Angle Estimation and Target Detection with Automotive Radar
Machine Learning and Compressive Sensing Approaches

Roldan Montero, I.

Publication date
2024

Citation (APA)
Roldan Montero, I. (2024). *Angle Estimation and Target Detection with Automotive Radar: Machine Learning and Compressive Sensing Approaches*. [Dissertation (TU Delft), Electrical Engineering, Mathematics and Computer Science].

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Angle Estimation and Target Detection with Automotive Radar

Machine Learning and Compressive Sensing Approaches

ANGLE ESTIMATION AND TARGET DETECTION WITH AUTOMOTIVE RADAR

MACHINE LEARNING AND COMPRESSIVE SENSING APPROACHES

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Monday 18 November 2024 at 17:30

by

Ignacio ROLDÁN MONTERO

Master of Science in Telecommunication Engineering,
Universidad Politécnica de Madrid, Spain
born in Madrid, Spain

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. A. Yarovoy	Delft University of Technology, promotor
Dr. F. Fioranelli	Delft University of Technology, promotor
Independent members:	
Prof. dr. ir. G.J.T. Leus	Delft University of Technology
Prof. dr. D.M. Gavrila	Delft University of Technology
Prof. dr. C. Silveira Vaucher	Delft University of Technology / NXP, NL
Prof. dr. C. Clemente	University of Strathclyde, Scotland
Reserve member:	
Dr. M.A. Zuñiga Zamalloa	Delft University of Technology

This research has been carried out at the Delft University of Technology in the Microwave, Signals, and Systems (MS3) group.

Keywords: Automotive Radar, Direction of Arrival Estimation, MIMO Radar, Radar Signal Processing, Machine Learning.

Printed by: Proefschiftspecialist, 1506RZ Zaandam, The Netherlands.

Front & Back: Design by Zuzanna Trybuchowicz.

ISBN/EAN 978-94-6384-677-6 (Paperback/Softback)
ISBN/EAN 978-94-6384-678-3 (E-Book/PDF)

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Author e-mail: ignacioroldanmontero@gmail.com

To all the people I love,
whose important moments I've missed by being far away.

CONTENTS

List of Acronyms

ADC	Analog to Digital Converter
AR	Auto Regressive
BCS	Bayesian Compressive Sensing
BP	Basis Pursuit
CA-CFAR	Cell Averaging Constant False Alarm Rate
CAD	Computer Aided Design
CFAR	Constant False Alarm Rate
CD	Chamfer Distance
CNN	Convolutional Neural Network
CPI	Coherent Processing Interval
CRB	Cramér-Rao Bound
CS	Compressive Sensing
DaS	Delay and Sum
DBF	Digital Beamforming
DoA	Direction of Arrival
EM	Expectation Maximization
ESPRIT	Estimation of Signal Parameters via Rotational Invariant Techniques
FFT	Fast Fourier Transform
FMCW	Frequency Modulated Continuous Wave
FoV	Field of View
FPN	Feature Pyramid Network
GGAN	Guided Generative Adversarial Network
GPS	Global Positioning System
IMU	Inertial measurement unit
IoU	Intersection over Union
LSTM	Long Short-Term Memory
MAP	Maximum A Posteriori
MLW	Main Lobe Width
MIMO	Multiple Input Multiple Output
MRA	Minimum Redundancy Array
MT-BCS	Multi Task Bayesian Compressive Sensing
MUSIC	Multiple Signal Classification
MVDR	Minimum Variance Distortionless Response
MSE	Mean Squared Error
NN	Neural Network
NUDFT	Non Uniform Discrete Fourier Transform
OS-CFAR	Ordered Statistics Constant False Alarm Rate
PC	Point Cloud
PDF	Probability Density Function

P-ISLR	Pseudo Integrated Side-lobe Level Ratio
PRI	Pulse Repetition Interval
PSF	Point Spread Function
RA	Range Azimuth
RAD	Range Azimuth Doppler
RAE	Range Azimuth Elevation
RAED	Range Azimuth Elevation Doppler
RCS	Radar Cross Section
RDC	Range Doppler Channel
RIP	Restricted Isometry Property
ROC	Receiver Operating Characteristic
RVM	Relevance Vector Machine
SAR	Synthetic Aperture Radar
SNR	Signal to Noise Ratio
SLL	Side Lobe Level
SS-MUSIC	Single Snapshot Multiple Signal Classification
TDMA	Time Division Multiple Access
TV	Total Variation
TV-CS	Total Variation Compressive Sensing
TVAL3	TV minimization by Augmented Lagrangian and Alternating direction Algorithm
ULA	Uniform Linear Array
URA	Uniform Rectangular Array

