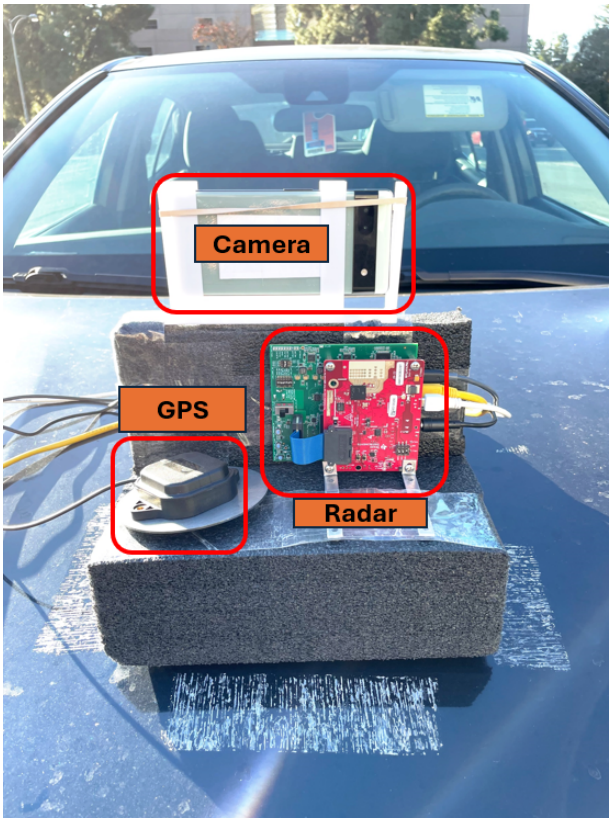
Using our experimental setup, we collected data from 19 scenarios in a parking lot at UC Davis. Each scenario contains 680 point cloud frames for the ego car (a total of 12,920 frames), 680 point cloud frames for the assistant car (12,920 frames), and 680 point cloud frames for the combined view (12,920 frames).

3.3.2 Radar Perception

For our 3D point cloud generation, we modified the open-source 3D point cloud generation code to process our measured raw ADC data [133]. A static clutter removal algorithm used to eliminate background noise in [133] was disabled to get rich point cloud data for static



(a) GPS-RTK system for our work. GPS-RTK board is connected with a serial basic breakout board to receive GPS correction signals from satellites.



(b) An RGB camera, FMCW radar, and GPS sensor are positioned together on the front side of the car to perceive the environment.

Figure 3.5: Experimental setup for cooperative perception data collection.

Table 3.2: Sensor Specification And Capture Rate For Each Vehicle

Sensor	**Specification**	**Capture frequency**
RGB Camera	Google Pixel 6a, 1920 x 1080	30 Hz
Radar	AWR1843	10 Hz
GPS	Sparkfun Neo-M8P-2	1 Hz

Table 3.3: FMCW Radar Configuration

Parameter	**Value**
Starting Frequency, f_c	77 GHz
Sweep Slope, S	30 MHz/μs
Number of frames	60
Frame Duration	100 ms
Number of samples per chirp	256
Number of Tx, Rx	3, 4
Sampling Frequency	10,000 ksps
Azimuth angle resolution	15°
Elevation angle resolution	60°

scenarios. We select the pixels with the highest values from the Doppler-FFT heatmap as candidate points for potential 3D detections rather than using the fixed threshold for CA-CFAR since this fixed threshold may cause considerable fluctuations in selected points across frames. For instance, one frame may generate hundreds of selected points, while another may produce very few or none, which can result in useless data for ML training. This issue is especially pronounced in dynamic test environments, where the energy distribution shifts significantly in occluded scenarios. Therefore, the 128 highest-value pixels from the Doppler-FFT heatmap are selected for each frame to maintain a consistent number of point clouds across different frames and environments. By using this filtering method, points affected by multi-path effects are efficiently filtered out, as signals that undergo multiple reflections typically have lower energy than those that are directly reflected. We also filtered out the points which are either too close or too far from the

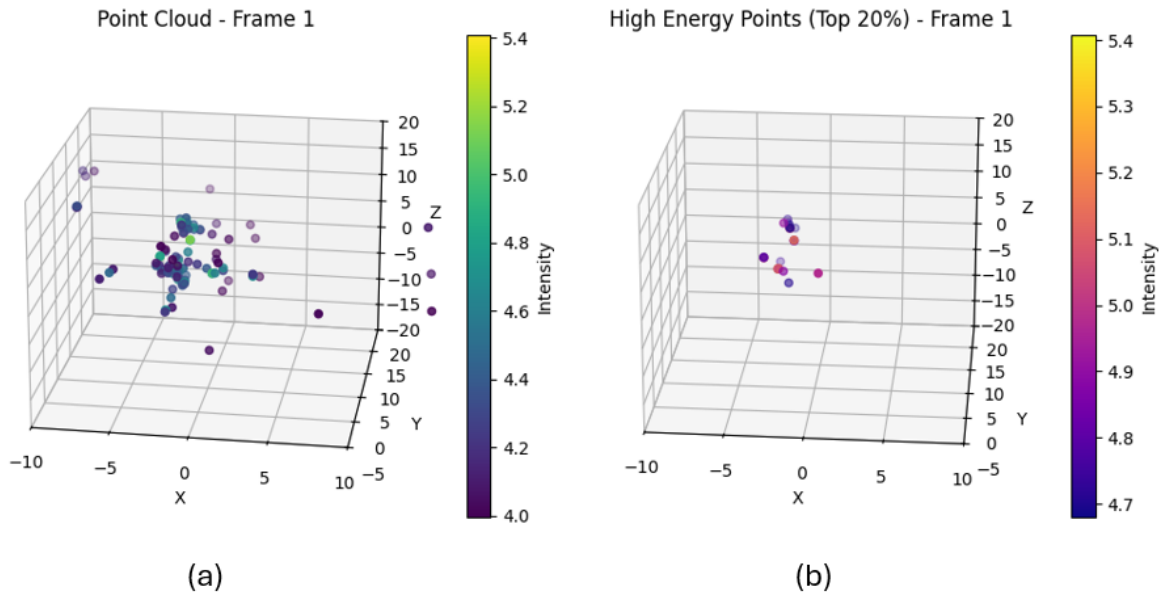
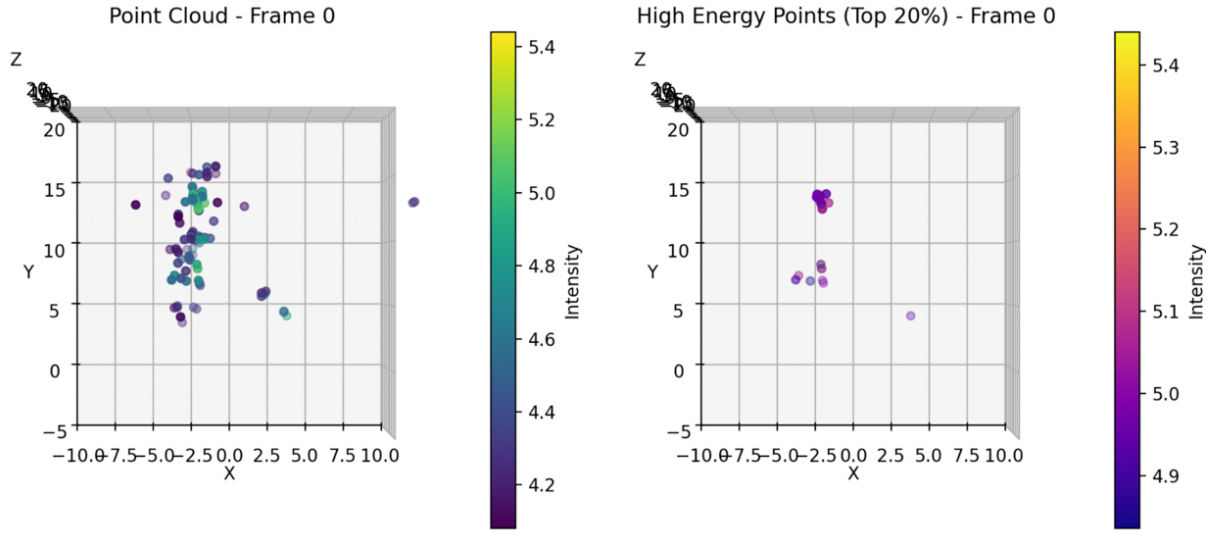


Figure 3.6: 3D point cloud data (a) The highest 128 points for static measurement. Each data point contains the range, velocity, angle, signal energy, x, y, and z coordinates. (b) Selecting only the point cloud data with the top 20 percent signal intensity to further filter out noise from the 128 points.

radar. After this noise removal, we calculated the x, y, and z coordinates, range, velocity, angle of arrival, and signal intensity for each point sample. A sample of our measurement for 3D point cloud data is shown in Fig. 3.6(a). We can further filter more points by only choosing the top 20% signal intensity points, as shown in Fig. 3.6(b).