LIST OF SYMBOLS AND NOTATIONS

a, A	Scalars and complex number are denoted by normal letters
$\mathbf{a}$	Vectors are denoted by lower-case bold-face letters
$\mathbf{A}$	Matrices are denoted by upper-case bold-face letters
$\mathbf{A}^{-1}$	Inverse of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{T}}$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{H}}$	Conjugate transpose of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	Returns the matrix with the absolute value of all entries of matrix $\mathbf{A}$
vec($\mathbf{A}$)	Vectorization of matrix $\mathbf{A}$ by stacking the columns together
diag($\mathbf{A}$)	Returns the diagonal entries of matrix $\mathbf{A}$
$\lvert\lvert\mathbf{a}\rvert\rvert_{\mathrm{p}}$	P norm of $\mathbf{a}$
a^*	Complex conjugate of a
$\lfloor \mathrm{a} \rfloor$	Floor operator returning the nearest integer less than or equal to a
$\circ$	Outer product
$\otimes$	Kroneker product
$\odot$	Hadamard product
$\lvert\langle , \rangle\rvert$	Inner product
$\mathbf{I}$	Identity matrix
$\mathbb{C}$	Complex number field
$\mathbb{R}$	Real number field
$\Re(\cdot)$	Real part extraction
$\Im(\cdot)$	Imaginary part extraction

SUMMARY

This thesis focuses on advancing radar technology to meet the growing demands of autonomous driving, particularly regarding angular resolution and target detection. The work acknowledges the critical role of automotive radar in achieving higher levels of vehicle autonomy, where reliable detection, classification, and tracking of objects like pedestrians, vehicles, and infrastructure are critical. The primary focus of the dissertation is to explore novel techniques for enhancing radar system performance through machine learning (ML) and compressive sensing (CS) approaches, addressing the challenges faced by current radar technologies, such as low angular resolution and inefficient target detection in dynamic environments.

The dissertation begins by outlining the challenges in automotive radar systems, especially the need for improved angular resolution without increasing the physical size and complexity of the radar devices. The importance of angular resolution is emphasized for both azimuth and elevation, as modern vehicles must be able to discriminate between various objects, such as distinguishing between two vehicles at a similar distance, or identifying an object's height to determine whether it can be driven under or must be avoided. Current methods for improving angular resolution, such as increasing the number of transmitters and receivers in multiple input multiple output (MIMO) radars, are costly and increase system complexity, thus requiring novel solutions to meet industry needs. Then, Chapter 2 briefly summarizes the theoretical background of MIMO radars and defines the terminology used in the rest of the dissertation. This is crucial since automotive radar is a multi-disciplinary topic with people from different backgrounds interacting, and often, the same concepts are named differently.

The first research chapter, Chapter 3, introduces a self-supervised learning framework designed to enhance the angular resolution of radar systems without the need for additional physical hardware. A neural network (NN) artificially expands the radar's aperture by predicting the response of additional antenna elements based on data from radars with larger arrays. This approach leverages the correlation between antenna elements to generate a more detailed angular profile from a smaller, low-resolution radar, allowing for a more accurate estimation of incoming signals' direction of arrival (DoA). Extensive simulations and experimental results demonstrate that this method significantly enhances radar performance in separating closely spaced objects, which is critical in automotive scenarios.

Another key contribution of the dissertation is presented in Chapter 4, with the application of Bayesian compressive sensing (BCS) to automotive radar, which exploits the sparse nature of the data in the DoA domain. The BCS approach uses probabilistic models to estimate the DoA while also providing uncertainty measures, offering both accuracy and reliability in angular estimation. The research further explores how array topologies can be optimized for BCS-based DoA estimation, demonstrating that a carefully designed antenna array can achieve better performance with fewer elements, thus

reducing system costs. Additionally, the work presents a computationally efficient BCS algorithm that dramatically reduces the time needed for DoA estimation without compromising accuracy. This is a crucial advancement for real-time applications, where fast processing is required for decision-making in autonomous driving.

In addition to BCS, in Chapter 5, total variation compressive sensing (TVCS) is applied to the problem of radar imaging. TVCS enforces sparsity in the gradient of the signal rather than in the signal itself, which proves particularly effective in estimating the shape of extended targets, such as vehicles or pedestrians. By applying TVCS to 2D and 3D radar data, the dissertation demonstrates that this method can reconstruct objects' shapes more accurately than traditional methods, thereby enhancing the radar's ability to classify and understand the surrounding environment. The application of TVCS marks a significant step forward in radar-based shape estimation, especially for imaging radars used in automotive systems.

The dissertation also addresses the limitations of conventional target detection methods, particularly the widely used window-based constant false alarm rate (CFAR) detectors. Window-based CFAR detectors struggle with dynamic and unpredictable environments, which are common in road traffic scenarios. Moreover, they are unsuitable for extended targets with very different sizes, such as the ones encountered in automotive radar. To overcome this, in Chapter 6, a deep learning-based detector is proposed, trained using a newly developed dataset, RaDelft, which includes synchronized radar and lidar data. This deep learning detector outperforms traditional CFAR detectors by significantly improving the probability of detection and the Chamfer distance, especially in complex and cluttered environments. The RaDelft dataset itself is another important contribution of the dissertation, providing the research community with a well-curated, large-scale, multi-sensor dataset for further exploration and development of radar-based detection and classification systems.