To calibrate the 3D point cloud data with the actual environment, we compared it with the RGB image, as shown in Figs. 3.7 and 3.8. The measured distance between the radar and the target car aligns well with the 3D point cloud data. This consistency validates the accuracy of the radar data in representing real-world positions. In Fig. 3.7b, the plot shows the filtering of high-energy points isolates the most prominent reflections, which are likely associated with more significant objects or surfaces directly facing the radar sensor. This selective view helps in identifying key objects by reducing noise and highlighting the most relevant points for perception. However, not all reflected points were captured by the receiver antennas, likely due to specular reflection, which causes some signals to reflect away from the radar's receiving sensors. Additionally, some of the

(a) RGB image when the target car is located left to the radar

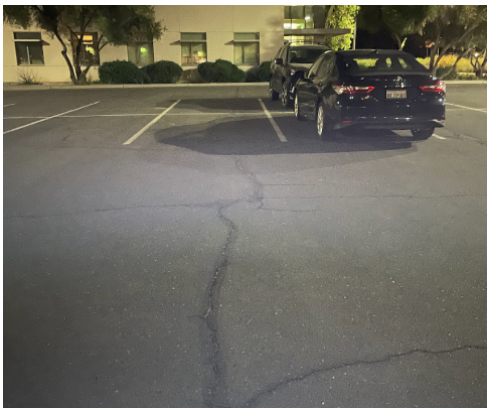


(b) Bird-eye view of the highest 128 points data (left) and the high 20 percent energy points data (right)

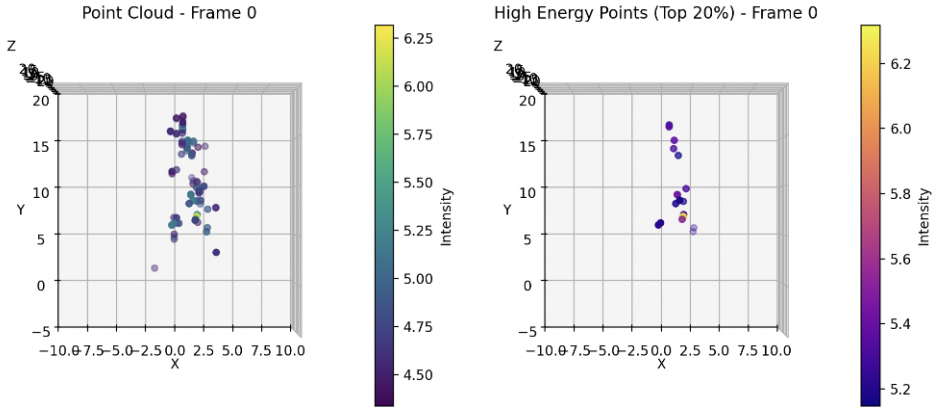
Figure 3.7: Calibration image and point cloud data for the static scenario where the target car is located in the upper-left relative to the radar.

detected points do not correspond to the target car due to noise.

Figs. 3.9 and 3.10 show how cooperative perception enables vehicles to share and validate sensory data, resulting in a more comprehensive and accurate environmental model. By leveraging multiple perspectives, cooperative perception enhances object detection and spatial coverage. For instance, Fig. 3.9 shows the ego car cannot detect the silver car blocked by the dark blue car. However, by receiving information from the assistant car,



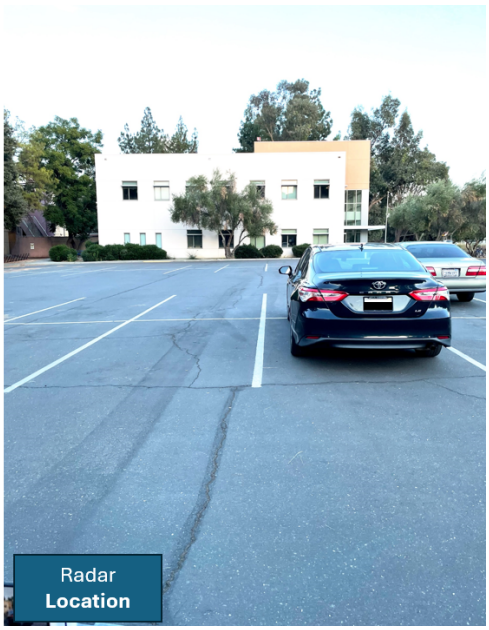
(a) RGB image when the target car is located right to the radar



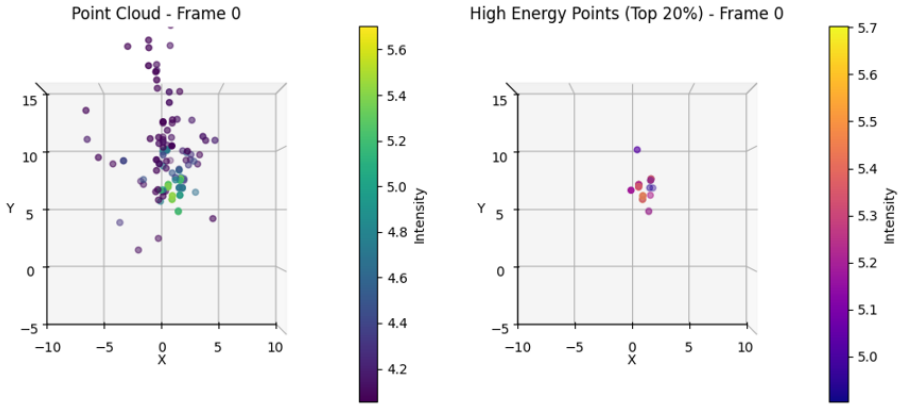
(b) Bird-eye view of the highest 128 points data (left) and the high 20 percent energy points data (right)

Figure 3.8: Calibration image and point cloud data for the static scenario in which the target car is located in the upper-right relative to the radar.

the ego car gains awareness of the blocked vehicle. Fig. 3.11 illustrates this process. In Fig. 3.11a, the point clouds from the ego and assistant cars are displayed in their original, unaligned states. The red points (assistant car) and blue points (ego car) are offset from each other, indicating a misalignment. This initial separation occurs due to differences in perspective between the two vehicles. Figure 3.11b shows the view after alignment. The assistant car's point cloud is brought into the ego car's reference frame by leveraging GPS coordinate differences. This alignment improves spatial consistency between the two point clouds, enhancing the accuracy of the combined environmental model in cooperative perception.



(a) RGB image from the perspective of the ego car



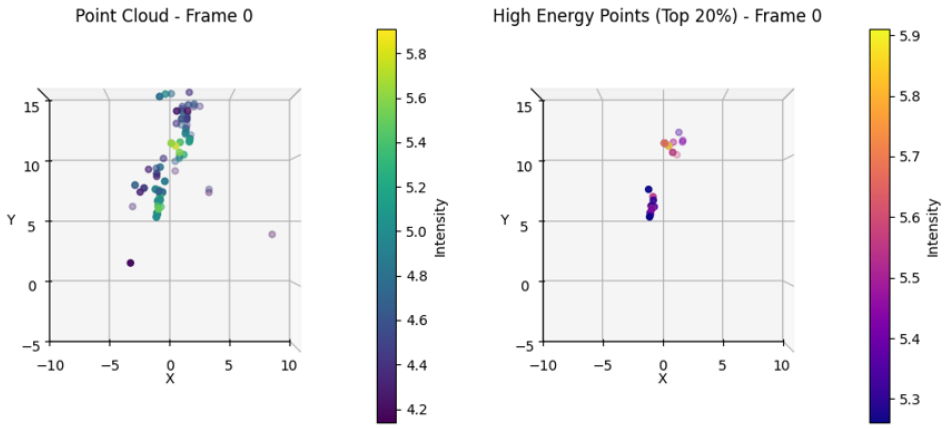
(b) Bird-eye view of the highest 128 points data (left) and the high 20 percent energy points data (right)

Figure 3.9: RGB image and point cloud from the perspective of the ego car. The ego car does not have a line of sight to the silver car located to the right of the dark blue car.

3.4 Summary

In this work, we introduce the first large-scale radar-based cooperative perception (V2VRP) dataset specifically designed for vehicle-to-vehicle (V2V) applications aimed at advancing research in autonomous driving. Our dataset includes over 39,000 annotated 3D ground-truth bounding boxes, synchronized radar point cloud frames, camera images, and GPS coordinates collected along driving routes, providing a rich resource for studying and

(a) RGB image from the perspective of the assistant car

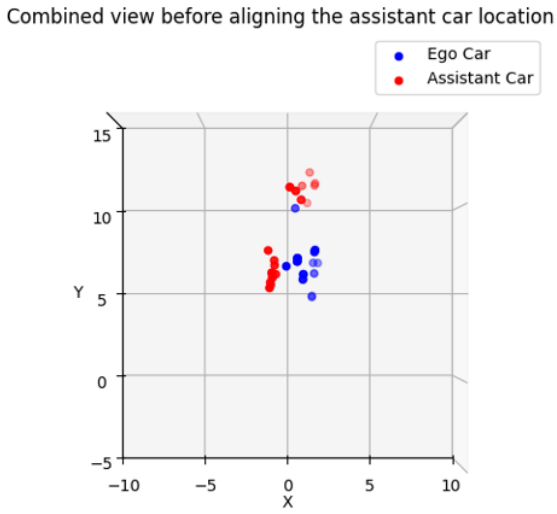


(b) Bird-eye view of the highest 128 points data (left) and the high 20 percent energy points data (right)

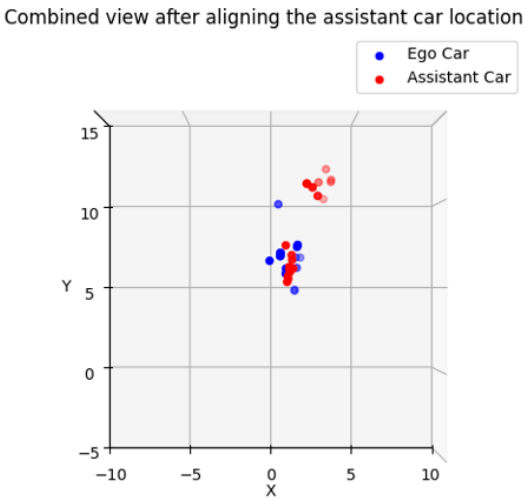
Figure 3.10: RGB image and point cloud from the perspective of the assistant car. The assistant car has a line of sight for both the silver and dark blue cars.

developing V2V cooperative perception systems.

This dataset is a significant contribution to the field, as it addresses the need for high-quality real-world radar point cloud data tailored for V2V communication. By enabling connected autonomous vehicles to share sensory data, it supports improved environmental awareness, object detection, and spatial consistency, all of which are critical for the safety



(a) Combined view of ego car and assistant car point clouds before aligning the assistant car's location.



(b) Combined view after aligning the assistant car's location to the ego car's frame.

Figure 3.11: Before and after merging the perspectives of the ego and assistant cars. Points from the ego car are displayed in blue, while points from the assistant car are displayed in red.

and efficiency of autonomous driving. Researchers can leverage this dataset to enhance cooperative perception methods, fostering advancements in the deployment of autonomous systems in real-world environments.

Through this dataset, we hope to accelerate progress in cooperative perception research, empowering autonomous vehicles to collaborate seamlessly and safely in dynamic settings. By enabling vehicles to share radar-based sensory data, our dataset aims to improve the robustness and accuracy of perception systems under diverse environmental conditions, including challenging weather and low visibility. This enhanced level of cooperation between autonomous vehicles is expected to lead to greater environmental awareness.

Chapter 4

3D Bounding Box Estimation using V2VRP Dataset

3D bounding box estimation has become a core task in the perception systems of autonomous vehicles, where accurately detecting and localizing objects such as other vehicles, pedestrians, and obstacles is crucial [134–145]. Traditional 2D object detection falls short in such environments, as it lacks depth information and spatial orientation. In contrast, 3D bounding box estimation provides a more comprehensive view by determining the precise location, orientation, and dimensions of objects within three-dimensional space. This information is essential for robust path planning and obstacle avoidance.

Machine learning models, particularly deep learning architectures, have significantly improved 3D bounding box estimation, allowing models to accurately process data from sensors such as LiDAR, radar, and stereo cameras [146–152]. In autonomous driving, datasets like KITTI, Waymo Open Dataset, and nuScenes have established benchmarks for 3D object detection research [120–122], offering real-world annotated data for training and evaluating models. However, these datasets rely on LiDAR point clouds, which are costly for data transmission due to their large file sizes, and LiDAR sensors themselves are expensive. Compared to LiDAR, FMCW radars are more affordable, produce smaller file sizes, and are already standard in many cars today. By leveraging FMCW radar point clouds, our approach ensures that models remain robust while also being cost-effective and data-efficient, enabling faster data processing and improved real-time object

detection. This allows us to develop perception systems that are both reliable and scalable for deployment in real-world autonomous vehicles, addressing critical requirements such as affordability, data efficiency, and performance across diverse environments.

In this work, we utilize the FMCW radar point cloud dataset, V2VRP, specifically designed for cooperative perception in 3D bounding box estimation. Our main contributions are as follows:

- We demonstrate that V2VRP can be effectively used with existing machine learning models for 3D bounding box estimation.
- We present preliminary results for 3D bounding box estimation using our cooperative dataset, achieving a maximum IoU of 0.67 and an mAP score of 0.6 at an IoU threshold of 0.1.
- We present comparative results demonstrating that combined view data samples outperform single view samples. The maximum IoU improved by 70% in the combined view compared to the single view, and the mAP score at an IoU threshold of 0.1 was 15% higher with the combined view approach.

4.1 Preliminaries

4.1.1 Ground Truth Label Annotation

Numerous labeling tools are available for generating ground truth annotations from RGB images, which are essential for training and validating machine learning models in computer vision. These tools enable users to precisely mark objects, regions, or features within images using bounding boxes, and segmentation masks, making them invaluable for tasks like object detection, semantic segmentation, and pose estimation. Popular tools such as LabelImg, Labelbox, and CVAT (Computer Vision Annotation Tool) support various annotation formats and offer user-friendly interfaces to streamline the labeling process. Additionally, some tools feature AI-assisted labeling, which accelerates annotation by making initial predictions that human annotators can quickly refine. This diverse range of labeling tools provides researchers and developers with robust options for creating

(a) Initial setup of the calibration chess board positioned outdoors.

(b) Detected corners on the chess board pattern, marked with colored dots, illustrating successful corner detection for camera calibration.

Figure 4.1: Calibration setup for camera calibration in ground truth annotation.

high-quality datasets tailored to their specific project needs. However, since these tools work with RGB images, they lack depth information, making them challenging to apply directly to 3D point cloud ground truth annotation.

Among recent advances, one model, Monocular 3D Object Detection with Depth-Aware Transformer (MonoDTR), introduces a novel approach to 3D object detection using only monocular images, addressing the complex challenge of estimating depth from single-view inputs [153]. In [153], it proposes a Depth-Aware Transformer (DTR) that enhances depth estimation by integrating image features with depth priors, improving spatial understanding and enabling accurate 3D perception. MonoDTR employs a two-stage detection pipeline: an initial region proposal network (RPN) generates candidate bounding boxes, and a depth-aware transformer refines these by incorporating depth information. This transformer-based framework uses self-attention mechanisms to learn depth representations, making it more accurate in estimating 3D object locations. By integrating depth-aware attention modules, MonoDTR achieves significant improvements in monocular 3D object detection accuracy, especially in complex scenes with varying object distances.

To obtain accurate depth information for use with MonoDTR, precise camera calibration is essential. This process, often performed using OpenCV, is critical for applications requiring accurate 3D information, such as image correction, depth estimation, and 3D reconstruction [154]. Camera calibration involves determining both intrinsic parameters (such as focal length, optical center, and distortion coefficients) and extrinsic parameters (rotation and translation vectors) for the camera. For example, as shown in Fig. 4.1, the process typically begins with a controlled setup using a calibration object, such as a chessboard or dot grid pattern, with a known number of rows, columns, and square sizes. A chessboard is often preferred for calibration as it provides well-defined corner points that are easily detected. Capturing multiple images of this pattern (typically 10-15) from various angles and distances improves calibration accuracy, ensuring the pattern is visible across a diverse field of view.

For each calibration image, 3D object points representing each corner in the calibration

pattern are defined, assuming fixed distances between points and usually setting $z = 0$ for a flat pattern. OpenCV's *cv2.findChessboardCorners* can then detect the 2D corner positions in each image, which can be further refined with *cv2.cornerSubPix* for improved precision in high-resolution images. These object points and image points are stored in corresponding lists for use in calibration. Using OpenCV's *cv2.calibrateCamera*, the function computes the camera matrix, distortion coefficients, both rotation and translation vectors. The camera matrix, which contains intrinsic parameters such as focal length (f_x, f_y) and optical center (c_x, c_y), along with distortion coefficients that correct lens distortions, enables accurate 3D modeling.

To verify calibration accuracy, an image can be undistorted with *cv2.undistort* using the calculated camera matrix and distortion coefficients. Additionally, reprojection error can be computed by projecting the 3D object points back to the image plane and comparing them with the detected 2D points. A lower reprojection error indicates a more accurate calibration. Once calibration is complete, the camera matrix and distortion coefficients can be saved for future use, streamlining the calibration process and enabling reliable camera parameters to support various 3D vision tasks.

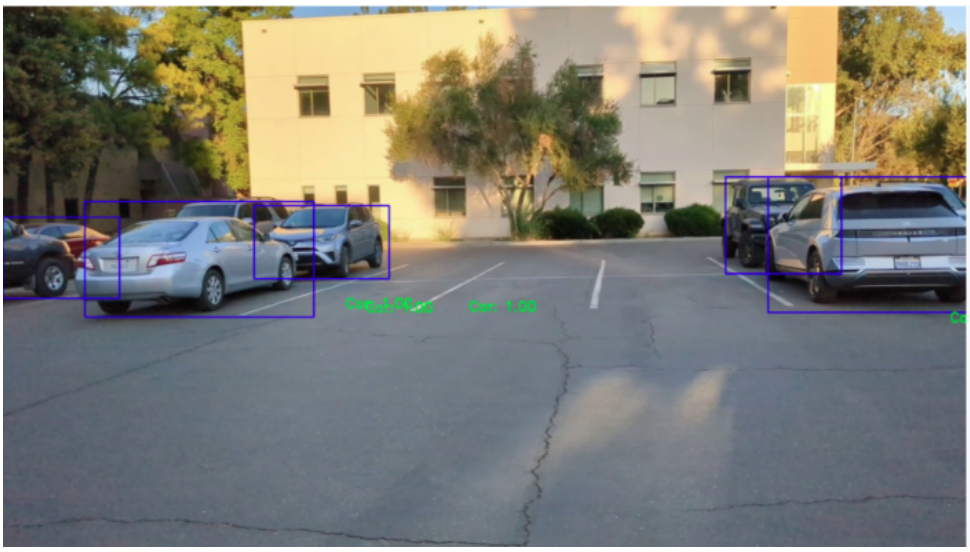
After getting camera calibration matrices, we can use MonoDTR to label by predicting 3D ground truth bounding box with our images as shown in Fig. 4.2.

4.1.2 Evaluation Parameters

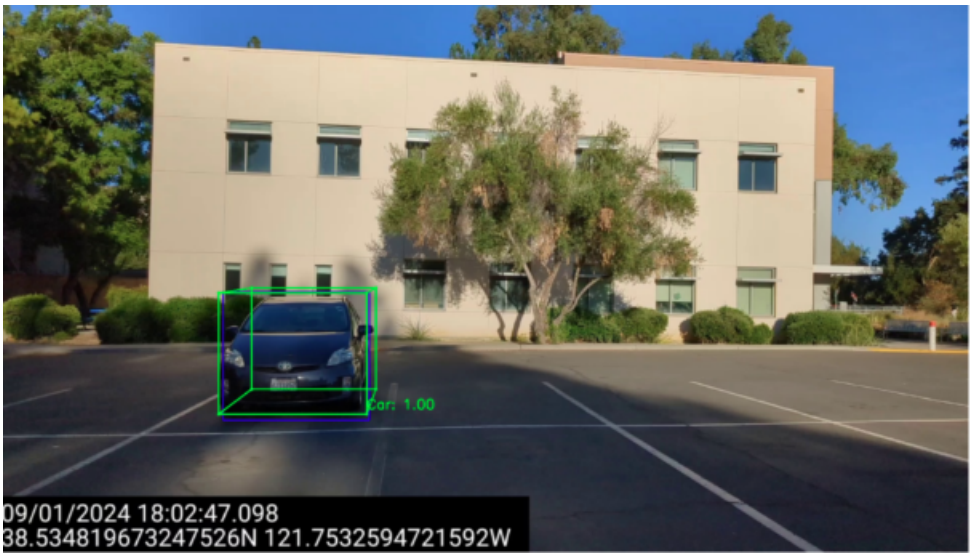
Intersection over Union (IoU) and Mean Average Precision (mAP) are standard evaluation metrics used to measure the accuracy of object detection models, assessing both the precision of bounding box placement and overall detection performance across classes. These metrics allowed us to evaluate the model's effectiveness in accurately locating and classifying objects within the scene.