In conclusion, this dissertation presents a comprehensive study of methods to enhance the angular resolution, detection capabilities, and efficiency of automotive radar systems through a combination of machine learning and compressive sensing. It provides practical solutions verified with experimental data to overcome existing limitations in automotive radar technology, particularly in the areas of angular resolution, target detection, and data processing speed. These advancements contribute to the broader goal of achieving fully autonomous driving by improving the ability of radar systems to perceive and interpret complex environments.

SAMENVATTING

Deze scriptie richt zich op de verbetering van radartechnologie om te voldoen aan de toenemende eisen van autonoom rijden, met name wat betreft hoekresolutie en doeldetectie. Het werk erkent de cruciale rol van automotive radar bij het bereiken van hogere niveaus van voertuigautonomie, waarbij het betrouwbaar detecteren, classificeren en volgen van objecten zoals voetgangers, voertuigen en infrastructuur essentieel zijn. De primaire focus van het proefschrift is het onderzoeken van nieuwe technieken om de prestaties van radarsystemen te verbeteren door middel van machine learning (ML) en compressieve sensing (CS), waarbij de uitdagingen van de huidige radartechnologieën worden aangepakt, zoals lage hoekresolutie en inefficiënte doeldetectie in dynamische omgevingen.

Het proefschrift begint met een overzicht van de uitdagingen in automotive radarsystemen, met name de behoefte aan verbeterde hoekresolutie zonder de fysieke grootte en complexiteit van de radarsystemen te vergroten. Het belang van hoekresolutie wordt benadrukt voor zowel azimut als elevatie, aangezien moderne voertuigen verschillende objecten moeten kunnen onderscheiden, zoals het onderscheiden van twee voertuigen op vergelijkbare afstand of het vaststellen van de hoogte van een object om te bepalen of er onderdoor gereden kan worden of dat het moet worden vermeden. Huidige methoden om de hoekresolutie te verbeteren, zoals het verhogen van het aantal zenders en ontvangers in multiple input multiple output (MIMO) radars, zijn kostbaar en vergroten de systeemcomplexiteit, waardoor nieuwe oplossingen nodig zijn om aan de industriële eisen te voldoen. Vervolgens vat Hoofdstuk 2 kort de theoretische achtergrond van MIMO-radars samen en definieert het de terminologie die in de rest van het proefschrift wordt gebruikt. Dit is van cruciaal belang, aangezien automotive radar een multidisciplinair onderwerp is waarin mensen uit verschillende achtergronden samenwerken en vaak dezelfde concepten anders worden genoemd.

Het eerste onderzoekshoofdstuk, Hoofdstuk 3, introduceert een zelflerend raamwerk dat is ontworpen om de hoekresolutie van radarsystemen te verbeteren zonder extra fysieke hardware. Een neuraal netwerk (NN) vergroot kunstmatig de radarapertuur door de reactie van extra antenne-elementen te voorspellen op basis van gegevens van radars met grotere arrays. Deze benadering maakt gebruik van de correlatie tussen antenne-elementen om een gedetailleerder hoekprofiel te genereren vanuit een kleinere radar met lage resolutie, waardoor een nauwkeurigere schatting mogelijk is van de richting van aankomst (DoA) van echosignalen. Uitgebreide simulaties en experimentele resultaten tonen aan dat deze methode de prestaties van radar aanzienlijk verbetert bij het onderscheiden van dicht bij elkaar gelegen objecten, wat van cruciaal belang is in verkeerstoepassingen.

Een andere belangrijke bijdrage van het proefschrift wordt gepresenteerd in Hoofdstuk 4, met de toepassing van Bayesian compressive Sensing (BCS) op automotive radar, dat gebruikmaakt van de ijle aard van de gegevens in het DoA-domein. De BCS-

benadering gebruikt probabilistische modellen om de DoA te schatten en biedt tegelijkertijd mates van onzekerheid, wat zowel nauwkeurigheid als betrouwbaarheid biedt in hoekschattingen. Het onderzoek verkent verder hoe array-topologieën geoptimaliseerd kunnen worden voor op BCS gebaseerde DoA-schatting, waarbij wordt aangetoond dat een zorgvuldig ontworpen antenne-array betere prestaties kan leveren met minder elementen, wat de systeemkosten verlaagt. Bovendien wordt een efficiënt BCS-algoritme gepresenteerd dat de tijd die nodig is voor DoA-schatting drastisch vermindert zonder nauwkeurigheid te verliezen. Dit is een cruciale vooruitgang voor realtime toepassingen, waar snelle verwerking vereist is voor besluitvorming bij autonoom rijden.

Naast BCS wordt in Hoofdstuk 5 total variation compressive sensing (TVCS) toegepast op het probleem van radarbeeldvorming. TVCS dwingt ijlheid af in de gradiënt van het signaal in plaats van in het signaal zelf, wat bijzonder effectief blijkt te zijn bij het schatten van de vorm van niet-puntachtige doelen, zoals voertuigen of voetgangers. Door TVCS toe te passen op 2D- en 3D-radargegevens, toont het proefschrift aan dat deze methode de vormen van objecten nauwkeuriger kan reconstrueren dan traditionele methoden, waardoor de radar beter in staat is om de omgeving te classificeren en te begrijpen. De toepassing van TVCS markeert een significante vooruitgang in radargebaseerde vormschatting, vooral voor beeldvormende radars die worden gebruikt in automotive systemen.