IoU measures the overlap between the predicted bounding box and the ground truth bounding box. It is defined as the ratio of the intersection area to the union area of two bounding boxes, and it can be written as

$$\text{2D IoU} = \frac{\text{Intersection Area}}{\text{Union Area}} \tag{4.1}$$



(a) 2D bounding box estimation of vehicles using MonoDTR, illustrating the model's ability to detect objects in the image plane.



(b) 3D bounding box estimation of a vehicle using MonoDTR, demonstrating depth estimation and spatial localization in the scene.

Figure 4.2: Example of 2D and 3D Bounding Box Estimation Using MonoDTR

For 3D IoU, it represents the overlap volume between the prediction and the ground truth, and it can be shown as

$$\text{3D IoU} = \frac{\text{Intersection Volume}}{\text{Union Volume}} \tag{4.2}$$

An IoU value of 1 indicates a perfect overlap, meaning the predicted box exactly matches the ground truth, while a value of 0 means there is no overlap. Higher IoU values represent better predictions, as they signify closer alignment with the ground truth. IoU is essential in object detection, as it helps determine whether a detected object qualifies as a true positive (if it surpasses a certain IoU threshold) or a false positive. If two boxes of equal size overlap by half, they would yield an IoU of 0.33. Consequently, an IoU of approximately 0.5 is typically seen as a satisfactory level of overlap.

Mean Average Precision (mAP) is a crucial evaluation metric in object detection that assesses the accuracy of a model in predicting object locations and classifications. The mAP score is calculated by taking the average of the Average Precision (AP) scores across all object classes. AP itself is derived by plotting the precision-recall curve for a particular class and calculating the area under this curve.

$$\textit{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4.3}$$

$$\textit{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4.4}$$

where TP is a true positive, FP is a false positive and FN is a false negative.

Precision measures how accurately the model's predictions match the actual objects (true positives versus false positives). In contrast, recall measures the model's ability to detect all instances of an object (true positives versus missed detections). The mAP score is often computed at multiple Intersection over Union (IoU) thresholds to test the model's robustness at different levels of localization accuracy. A higher mAP score indicates better overall model performance, as it demonstrates both high precision and recall across classes.

4.2 Related Work

3D perception is essential in autonomous driving, enabling vehicles to detect, locate, and classify objects in 3D space. Sensors like LiDAR, radar, and stereo cameras provide the spatial data needed for this task. Large datasets such as KITTI, Waymo Open Dataset, and nuScenes have accelerated 3D perception model development by providing annotated 3D data for benchmarking [120–122].

Early 3D models relied on LiDAR for its high spatial accuracy and depth resolution. Models like PointRCNN and Voxelnet use raw LiDAR point clouds to predict object locations [155–161]. Despite its accuracy, LiDAR is costly and data-intensive, posing challenges for real-time processing. Though depth estimation is more challenging, monocular and stereo cameras offer a cost-effective alternative to LiDAR. For example, MonoDTR use depth-aware transformers to improve depth estimation, achieving strong results in monocular 3D detection, though they still fall short of LiDAR accuracy, especially at long distances [153]. FMCW radar has emerged as a viable 3D perception sensor due to its affordability, robustness in poor weather, and efficient data handling. FMCW radar provides lower-resolution point clouds than LiDAR but allows faster processing, making it suitable for real-time applications [162–168]. However, no existing work has explored the use of cooperative radar point cloud datasets for 3D object detection with dual views from different vehicles.

Our work leverages FMCW radar point clouds for 3D bounding box estimation, focusing on cost efficiency and data speed. Using the FMCW radar point cloud, our approach provides an affordable, scalable solution for real-time 3D perception, contributing to robust, accessible autonomous driving systems.

4.3 3D Bounding Box Estimation

4.3.1 Network Architecture

We adapted the deep learning architecture, RP-net, to evaluate our dataset [5]. RP-Net is similar to PointNet in that it processes point clouds, but it enhances the model by focusing on relational features between points [169]. This network innovatively generates

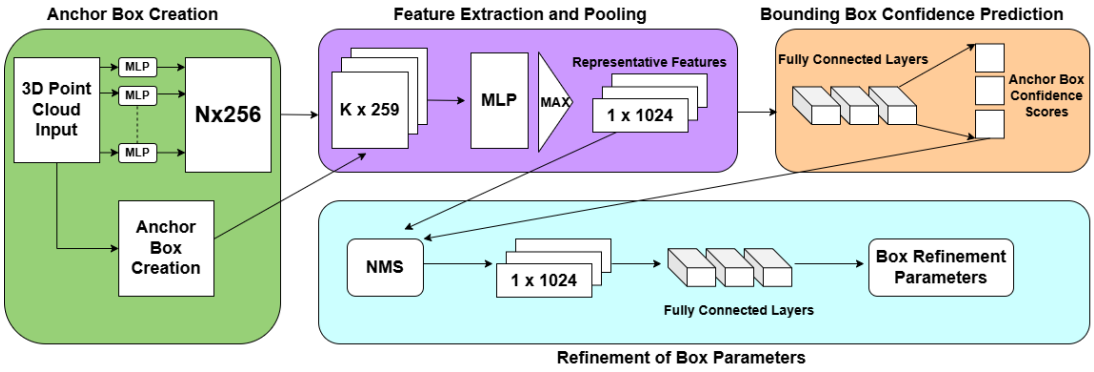


Figure 4.3: RP-net model architecture. Adapted from [5].

region proposals (anchor boxes) based on radar point clouds related to vehicle geometry. The sparsity of radar point cloud data and that most points originate from vehicle surfaces enable us to define region proposals based on individual points. These fixed-size anchor boxes, representing initial estimates of 3D bounding boxes, are sized according to the average vehicle dimensions in the training dataset. Instead of processing the entire point cloud at once, our network operates on each anchor box individually. Figure 4.3 The model comprises the following key components:

Anchor Box Creation: For the object with the highest reflection in scenes, this model defines a region proposal by placing anchor boxes around the object. A point-based anchor box generation method that leverages radar responses influenced by the object's geometry and spatiotemporal coherence. For the designated anchor point, five anchor boxes are placed around it, with each box's orientation angle determined by the spatiotemporal coherence values associated with the anchor point.

Segmentation through Feature Extraction and Pooling: The network's goal is to perform classification and 3D bounding box parameter regression by learning meaningful feature representations from the point cloud data [169]. RP-net extracts features in two stages. In the first stage, a PointNet encoder with shared MLP layers extracts features from the entire point cloud . In the second stage, anchor boxes are generated for each point. An RoI (Region of Interest) feature pooling block gathers features from all points within each anchor box. These features are passed through another PointNet layer and then max-pooled to create a single representative feature for each anchor box.

Bounding Box Confidence Prediction: The representative features of all anchor boxes from the previous block are passed through a classification network composed of fully connected layers. This network maps the anchor box features to a confidence score for each box. Performing classification on RoI-based max-pooled features ensures that the network captures contextual information from all nearby points within each anchor box, enhancing classification accuracy. By classifying each anchor box, the network identifies the boxes most likely to contain an object.

Refinement of Box Parameters: Initially, the anchor boxes are rough estimates of object dimensions, center, and orientation based on fixed-size assumptions. To achieve more accurate bounding boxes, we refine these parameters. After classification, confidence scores are available for all anchor boxes, but overlapping high-confidence boxes may represent the same object. Non-maximal suppression (NMS) is applied to eliminate overlapping boxes with lower confidence. The remaining anchor box features are passed through three fully connected layers to predict a refined tuple $[h', w', l', x', y', z', \theta']$, representing adjustments for length, width, height, center coordinates, and orientation angle. These refinements are added to the original anchor box parameters, resulting in the final 3D bounding box prediction.

Loss Functions The classification of anchor boxes in the first stage is a binary classification problem, handled with cross-entropy loss, defined as:

$$L_{\text{RPN}} = \sum_{i=1}^{N} - \left(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)\right) \tag{4.5}$$

where y_i is the ground truth label $[0, 1]$ and p_i is the predicted confidence score. The bounding box refinement stage is treated as a regression problem using Smooth-L1 loss:

$$L_{\text{refinement}}(r, r') = \begin{cases} \frac{1}{2}(r - r')^2 & \text{if } |r - r'| < 1, \\ \delta|r - r'| - \frac{1}{2} & \text{otherwise,} \end{cases} \tag{4.6}$$

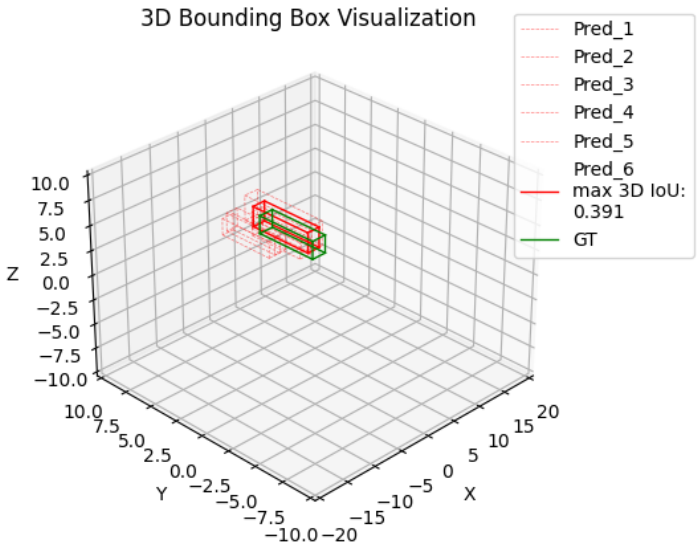
where r and r' represent the ground truth and regressed refinement values, respectively, for each parameter $[h', w', l', x', y', z', \theta']$.