Het proefschrift behandelt ook de beperkingen van conventionele doeldetectiemethoden, met name de veelgebruikte venster-gebaseerde constant false alarm rate (CFAR)-detectoren. Venster-gebaseerde CFAR-detectoren hebben moeite met dynamische en onvoorspelbare omgevingen, die vaak voorkomen in verkeerssituaties. Bovendien zijn ze ongeschikt voor niet-puntachtige doelen met zeer verschillende afmetingen, zoals die worden aangetroffen in automotive radar. Om dit te verhelpen wordt in Hoofdstuk 6 een op deep learning gebaseerde detector voorgesteld, getraind met behulp van een nieuw ontwikkelde dataset, RaDelft, die gesynchroniseerde radar- en lidar-gegevens bevat. Deze deep learning-detector presteert aanzienlijk beter dan traditionele CFAR-detectoren door de detectiekans en de Chamfer-afstand aanzienlijk te verbeteren, vooral in complexe omgevingen met ongewenste echo's. De RaDelft-dataset zelf is een andere belangrijke bijdrage van het proefschrift, aangezien het de onderzoekswereld voorziet van een goed gecureerde, grootschalige, multisensordataset voor verdere verkenning en ontwikkeling van op radar gebaseerde detectie- en classificatiesystemen.

Concluderend presenteert dit proefschrift een uitgebreide studie van methoden om de hoekresolutie, detectiemogelijkheden en efficiëntie van automotive radarsystemen te verbeteren door een combinatie van machine learning en compressieve sensing. Het biedt praktische oplossingen, geverifieerd met experimentele gegevens, om bestaande beperkingen in de automotive radartechnologie te overwinnen, met name op het gebied van hoekresolutie, doeldetectie en datasnelheid. Deze vooruitgangen dragen bij aan het bredere doel van volledig autonoom rijden door de mogelijkheid van radarsystemen te verbeteren om complexe omgevingen waar te nemen en te interpreteren.

1

INTRODUCTION

I believe automotive radar will be a critical component of future vehicles, but always in cooperation with other sensors for redundancy. In this chapter, I mention some of the outstanding challenges that automotive radar must solve before it can be safely used. In this thesis, I propose some new approaches that may help us get one step closer to solving these problems.

1

1.1. AUTOMOTIVE RADAR: MOTIVATION AND CHALLENGES

Autonomous driving is one of the most relevant trends in the automotive industry, and a race for reaching driver-assistance level 5 has begun between all the major car manufacturers. To achieve this, the sensing suite in autonomous vehicles must provide the most reliable and dense information about the surroundings. Therefore, reliable detection and classification of very different objects, such as pedestrians, cars, potholes, or speed bumps, amongst others, should be performed in real-time. Moreover, the system must understand which items can be driven-over, such as small debris on the road or speed bumps, which objects can be driven under, such as bridges or tunnel entrances, and which objects should be avoided as significant obstacles on the road that could damage the vehicle.

This difficult task cannot be run successfully by any single sensor, and a combination of radars, cameras, and lidar is the most used approach in current autonomous driving research. With higher levels of automation, more sensors are expected to be integrated into a vehicle, as seen in Figure 1.1. However, radars have some advantages over other sensors. First, they work in adverse weather conditions, such as fog, rain, or low-light conditions [1], with little drop in performance. Second, radar can accurately and directly measure objects' velocity via the Doppler effect. Finally, it can be mounted behind the chassis and, therefore, is less susceptible to damage from scratches, dust, and other environmental factors, making it more reliable and long-lasting in various driving conditions. All these reasons make radar a crucial sensor for vehicular autonomy [2].

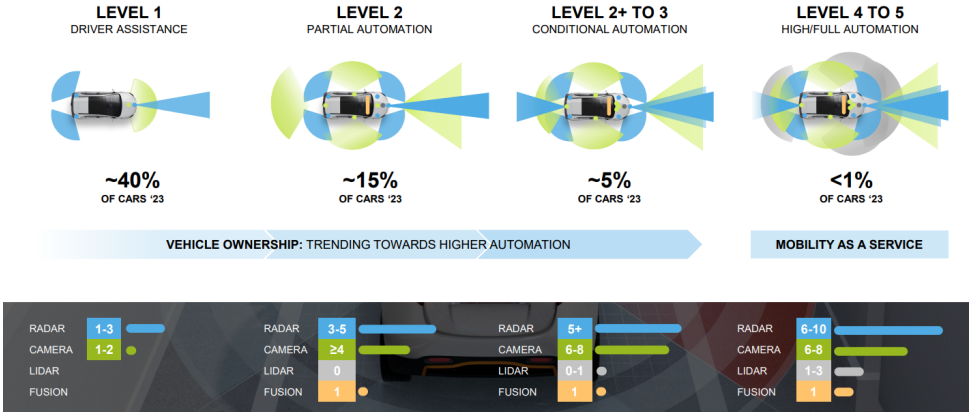


Figure 1.1: Number of sensors for the different levels of vehicle automation. It can be seen how fast the number of radars increases. Taken from [3].

A notable trend in automotive radar is the shift towards imaging radar, which achieves a high angular resolution in both azimuth and elevation angles by leveraging a more significant number of antennas and, thus, a larger aperture, combined with using multiple-input multiple-output (MIMO) methods [4]. However, many open challenges still need to be solved in imaging radars to make them suitable for safe, fast, and reliable environmental perception. These will be discussed in the following sub-section.

1.1.1. CHALLENGES

ANGULAR RESOLUTION

Automotive radars must have enough angle resolution in both azimuth and elevation to provide dense and valuable information about the surroundings for subsequent detection and classification. With the development of MIMO radars, the angular resolution has been improved significantly, but it still needs to meet the requirements of the automotive industry. Current high resolution in range and Doppler may seem enough to separate objects, but many everyday automotive situations will not be resolved. For example, Figure 1.2 shows a vehicle equipped with an automotive radar on a highway, where two cars are at the same distance and have the same speed. For the automotive radar to resolve two targets in this scene, around 1-degree azimuth resolution is needed. Moreover, Figure 1.3 shows another situation where a vehicle is stopped at the entrance of a tunnel. Similarly, around 1-degree elevation resolution is required to distinguish between the car and tunnel entrance.

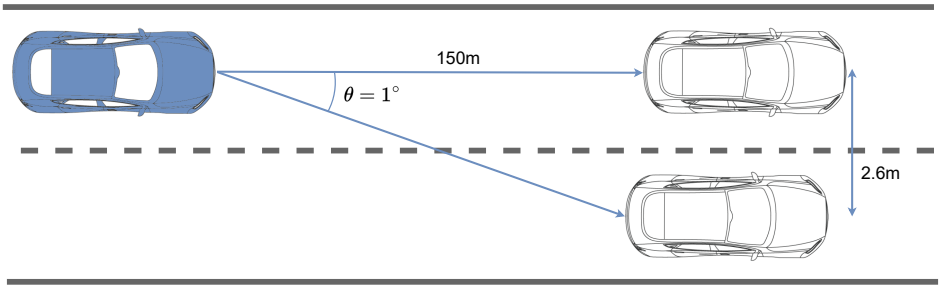


Figure 1.2: Representation of two vehicles in two different lanes moving at the same speed and the same range with respect to the ego vehicle. High azimuth resolution would be needed to resolve both targets.

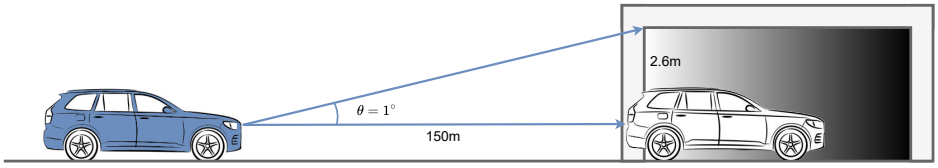
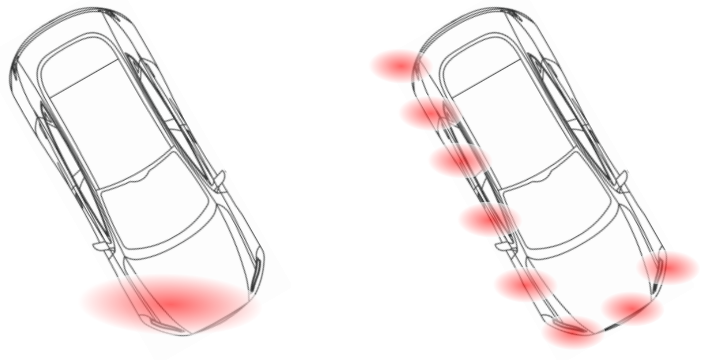


Figure 1.3: Representation of a vehicle stopped at a tunnel entrance. A high elevation resolution would be needed to detect it as an independent target.

Also, high angular resolution is needed for object classification [5]. Automotive targets have many scattering centers due to their different curvatures, corners, and slots [6–8]. If a low angular resolution radar is used, the power reflected from all these scatterer points will combine in a single cell, losing all the target's shape information, as seen in Figure 1.4a. On the other hand, a high angular resolution radar will generate more detection points per object, being able to perceive shape information crucial for target classification, as illustrated in Figure 1.4b.

The approach to address this problem is designing automotive radars with more transmitters and receivers. While this approach directly improves the angular resolu-



(a) (b)

Figure 1.4: On (a), a single detection is perceived by a low-resolution radar, while in (b), many smaller detections are generated by a high-resolution radar. As can be seen, the shape of the target is preserved in case (b), and therefore, the classification of the object will be better.

tion, it creates other practical issues, such as an increase in the system cost, a massive increase in the data rate and the processing capabilities to handle that data, thermal issues, and complicated multiplexing schemes, among others. In this thesis, two different approaches to increase the angular resolution without adding extra transmitters or receivers have been proposed, one based on neural networks in Chapter 3, and one based on compressive sensing in Chapter 4.