4.3.2 Evaluation

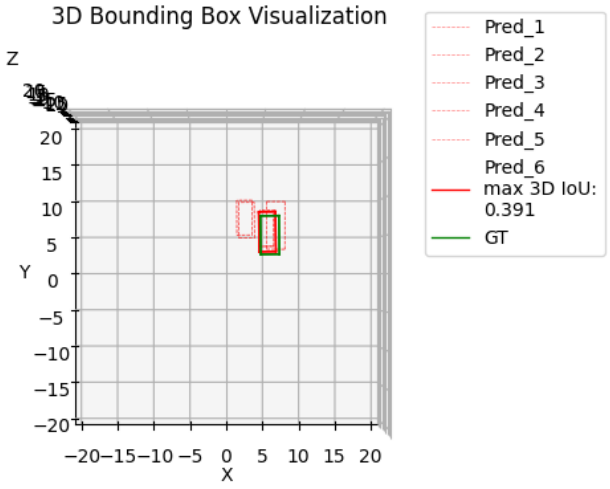
The deep learning model was implemented using the PyTorch framework. The Adam optimizer was used with a learning rate of 0.00002 and a momentum of 0.9. The model was trained over 300 epochs in two different cases to compare single-view data with combined-view data. In the first case, we used the public dataset provided by Pointillism [5], which includes 48 scenes with 14,208 combined radar frames, alongside our single-view dataset containing 19 scenes with 12,920 radar frames from a single vehicle. In the second case, we repeated the training but combined 12,920 radar frames by merging perspectives from both the ego and assistant vehicles rather than using data from a single vehicle. We split the data into training and testing sets with a 9:1 ratio.

Figure 4.4 shows the 3D bounding box estimation with a single view. Both the 3D camera view and bird's eye view display the ground truth bounding box alongside the predicted bounding boxes, with the predicted box achieving the highest 3D IoU score highlighted in a solid red line. The maximum IoU obtained here is 0.39, and the mAP score is 0.52 for an IoU threshold of 0.1. This relatively low IoU score indicates limited overlap between the predicted box and the ground truth, likely due to the absence of additional spatial information that multiple perspectives could have provided.

Figure 4.5 illustrates the 3D bounding box estimation for cooperative perception, demonstrating the model's ability to more accurately localize objects by leveraging multiple viewpoints. The use of cooperative perception in this figure demonstrates a clear advantage over single-view data. By combining perspectives, the model achieves better spatial coherence and alignment, as shown by the higher IoU score of 0.67 and the mAP score of 0.60 for an IoU threshold of 0.1. This outperforming result in 3D IoU suggests that cooperative perception significantly enhances the model's ability to localize objects accurately, which is less achievable in single-view setups where occlusions and limited viewpoints may reduce accuracy.

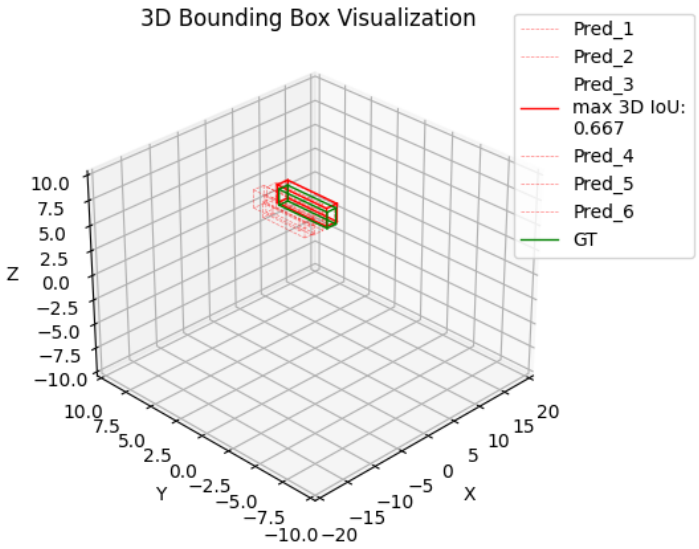


(a) 3D camera view of 3D bounding boxes.

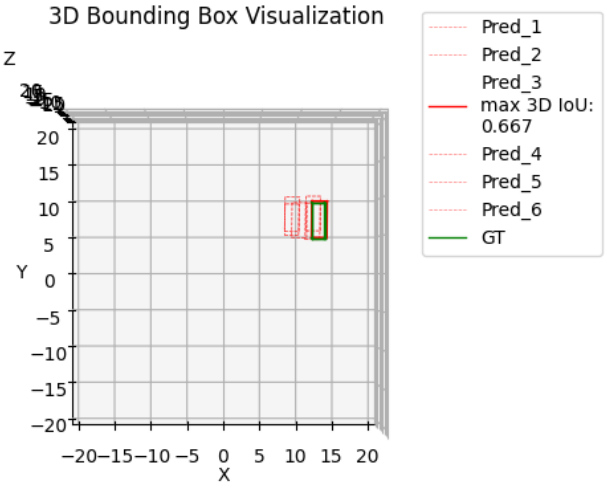


(b) Bird's eye view of the same bounding boxes.

Figure 4.4: Visualization of 3D bounding box estimation for single radar perception. The predicted box with the maximum 3D IoU (0.391) is highlighted with a solid red line, while other predictions are shown with dashed lines.



(a) 3D camera view of 3D bounding boxes



(b) Bird's eye view of the same bounding boxes.

Figure 4.5: Visualization of 3D bounding box estimation for cooperative perception. The predicted box with the maximum 3D IoU (0.667) is highlighted with a solid red line, while other predictions are shown with dashed lines.

4.4 Summary

In conclusion, this work leverages the V2VRP FMCW radar point cloud dataset, specially designed for cooperative perception in 3D bounding box estimation. We have shown that V2VRP can be effectively utilized with the existing machine learning model for 3D object localization, providing valuable insights into the benefits of cooperative perception. Preliminary results demonstrate promising performance, with a maximum IoU of 0.67 and an mAP score of 0.59 at an IoU threshold of 0.1. Additionally, our comparative analysis between single-view and combined-view 3D bounding box estimation shows a 70% improvement in max 3D IoU and a 15% improvement in mAP at the same IoU threshold (0.1) for the combined view. This significant improvement highlights the benefits of a cooperative perception dataset, enhancing model accuracy and robustness in 3D perception tasks.

Chapter 5

Conclusion

This dissertation explores the potential of Frequency Modulated Continuous Wave (FMCW) radar to enhance millimeter wave (mmWave) sensing for vehicular communication, focusing on improving communication efficiency and environmental perception. One of the key findings is the effectiveness of integrating FMCW radar data into beamforming protocols for Vehicle-to-Vehicle (V2V) communication. This approach addresses challenges such as mobility and path blockage, which are common in dynamic vehicular scenarios. The radar-assisted beamforming protocol demonstrated significant improvements in communication throughput during real-world testing, highlighting the value of using environmental data to optimize mmWave communication links.

Another primary inference from this thesis is developing a cooperative perception framework to address the limitations of sparse point cloud data from FMCW radar. By leveraging synchronized data from multiple vehicles, the framework achieved a more comprehensive and robust understanding of the environment. This approach highlights the benefits of cooperative perception in enhancing situational awareness and safety for autonomous vehicles, allowing connected vehicles to share sensory data for a more accurate representation of their surroundings. The cooperative perception framework is supported by introducing a novel radar point cloud dataset tailored for V2V communication, filling a critical gap in autonomous driving research. This dataset, which includes real-world radar point cloud data synchronized with GPS and camera information, offers valuable resources for advancing machine learning models for autonomous perception. Valida-

tion of the cooperative perception dataset using a 3D bounding box estimation model demonstrated its effectiveness in advancing machine learning applications for autonomous driving. The results showed that the cooperative perception approach significantly outperformed single-view methods, achieving an improvement in maximum Intersection over Union (IoU). These findings illustrate the potential of cooperative perception to enhance object detection and tracking capabilities in dynamic and complex environments.

However, based on these inferences, we also identified limitations of using FMCW radars. Since FMCW radar provides only limited point cloud data representing the environment, it may not effectively pinpoint the receiver module for the target vehicle. Additionally, while cooperative sensing of radar point clouds increases overall density, the data still remains sparser compared to LiDAR point clouds. Future work could focus on addressing these challenges in several key areas. One avenue is the development of advanced signal processing algorithms to more robustly reconstruct environmental details. Another is the development of deep learning models specifically tailored for radar point clouds, which could further enhance object detection, tracking, and classification in dynamic environments. Lastly, extending the cooperative perception framework to enable real-time data processing and sharing across multiple vehicles would be highly beneficial. Achieving this would require the development of efficient V2V communication protocols to overcome challenges related to bandwidth and latency.

By addressing these areas, the findings from my thesis lay the groundwork for more reliable, scalable, and efficient autonomous driving systems. These advancements not only improve road safety and traffic management but also contribute to the broader adoption of autonomous technologies in real-world scenarios. Moreover, this research opens avenues for innovative applications in other domains, such as healthcare and urban mobility, showcasing the versatility and potential of FMCW radar technology.

5.1 Future Work

Millimeter wave sensing with FMCW radar has applications that extend well beyond vehicular communication, including promising uses in human health monitoring. One

compelling area of future research is using this technology to monitor early-stage neurodegenerative disorders. Conditions like Parkinson's disease and Alzheimer's disease gradually impair motor function and cognitive abilities. Early detection of these conditions is crucial for managing these conditions, but traditional methods often rely on invasive procedures. FMCW radar, with its capability to track fine movements and vital signs remotely, presents a non-invasive monitoring.

FMCW radar can detect subtle changes in movement patterns, gait, and posture—often early indicators of neurodegenerative disease progression [170]. Additionally, FMCW radar can measure vital signs such as respiration and heart rate, which are important for neurodegenerative conditions that impact autonomic function. However, substantial research in this area remains limited.

Future work could involve integrating FMCW radar with machine learning models to detect patterns indicative of neurodegenerative disorders through real-time data analysis. Combining radar data with other health metrics, such as cognitive assessments and sleep patterns, could create a comprehensive monitoring system tailored to neurodegenerative disorders.

References

[1] MATLAB, "Constant False Alarm Rate (CFAR) Detection." Website: https://www.mathworks.com/help/phased/ug/constant-false-alarm-rate-cfar-detection.html.

[2] RohdeSchwarz, "802.11ad - WLAN at 60 GHz: A Technology Introduction." Website: https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/dl_application/application_notes/1ma220/1MA220_3e_WLAN_11ad_WP.pdf.