RADAR DETECTION

A primary challenge in the context of target detection with automotive radar is the use of the well-known window-based constant false alarm rate (CFAR) detectors for generating radar point clouds from the dense radar data cube. While CFAR detectors have proven optimal in other environments [9], their application in the dynamic and unpredictable conditions of road traffic scenarios suffers from poor performance [10, 11]. Namely, they are designed to maintain a constant rate of false alarms amidst varying clutter, but they struggle to adapt to the rapidly changing environments typical of roadways. Complications such as non-uniform clutter (or the lack of reliable clutter models for this task), target masking, and shadowing can significantly reduce the effectiveness of CFAR detectors in automotive radar settings. Additionally, CFAR detectors are constrained by a fundamental limitation: they typically assume a fixed, expected target size based on predefined guard and training cell hyperparameters. However, in an automotive context, this assumption needs to be revised as the size of potential targets can widely vary, ranging from medium-sized objects, such as pedestrians, to large vehicles, such as trucks or buses. Moreover, the perceived size of these targets in the radar's angular dimension changes with distance. Large objects occupying multiple cells at close range can appear as more straightforward point-like targets at further distances. This relationship between angular target size and distance adds another layer of complexity to using

window-based CFAR detectors in automotive radar, necessitating alternative solutions to accurately detect and then classify objects under varying road conditions. This thesis introduces a novel deep learning radar detector in Chapter 6 to tackle this issue specifically.

SCENE UNDERSTANDING

In terms of basic functionalities, radar sensors can be used just to detect objects and report their position and velocity. However, more is needed to achieve autonomous driving. Target classification is necessary for modern radars since decision systems require information about the objects' characteristics, such as whether they are drivable-through, drivable-under, a vulnerable road user, or part of the road infrastructure. This has become only possible in the last few years with the increased popularity of machine learning methods and the improvements in resolution. However, radar data is not directly interpretable for humans. Hence, labeling radar data is not trivial, and this is one of the main bottlenecks when designing neural network based classifiers, which are inherently data-hungry. Moreover, radar data have unique characteristics, such as the measurement of Doppler or the extensive dynamic range in the measured backscattered power, that need specific classifier designs to be applied. On top of that, access to high-quality radar data is still not trivial, and it is usually the practice of collecting specific data for each developed or proposed algorithm. While there are open datasets with widely accepted benchmarks in the computer vision community, this is a relatively new trend in the radar community [12]. Also, the differences in the type of sensor and waveform, including the effects of the fabrication and mounting of the radar sensors, make it challenging to design a generic classifier that is truly device-agnostic. During this PhD work, a large-scale multi-sensor dataset has been collected and shared publicly, aiming to ease access to radar data.

REAL-TIME RADAR PROCESSING

In the automotive context, the whole radar processing pipeline, from the digitization of the signals to the report of the targets, must be performed in the order of 50ms to guarantee safety in the higher levels of automation. With the increase of receiver elements in modern automotive systems, the amount of data to process is also increasing, and this is especially problematic for super-resolution direction of arrival (DoA) algorithms that need to be implemented in embedded hardware with low computational resources. Therefore, the computational complexity of DoA algorithms should be considered, and new low-complexity methods are needed to deal with the increasing number of antennas. This thesis presents a framework in Chapter 4 that, in combination with Bayesian compressive sensing (BCS), significantly reduces the computational time for the DoA.

INTERFERENCE MITIGATION

Figure 1.1 shows that the number of radars per vehicle rapidly increases, with up to 10 onboard radars to build the whole environmental perception system. This means that the number of radars working simultaneously will quickly escalate in a crowded environment. In that dense environment, interference mitigation is a crucial component that should be addressed. While interference mitigation is outside of the scope of this PhD thesis, it is an issue that should be considered when building new radar solutions.

1

1.2. RESEARCH OBJECTIVES AND MAIN RESULTS

This thesis aims to design and evaluate novel signal and data processing algorithms for automotive imaging radars at different levels of the processing chain. The need for high angular resolution brings the following research questions:

Q1: Is it possible to increase the angular resolution of a MIMO radar system without increasing the number of physical transmitters or receivers?

Q2: Is preserving the detected objects' shape possible when performing DoA processing?

Q3: Is it possible to design a MIMO array topology optimized for the DoA algorithm to be specifically used in the system?

However, with the increase in angular resolution of the most recent radar systems, some new challenges and questions arise, and this PhD thesis has tackled them in several ways, namely:

Q4: Could efficient DoA estimation methods be designed to process the high-dimensional data sampled by modern imaging radars?

Q5: With the extended nature of targets due to the high angular resolution and the diverse and dynamic automotive scenes, is it possible to design a more suitable detector than conventional, window-based state-of-the-art radar detectors?