[3] OXTS, "What is RTK?." Website: https://www.oxts.com/rtk/, 2020.

[4] UNAVCO, "UNAVCO Network Monitoring." Website: https://www.unavco.org/instrumentation/networks/map/.

[5] K. Bansal, K. Rungta, S. Zhu, and D. Bharadia, "Pointillism: Accurate 3d bounding box estimation with multi-radars," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pp. 340–353, 2020.

[6] NHTSA, "Early estimate of motor vehicle traffic fatalities for the first quarter of 2024," *NHTSA*.

[7] A. Venon, Y. Dupuis, P. Vasseur, and P. Merriaux, "Millimeter wave fmcw radars for perception, recognition and localization in automotive applications: A survey," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 533–555, 2022.

[8] L. Xu, S. Sun, K. V. Mishra, and Y. D. Zhang, "Automotive fmcw radar with difference co-chirps," *IEEE Transactions on Aerospace and Electronic Systems*, 2023.

[9] F. Folster, H. Rohling, and U. Lubbert, "An automotive radar network based on 77 ghz fmcw sensors," in *IEEE International Radar Conference, 2005.*, pp. 871–876, IEEE, 2005.

[10] M. Rameez, M. I. Pettersson, and M. Dahl, "Interference compression and mitigation for automotive fmcw radar systems," *IEEE Sensors Journal*, vol. 22, no. 20, pp. 19739–19749, 2022.

[11] A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, and U. J. Nunes, "Depthcn: Vehicle detection using 3d-lidar and convnet," in *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)*, pp. 1–6, IEEE, 2017.

[12] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, "A lidar and vision-based approach for pedestrian and vehicle detection and tracking," in *2007 IEEE intelligent transportation systems conference*, pp. 1044–1049, IEEE, 2007.

[13] F. Zhang, D. Clarke, and A. Knoll, "Vehicle detection based on lidar and camera fusion," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1620–1625, IEEE, 2014.

[14] A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, and U. J. Nunes, "Multimodal vehicle detection: fusing 3d-lidar and color camera data," *Pattern Recognition Letters*, vol. 115, pp. 20–29, 2018.

[15] J. Cheng, Z. Xiang, T. Cao, and J. Liu, "Robust vehicle detection using 3d lidar under complex urban environment," in *2014 IEEE international conference on robotics and automation (ICRA)*, pp. 691–696, IEEE, 2014.

[16] H. Wang and X. Zhang, "Real-time vehicle detection and tracking using 3d lidar," *Asian Journal of Control*, vol. 24, no. 3, pp. 1459–1469, 2022.

[17] K. Qian, S. Zhu, X. Zhang, and L. E. Li, "Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 444–453, 2021.

[18] H. Wang, X. Lou, Y. Cai, Y. Li, and L. Chen, "Real-time vehicle detection algorithm based on vision and lidar point cloud fusion," *Journal of Sensors*, vol. 2019, no. 1, p. 8473980, 2019.

[19] G. Wang, J. Wu, T. Xu, and B. Tian, "3d vehicle detection with rsu lidar for autonomous mine," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 344–355, 2021.

[20] L. Guan, Y. Chen, G. Wang, and X. Lei, "Real-time vehicle detection framework based on the fusion of lidar and camera," *Electronics*, vol. 9, no. 3, p. 451, 2020.

[21] V. Vaquero, I. del Pino, F. Moreno-Noguer, J. Solà, A. Sanfeliu, and J. Andrade-Cetto, "Dual-branch cnns for vehicle detection and tracking on lidar data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6942–6953, 2020.

[22] X. Jin, H. Yang, X. He, G. Liu, Z. Yan, and Q. Wang, "Robust lidar-based vehicle detection for on-road autonomous driving," *Remote Sensing*, vol. 15, no. 12, p. 3160, 2023.

[23] A. Ceccarelli and F. Secci, "Rgb cameras failures and their effects in autonomous driving applications," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 4, pp. 2731–2745, 2022.

[24] K. El Madawi, H. Rashed, A. El Sallab, O. Nasr, H. Kamel, and S. Yogamani, "Rgb and lidar fusion based 3d semantic segmentation for autonomous driving," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 7–12, IEEE, 2019.

[25] Q. He, Z. Wang, H. Zeng, Y. Zeng, Y. Liu, S. Liu, and B. Zeng, "Stereo rgb and deeper lidar-based network for 3d object detection in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 152–162, 2022.

[26] B. Li, W. Ouyang, L. Sheng, X. Zeng, and X. Wang, "Gs3d: An efficient 3d object detection framework for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1019–1028, 2019.

[27] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li, and D. Cao, "Deep learning for image and point cloud fusion in autonomous driving: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 722–739, 2021.

[28] F. Secci and A. Ceccarelli, "Rgb cameras failures and their effects in autonomous driving applications," *arXiv preprint arXiv:2008.05938*, 2020.

[29] X. Ma, Z. Wang, H. Li, P. Zhang, W. Ouyang, and X. Fan, "Accurate monocular 3d object detection via color-embedded 3d reconstruction for autonomous driving," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6851–6860, 2019.

[30] D. E. Barrick, *FM/CW radar signals and digital processing*, vol. 55. Environmental Research Laboratories, 1973.

[31] M. Khan, P. Mahajan, G. N. Khan, D. Chaudhary, J. Benny, M. Wajid, and A. Srivastava, "Design and implementation of fpga based system for object detection and range estimation used in adas applications utilizing fmcw radar," in *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, IEEE, 2024.

[32] S. Biswas, B. Bartlett, J. E. Ball, and A. C. Gurbuz, "Classification of traffic signaling motion in automotive applications using fmcw radar," in *2023 IEEE Radar Conference (RadarConf23)*, pp. 1–6, IEEE, 2023.

[33] Z. E. Ekolle, R. Kohno, H. Ochiai, S. Sadamasa, I. Ikenji, I. Hiroshi, and O. Naomi, "A reliable 79ghz band ultra-short range radar for adas/ad vehicles using fmcw technology," in *2023 IEEE International Automated Vehicle Validation Conference (IAVVC)*, pp. 1–6, IEEE, 2023.

[34] C. Zeintl, F. Eibensteiner, and J. Langer, "Evaluation of fmcw radar for vibration sensing in industrial environments," in *2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–5, IEEE, 2019.

[35] M. van Delden, C. Guzy, and T. Musch, "Investigation on a system for positioning of industrial robots based on ultra-broadband millimeter wave fmcw radar," in *2019 IEEE Asia-Pacific Microwave Conference (APMC)*, pp. 744–746, IEEE, 2019.

[36] R. Rouveure, P. Faure, and M.-O. Monod, "Pelican: Panoramic millimeter-wave radar for perception in mobile robotics applications, part 1: Principles of fmcw radar and of 2d image construction," *Robotics and Autonomous Systems*, vol. 81, pp. 1–16, 2016.

[37] K. Harlow, H. Jang, T. D. Barfoot, A. Kim, and C. Heckman, "A new wave in robotics: Survey on recent mmwave radar applications in robotics," *IEEE Transactions on Robotics*, 2024.

[38] J.-J. Lin, Y.-P. Li, W.-C. Hsu, and T.-S. Lee, "Design of an fmcw radar baseband signal processing system for automotive application," *SpringerPlus*, vol. 5, pp. 1–16, 2016.

[39] F. Uysal, "Phase-coded fmcw automotive radar: System design and interference mitigation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 270–281, 2019.

[40] L. Yi, W. Hongxian, X. Mengdao, and B. Zheng, "Imaging study of high squint sar based on fmcw," in *2007 1st Asian and Pacific Conference on Synthetic Aperture Radar*, pp. 6–9, IEEE, 2007.

[41] A. Meta, P. Hoogeboom, and L. P. Ligthart, "Signal processing for fmcw sar," *IEEE Transactions on geoscience and remote sensing*, vol. 45, no. 11, pp. 3519–3532, 2007.

[42] M. A. Richards *et al.*, *Fundamentals of radar signal processing*, vol. 1. Mcgraw-hill New York, 2005.

[43] P. P. Gandhi and S. A. Kassam, "Analysis of cfar processors in nonhomogeneous background," *IEEE Transactions on Aerospace and Electronic systems*, vol. 24, no. 4, pp. 427–445, 1988.

[44] V. G. Hansen, "Constant false alarm rate processing in search radars. in radar—present and future," in *IEE Conf Publ*, vol. 105, p. 325, 1973.

[45] TexasInstruments, "AWR1843BOOST and IWR1843BOOST Single-Chip mmWave Sensing Solution User's Guide (Rev. B)." Website: `https://www.ti.com/lit/ug/spruim4b/spruim4b.pdf?ts=1731303098383`.

[46] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6g: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[47] H. Viswanathan and P. E. Mogensen, "Communications in the 6g era," *IEEE access*, vol. 8, pp. 57063–57074, 2020.

[48] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6g be?," *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[49] M. Alsabah, M. A. Naser, B. M. Mahmmod, S. H. Abdulhussain, M. R. Eissa, A. Al-Baidhani, N. K. Noordin, S. M. Sait, K. A. Al-Utaibi, and F. Hashim, "6g wireless communications networks: A comprehensive survey," *IEEE Access*, vol. 9, pp. 148191–148243, 2021.

[50] O. Abari, D. Bharadia, A. Duffield, and D. Katabi, "Cutting the cord in virtual reality," in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, pp. 162–168, 2016.

[51] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, *et al.*, "Towards fully autonomous driving: Systems and algorithms," in *2011 IEEE intelligent vehicles symposium (IV)*, pp. 163–168, IEEE, 2011.

[52] B. Yang, X. Cao, K. Xiong, C. Yuen, Y. L. Guan, S. Leng, L. Qian, and Z. Han, "Edge intelligence for autonomous driving in 6g wireless system: Design challenges and solutions," *IEEE Wireless Communications*, vol. 28, no. 2, pp. 40–47, 2021.

[53] L.-H. Shen, K.-T. Feng, and L. Hanzo, "Five facets of 6g: Research challenges and opportunities," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–39, 2023.