The aforementioned research questions lead to the following primary outcomes for this thesis:

- A novel framework leveraging neural networks was introduced to enhance the angular resolution of automotive radars. This framework artificially expands the antenna aperture through a self-supervised scheme. Specifically, data from a high angular resolution radar, characterized by a large aperture antenna, was employed to train a deep neural network to extrapolate the antenna element response of a compact, low angular resolution radar. The proposed framework has been extensively validated using both simulated and experimental data collected from a commercial automotive radar. A case study involving two pedestrians walking towards the radar at identical speed and range demonstrates that using an array of 86 virtual elements (with an aperture of 43 wavelengths) allows for resolving the two pedestrians 94.5% of the time, compared to only 30.1% with an array of 44 virtual elements (with an aperture of 22 wavelengths). When the proposed method was applied before angle estimation, the probability of resolution increased to 55.6% (i.e., the probability of detecting two targets instead of a single one). These findings indicate that the proposed method significantly enhances the radar's ability to distinguish between objects, which can directly improve the accuracy and response time of the vehicle's planning and decision-making systems.

- The angular estimation for extended targets using a single snapshot in short-range imaging radar was thoroughly analyzed. Traditional compressive sensing methods are inapplicable due to the scene's non-sparsity characteristics, and super-resolution algorithms based on subspace methods are unsuitable due to the problem's constraints (single snapshot and coherent sources). Consequently, two novel methods were proposed based on total variation compressive sensing (TVCS), where sparsity is enforced in the signal's gradient. The first method employs an independent azimuth-elevation estimation for each range cell, whereas the second method introduces a joint 3D range-azimuth-elevation optimization problem. These proposed algorithms were validated using realistically simulated data from 3D CAD models of various cars, pedestrians, and bicycles. The performance of these methods was compared to that of a conventional Fourier beamformer and the standard Compressive Sensing algorithm in shape reconstruction. The intersection over union (IoU) between the ground truth shape and the estimated shape produced by the four methods has been calculated to evaluate shape reconstruction capabilities. The proposed methods consistently outperform the FFT-based and conventional CS methods, achieving a maximum IoU increase of 3.5 times under favorable SNR conditions for the 2D version.

- Bayesian compressive sensing (BCS) is a family of algorithms that addresses the compressive sensing (CS) problem from a probabilistic perspective, providing both estimations and their associated uncertainties. Leveraging this uncertainty, a novel algorithm was developed to generate the optimal array topology for angle estimation using BCS in MIMO radars. This algorithm utilizes the differential entropy of the measurements and has been executed offline with simulated data to generate the array topology. It was subsequently tested with real measurements from a commercial automotive MIMO radar. The results indicate that the generated array outperforms the average random array, which is typically considered the best strategy for generating CS sensing matrices. Furthermore, the proposed method can accurately resolve closely spaced targets using only 50% of the elements compared to a fully dense uniform linear array (ULA).

- An algorithm was designed based on a split-and-merge approach followed by a correction stage, demonstrating a tenfold reduction in computational time compared to standard Bayesian compressive sensing (BCS). Significantly, this reduction in computational time does not compromise accuracy or the probability of resolving closely spaced targets. Experimental measurements validated that the proposed method outperforms FFT-based approaches and the single-snapshot multiple signal classification (MUSIC) algorithm in terms of accuracy and resolution probability, all while operating ten times faster than the MUSIC algorithm.

- A novel large-scale, real-world multi-sensory dataset was recorded in various driving scenarios in the city of Delft, which is publicly shared for non-commercial purposes. The *RaDelft* dataset contains over 30 minutes of actual driving scenarios collected using a vehicle equipped with lidar and radar sensors, resulting in 16975 radar frames paired with corresponding lidar ground truth. Compared with other existing datasets, *RaDelft* provides raw data from a commercial 4D imaging

radar needed by radar practitioners for many research lines. Moreover, it contains data processed at other levels (e.g., radar cubes and point clouds) suitable for researchers with different backgrounds and interests.

- Using the *RaDelft* dataset, a novel data-driven approach was proposed for radar detection, utilizing unlabeled synchronized lidar data as ground truth to train a deep neural network with radar data as the sole input. Two performance metrics were employed to validate the method: the conventional probability of detection and probability of false alarm, as well as the Chamfer distance, a point cloud-level metric that captures spatial relationships and similarities between point clouds. The proposed method achieves a 77% reduction in Chamfer distance (4.2 meters) compared to conventional OSCFAR detectors and a 28% reduction (0.62 meters) compared to the state-of-the-art. Additionally, it significantly increases the probability of detection. An ablation study demonstrated that incorporating temporal information is crucial, with Doppler information being particularly vital for the model's performance. Results indicate an increase in the probability of detection from 50.44% to 62.13%, and a 27% reduction in Chamfer distance when using Doppler information.

1.3. THESIS OUTLINE

Chapter 2 - FMCW MIMO Radar Fundamentals and Terminology.
Automotive radar has become part of a wider multidisciplinary field in autonomous vehicles where scientists from different backgrounds are cooperating. As different research communities might use different terms [13, 14], a list of definitions used in this thesis is provided in this chapter. Moreover, the standard MIMO FMCW signal model is reviewed, highlighting the relevant steps for the following chapters.