[54] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[55] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6g wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[56] H. Guo, X. Zhou, J. Liu, and Y. Zhang, "Vehicular intelligence in 6g: Networking, communications, and computing," *Vehicular Communications*, vol. 33, p. 100399, 2022.

[57] M. Banafaa, I. Shayea, J. Din, M. H. Azmi, A. Alashbi, Y. I. Daradkeh, and A. Alhammadi, "6g mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities," *Alexandria Engineering Journal*, vol. 64, pp. 245–274, 2023.

[58] S. I. Loutfi, I. Shayea, U. Tureli, A. A. El-Saleh, and W. Tashan, "An overview of mobility awareness with mobile edge computing over 6g network: Challenges and future research directions," *Results in Engineering*, p. 102601, 2024.

[59] K. Hosoya, N. Prasad, K. Ramachandran, N. Orihashi, S. Kishimoto, S. Rangarajan, and K. Maruhashi, "Multiple sector id capture (midc): A novel beamforming technique for 60-ghz band multi-gbps wlan/pan systems," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 1, pp. 81–96, 2014.

[60] S. Sur, V. Venkateswaran, X. Zhang, and P. Ramanathan, "60 ghz indoor networking through flexible beams: A link-level profiling," in *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 71–84, 2015.

[61] Y. Zhu, Z. Zhang, Z. Marzi, C. Nelson, U. Madhow, B. Y. Zhao, and H. Zheng, "Demystifying 60ghz outdoor picocells," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, pp. 5–16, 2014.

[62] J. He, K. Yang, and H.-H. Chen, "6g cellular networks and connected autonomous vehicles," *IEEE Network*, vol. 35, no. 4, pp. 255–261, 2020.

[63] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou, *et al.*, "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9710–9719, 2021.

[64] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE access*, vol. 8, pp. 58443–58469, 2020.

[65] S. Jiang, G. Charan, and A. Alkhateeb, "Lidar aided future beam prediction in real-world millimeter wave v2i communications," *IEEE Wireless Communications Letters*, 2022.

[66] S. Jiang and A. Alkhateeb, "Computer vision aided beam tracking in a real-world millimeter wave deployment," in *2022 IEEE Globecom Workshops (GC Wkshps)*, pp. 142–147, IEEE, 2022.

[67] U. Demirhan and A. Alkhateeb, "Radar aided 6g beam prediction: Deep learning algorithms and real-world demonstration," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2655–2660, IEEE, 2022.

[68] H. Ku, J. Song, D. Zhang, P. Mohapatra, and P. Pathak, "Characterizing real-time radar-assisted beamforming in mmwave v2v links," in *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 168–176, IEEE, 2023.

[69] D. Zhang, P. S. Santhalingam, P. Pathak, and Z. Zheng, "Characterizing interference mitigation techniques in dense 60 ghz mmwave wlans," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–9, IEEE, 2019.

[70] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, pp. 226–231, 1996.

[71] G. Hamerly and C. Elkan, "Learning the k in k-means," *Advances in neural information processing systems*, vol. 16, 2003.

[72] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020.

[73] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: why and how you should (still) use dbscan," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[74] 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*, pp. 1–6, IEEE, 2019.

[75] D. Zhang, P. S. Santhalingam, P. Pathak, and Z. Zheng, "Networked beamforming in dense mmwave wlans," in *Proceedings of the 23rd Annual International Workshop on Mobile Computing Systems and Applications*, pp. 102–108, 2022.

[76] Z. Yang, P. H. Pathak, J. Pan, M. Sha, and P. Mohapatra, "Sense and deploy: Blockage-aware deployment of reliable 60 ghz mmwave wlans," in *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 397–405, IEEE, 2018.

[77] D. Zhang, M. Garude, and P. H. Pathak, "mmchoir: Exploiting joint transmissions for reliable 60ghz mmwave wlans," in *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 251–260, 2018.

[78] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmwave mimo systems," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1140–1143, 2016.

[79] H. Li, M. Li, Q. Liu, and A. L. Swindlehurst, "Dynamic hybrid beamforming with low-resolution pss for wideband mmwave mimo-ofdm systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2168–2181, 2020.

[80] J. Zhan and X. Dong, "Interference cancellation aided hybrid beamforming for mmwave multi-user massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2322–2336, 2021.

[81] W. Ma, C. Qi, and G. Y. Li, "Machine learning for beam alignment in millimeter wave massive mimo," *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 875–878, 2020.

[82] D. Zhang, B. Han, P. Pathak, and H. Wang, "Innovating multi-user volumetric video streaming through cross-layer design," in *Proceedings of the Twentieth ACM Workshop on Hot Topics in Networks*, pp. 16–22, 2021.

[83] L. Yan, H. Ding, L. Zhang, J. Liu, X. Fang, Y. Fang, M. Xiao, and X. Huang, "Machine learning-based handovers for sub-6 ghz and mmwave integrated vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4873–4885, 2019.

[84] N. J. Myers, Y. Wang, N. González-Prelcic, and R. W. Heath, "Deep learning-based beam alignment in mmwave vehicular networks," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8569–8573, IEEE, 2020.

[85] I. Rasheed and F. Hu, "Intelligent super-fast vehicle-to-everything 5g communications with predictive switching between mmwave and thz links," *Vehicular Communications*, vol. 27, p. 100303, 2021.

[86] Y. Wu, L. Yan, and X. Fang, "A low-latency content dissemination scheme for mmwave vehicular networks," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7921–7933, 2019.

[87] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, and K. Chowdhury, "Deep learning on multimodal sensor data at the wireless edge for vehicular network," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7639–7655, 2022.

[88] W. Xu, F. Gao, X. Tao, J. Zhang, and A. Alkhateeb, "Computer vision aided mmwave beam alignment in v2x communications," *IEEE Transactions on Wireless Communications*, 2022.

[89] W. Yuan, F. Liu, C. Masouros, J. Yuan, D. W. K. Ng, and N. González-Prelcic, "Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1442–1456, 2020.

[90] B. Liu, J. Liu, and N. Kato, "Optimal beamformer design for millimeter wave dual-functional radar-communication based v2x systems," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 10, pp. 2980–2993, 2022.

[91] F. Liu, W. Yuan, C. Masouros, and J. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7704–7719, 2020.

[92] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, and M. Hollick, "Compressive millimeter-wave sector selection in off-the-shelf ieee 802.11 ad devices," in *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, pp. 414–425, 2017.

[93] J. Palacios, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer, "Adaptive codebook optimization for beam training on off-the-shelf ieee 802.11 ad devices," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pp. 241–255, 2018.

[94] Git Repository [Online]. Available: `https://github.com/ibaiGorordo/AWR1843-Read-Data-Python-MMWAVE-SDK-3-`.

[95] S. Hinderer, "What is adaptive cruise control?," *Kelley Blue Book*.

[96] R. Heaps, "Lane-keeping assist: Everything you need to know," *Kelley Blue Book*.

[97] R. Heaps, "Blind-spot monitors: Everything you need to know," *Kelley Blue Book*.

[98] A. Baker, "Zoox driverless robotaxis begin testing on streets of san francisco," *KRON4*.

[99] J. Elias, "Waymo opens robotaxi service to anyone in los angeles, marking its largest expansion yet," *CNBC*.

[100] F. Lambert, "Tesla spotted testing cybercab robotaxi on giga texas roads," *Electrek*.

[101] J. Van Brummelen, M. O'brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transportation research part C: emerging technologies*, vol. 89, pp. 384–406, 2018.

[102] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

[103] P. Koopman and M. Wagner, "Autonomous vehicle safety: An interdisciplinary challenge," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 1, pp. 90–96, 2017.

[104] A. Faisal, M. Kamruzzaman, T. Yigitcanlar, and G. Currie, "Understanding autonomous vehicles," *Journal of transport and land use*, vol. 12, no. 1, pp. 45–72, 2019.

[105] D. Parekh, N. Poddar, A. Rajpurkar, M. Chahal, N. Kumar, G. P. Joshi, and W. Cho, "A review on autonomous vehicles: Progress, methods and challenges," *Electronics*, vol. 11, no. 14, p. 2162, 2022.

[106] J. K. Choi and Y. G. Ji, "Investigating the importance of trust on adopting an autonomous vehicle," *International Journal of Human-Computer Interaction*, vol. 31, no. 10, pp. 692–702, 2015.

[107] K. Kritayakirana and J. C. Gerdes, "Autonomous vehicle control at the limits of handling," *International Journal of Vehicle Autonomous Systems*, vol. 10, no. 4, pp. 271–296, 2012.

[108] W. G. Najm, R. Ranganathan, G. Srinivasan, J. D. Smith, S. Toma, E. D. Swanson, A. Burgett, *et al.*, "Description of light-vehicle pre-crash scenarios for safety applications based on vehicle-to-vehicle communications," tech. rep., United States. Department of Transportation. National Highway Traffic Safety ..., 2013.

[109] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2583–2589, IEEE, 2022.

[110] Y. Li, Z. An, Z. Wang, Y. Zhong, S. Chen, and C. Feng, "V2x-sim: A virtual collaborative perception dataset for autonomous driving," *arXiv preprint arXiv:2202.08449*, 2022.

[111] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2x-vit: Vehicle-to-everything cooperative perception with vision transformer," in *European conference on computer vision*, pp. 107–124, Springer, 2022.

[112] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of sumo-simulation of urban mobility," *International journal on advances in systems and measurements*, vol. 5, no. 3&4, 2012.

[113] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*, pp. 1–16, PMLR, 2017.

[114] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan, *et al.*, "Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21361–21370, 2022.

[115] R. Xu, X. Xia, J. Li, H. Li, S. Zhang, Z. Tu, Z. Meng, H. Xiang, X. Dong, R. Song, *et al.*, "V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13712–13722, 2023.

[116] T. Takasu and A. Yasuda, "Development of the low-cost rtk-gps receiver with an open source program package rtklib," in *International symposium on GPS/GNSS*, vol. 1, pp. 1–6, International Convention Center Jeju Korea Seogwipo-si, Republic of Korea, 2009.

[117] X. Li, X. Zhang, X. Ren, M. Fritsche, J. Wickert, and H. Schuh, "Precise positioning with current multi-constellation global navigation satellite systems: Gps, glonass, galileo and beidou," *Scientific reports*, vol. 5, no. 1, p. 8328, 2015.

[118] Sparkfun, "What is gps rtk?," *Sparkfun*.

[119] H. T. Berglund, K. M. Hodgkinson, F. Blume, D. Mencin, D. A. Phillips, C. M. Meertens, and G. S. Mattioli, "Unavco real-time gnss positioning: High-precision static and kinematic testing of the next generation gnss network.," in *AGU Fall Meeting Abstracts*, vol. 2014, pp. G41B–0485, 2014.

[120] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 3354–3361, IEEE, 2012.

[121] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11621–11631, 2020.

[122] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2446–2454, 2020.

[123] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun, "V2vnet: Vehicle-to-vehicle communication for joint perception and prediction," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pp. 605–621, Springer, 2020.

[124] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3234–3243, 2016.

[125] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

[126] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2636–2645, 2020.

[127] J. Seo, J. Lee, J. Park, H. Kim, and S. You, "Distributed two-dimensional music for joint range and angle estimation with distributed fmcw mimo radars," *Sensors*, vol. 21, no. 22, p. 7618, 2021.

[128] M. Ren, P. He, and J. Zhou, "Improved shape-based distance method for correlation analysis of multi-radar data fusion in self-driving vehicle," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24771–24781, 2021.

[129] M. Rykunov, E. De Greef, K. Aziz, A. Bourdoux, H. Sahli, *et al.*, "Multi-radar fusion for failure-tolerant vulnerable road users classification," in *2021 18th European Radar Conference (EuRAD)*, pp. 337–340, IEEE, 2022.

[130] L. Favalli, P. Gamba, T. Gatti, and A. Mecocci, "Multi-radar data fusion for object tracking and shape estimation," *Signal processing*, vol. 48, no. 3, pp. 235–239, 1996.

[131] I. Jouny, "Target identification using multi-radar fusion," in *Automatic Target Recognition XVII*, vol. 6566, pp. 192–197, SPIE, 2007.

[132] M. N. Petsios, E. G. Alivizatos, and N. K. Uzunoglu, "Solving the association problem for a multistatic range-only radar target tracker," *Signal Processing*, vol. 88, no. 9, pp. 2254–2277, 2008.

[133] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su, "mmmesh: Towards 3d real-time dynamic human mesh construction using millimeter-wave," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 269–282, 2021.

[134] A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka, "3d bounding box estimation using deep learning and geometry," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 7074–7082, 2017.

[135] D. Xu, D. Anguelov, and A. Jain, "Pointfusion: Deep sensor fusion for 3d bounding box estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 244–253, 2018.

[136] C. Sahin, G. Garcia-Hernando, J. Sock, and T.-K. Kim, "A review on object pose recovery: From 3d bounding box detectors to full 6d pose estimators," *Image and Vision Computing*, vol. 96, p. 103898, 2020.

[137] M. M. Rahman, Y. Tan, J. Xue, L. Shao, and K. Lu, "3d object detection: Learning 3d bounding boxes from scaled down 2d bounding boxes in rgb-d images," *Information Sciences*, vol. 476, pp. 147–158, 2019.

[138] W. Ali, S. Abdelkarim, M. Zidan, M. Zahran, and A. El Sallab, "Yolo3d: End-to-end real-time 3d oriented object bounding box detection from lidar point cloud," in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

[139] Z. Deng and L. Jan Latecki, "Amodal detection of 3d objects: Inferring 3d bounding boxes from 2d ones in rgb-depth images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5762–5770, 2017.

[140] B. Yang, J. Wang, R. Clark, Q. Hu, S. Wang, A. Markham, and N. Trigoni, "Learning object bounding boxes for 3d instance segmentation on point clouds," *Advances in neural information processing systems*, vol. 32, 2019.

[141] V. Kocur and M. Ftáčnik, "Detection of 3d bounding boxes of vehicles using perspective transformation for accurate speed measurement," *Machine Vision and Applications*, vol. 31, no. 7, p. 62, 2020.

[142] J. Fang, L. Zhou, and G. Liu, "3d bounding box estimation for autonomous vehicles by cascaded geometric constraints and depurated 2d detections using 3d results," *arXiv preprint arXiv:1909.01867*, 2019.

[143] Y. Liu, N. Mishra, M. Sieb, Y. Shentu, P. Abbeel, and X. Chen, "Autoregressive uncertainty modeling for 3d bounding box prediction," in *European Conference on Computer Vision*, pp. 673–694, Springer, 2022.

[144] H. Siwei and L. Baolong, "Review of bounding box algorithm based on 3d point cloud," *International Journal of Advanced Network, Monitoring and Controls*, vol. 6, no. 1, pp. 18–23, 2021.

[145] X. Liu, C. Zheng, K. B. Cheng, N. Xue, G.-J. Qi, and T. Wu, "Monocular 3d object detection with bounding box denoising in 3d by perceiver," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6436–6446, 2023.

[146] Y. Li, L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao, and J. Li, "Deep learning for lidar point clouds in autonomous driving: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3412–3432, 2020.

[147] D. Wu, Z. Liang, and G. Chen, "Deep learning for lidar-only and lidar-fusion 3d perception: A survey," *Intelligence & Robotics*, vol. 2, no. 2, pp. 105–129, 2022.

[148] G. Zamanakos, L. Tsochatzidis, A. Amanatiadis, and I. Pratikakis, "A comprehensive survey of lidar-based 3d object detection methods with deep learning for autonomous driving," *Computers & Graphics*, vol. 99, pp. 153–181, 2021.

[149] H. Guan, Y. Yu, Z. Ji, J. Li, and Q. Zhang, "Deep learning-based tree classification using mobile lidar data," *Remote Sensing Letters*, vol. 6, no. 11, pp. 864–873, 2015.

[150] X. Du, M. H. Ang, and D. Rus, "Car detection for autonomous vehicle: Lidar and vision fusion approach through deep learning framework," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 749–754, IEEE, 2017.

[151] G. Melotti, C. Premebida, and N. Gonçalves, "Multimodal deep-learning for object recognition combining camera and lidar data," in *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pp. 177–182, IEEE, 2020.

[152] M. Mekala, W. Park, G. Dhiman, G. Srivastava, J. H. Park, and H.-Y. Jung, "Deep learning inspired object consolidation approaches using lidar data for autonomous driving: a review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2579–2599, 2022.

[153] K.-C. Huang, T.-H. Wu, H.-T. Su, and W. H. Hsu, "Monodtr: Monocular 3d object detection with depth-aware transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4012–4021, 2022.

[154] G. Bradski, "Learning opencv: Computer vision with the opencv library," *O'REILLY google schola*, vol. 2, pp. 334–352, 2008.

[155] C. Yi, Y. Zhang, Q. Wu, Y. Xu, O. Remil, M. Wei, and J. Wang, "Urban building reconstruction from raw lidar point data," *Computer-Aided Design*, vol. 93, pp. 1–14, 2017.

[156] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4490–4499, 2018.

[157] S. Shi, X. Wang, and H. Li, "Pointrcnn: 3d object proposal generation and detection from point cloud," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 770–779, 2019.

[158] J. Wang, Y. Xu, O. Remil, X. Xie, N. Ye, C. Yi, and M. Wei, "Automatic modeling of urban facades from raw lidar point data," in *Computer Graphics Forum*, vol. 35, pp. 269–278, Wiley Online Library, 2016.

[159] J. Wang and K. Xu, "Shape detection from raw lidar data with subspace modeling," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 9, pp. 2137–2150, 2016.

[160] M. Awrangjeb and C. S. Fraser, "Automatic segmentation of raw lidar data for extraction of building roofs," *Remote Sensing*, vol. 6, no. 5, pp. 3716–3751, 2014.

[161] G. Forlani, C. Nardinocchi, M. Scaioni, and P. Zingaretti, "C omplete classification of raw lidar data and 3d reconstruction of buildings," *Pattern analysis and applications*, vol. 8, pp. 357–374, 2006.

[162] R. Nabati and H. Qi, "Centerfusion: Center-based radar and camera fusion for 3d object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1527–1536, 2021.

[163] K. Hu, X. Hu, L. Qi, G. Lu, Y. Zhong, and Y. Han, "Radnet: A radar detection network for target detection using 3d range-angle-doppler tensor," in *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, pp. 1–6, IEEE, 2023.

[164] Y. Cheng, J. Su, H. Chen, and Y. Liu, "A new automotive radar 4d point clouds detector by using deep learning," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8398–8402, IEEE, 2021.

[165] N. Scheiner, F. Kraus, N. Appenrodt, J. Dickmann, and B. Sick, "Object detection for automotive radar point clouds–a comparison," *AI Perspectives*, vol. 3, no. 1, p. 6, 2021.

[166] B. Tan, Z. Ma, X. Zhu, S. Li, L. Zheng, S. Chen, L. Huang, and J. Bai, "3-d object detection for multiframe 4-d automotive millimeter-wave radar point cloud," *IEEE Sensors Journal*, vol. 23, no. 11, pp. 11125–11138, 2022.

[167] P. Svenningsson, F. Fioranelli, and A. Yarovoy, "Radar-pointgnn: Graph based object recognition for unstructured radar point-cloud data," in *2021 IEEE Radar Conference (RadarConf21)*, pp. 1–6, IEEE, 2021.

[168] G. Lu, Z. He, S. Zhang, Y. Huang, Y. Zhong, Z. Li, and Y. Han, "A novel method for improving point cloud accuracy in automotive radar object recognition," *IEEE Access*, vol. 11, pp. 78538–78548, 2023.

[169] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

[170] E. Cardillo, G. Sapienza, C. Li, and A. Caddemi, "Head motion and eyes blinking detection: A mm-wave radar for assisting people with neurodegenerative disorders," in *2020 50th European Microwave Conference (EuMC)*, pp. 925–928, IEEE, 2021.