Chapter 3 - Angular Resolution Enhancement Using Self-Supervised Learning.
This chapter shows the first main result listed in Section 1.2, where a machine learning framework is used to enhance the angular resolution of automotive radar systems. It answers the first research question **Q1**, since the angular resolution is increased without increasing the number of physical antennas. The results presented in this chapter led to the following publications:
I. Roldan, F. Fioranelli and A. Yarovoy, "Enhancing Angular Resolution Using Neural Networks in Automotive Radars," 2021 18th European Radar Conference (EuRAD), London, United Kingdom, 2022, pp. 58-61.
I. Roldan, F. Fioranelli and A. Yarovoy, "Self-Supervised Learning for Enhancing Angular Resolution in Automotive MIMO Radars," in IEEE Transactions on Vehicular Technology, vol. 72, no. 9, pp. 11505-11514, Sept. 2023.

Chapter 4 - Bayesian Compressive Sensing Applied to Automotive Radar.
In this chapter, two novel ideas are presented under the umbrella of BCS. The first one addresses the placement of the antenna array elements, given that BCS will be used for DoA estimation. Also, the chapter presents a computationally efficient BCS method for DoA estimation. The research questions **Q3** and **Q4** are answered positively in this chap-

ter by designing an array topology specially tailored for BCS and by implementing BCS efficiently. The results of this chapter were presented in two conference papers:
I. Roldan, L. Lamberti, F. Fioranelli and A. Yarovoy, "Low Complexity Single-Snapshot DoA Estimation via Bayesian Compressive Sensing," 2023 IEEE Radar Conference (RadarConf23), San Antonio, TX, USA, 2023, pp. 1-6.
L. Lamberti, I. Roldan, A. Yarovoy and F. Fioranelli, "Sparse Array Placement for Bayesian Compressive Sensing Based Direction of Arrival Estimation," 2024 IEEE Radar Conference (RadarConf24), Denver, CO, USA, 2024, pp. 1-6.

Chapter 5 - Total Variation Compressive Sensing Applied to Automotive Radar.
This chapter applies for the first time total variation regularization in compressive sensing for shape estimation in imaging radars. Research question **Q2** is answered in this chapter by leveraging TVCS in the angular estimation process, which enforces high-quality shape estimations. This research led to two publications, one conference, and one journal paper:
I. Roldan, F. Fioranelli and A. Yarovoy, "Total Variation Compressive Sensing for Extended Targets in MIMO Radar," 2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM), Trondheim, Norway, 2022, pp. 61-65.
I. Roldan, F. Fioranelli and A. Yarovoy, "Total Variation Compressive Sensing for 3D Shape Estimation in Short-Range Imaging Radars," in IEEE Transactions on Radar Systems, vol. 1, pp. 583-592, 2023.

Chapter 6 - Deep Automotive Radar Detector and *RaDelft* Dataset.
This chapter introduces the novel collected *RaDelft* dataset, showing how conventional window-based CFAR detectors fail to perform well in real automotive scenes. Also, a novel radar detector based on deep learning is presented. Since the proposed detector outperforms window-based CFAR detectors, research question **Q5** is positively answered in this chapter. This research resulted in two publications where the detector and the collected dataset are presented and made available for researchers:
I. Roldan, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli and A. Yarovoy, "See Further Than CFAR: a Data-Driven Radar Detector Trained by Lidar," 2024 IEEE Radar Conference (RadarConf24), Denver, CO, USA, 2024, pp. 1-6.
I. Roldan, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli and A. Yarovoy, "A Deep Automotive Radar Detector using the RaDelft Dataset," in IEEE Transactions on Radar Systems.

2

FMCW MIMO Radar Fundamentals and Terminology

While completing this PhD thesis, I worked at different levels of the radar processing chain and read papers from researchers with very different backgrounds. Sometimes, various concepts are called by the same name, and sometimes, the same concept is called differently. To give an example that often triggered some misunderstandings: in the computer vision community, 'detection' means assigning a label/class to a region of pixels such as 'human' or 'car.' In contrast, within the radar community, 'detection' is a binary decision to determine whether a target is present or not. Some terms are ambiguous and context-dependent even within the radar community, and even within the same radar research group.

In this short chapter, I have briefly reviewed the relevant radar concepts and provided a list of definitions I will use throughout the thesis. If, at some point in the reading of other chapters, you find some term and are unsure what I am referring to, hopefully, you can find the answer in the following pages.

2.1. SIGNAL MODEL

2.1.1. FMCW

Current standard automotive radars are frequency-modulated continuous-wave (FMCW), which modulate the transmitted frequency with a sawtooth pattern. The main advantages of FMCW over other waveforms include simultaneous range and Doppler estimation, low sampling frequency, narrow-band processing, and simplicity, among others. An example of FMCW signal in the time-frequency domain can be seen in Figure 2.1, where the chirp duration is denoted with T_0. As it can be seen, there is some idle time after the frequency reaches the maximum value to allow the hardware components to restart the correct initial frequency; this is representative of what happens in mm-wave automotive radar chips, in contrast to more ideal models of FMCW radars with instantaneous frequency transition.

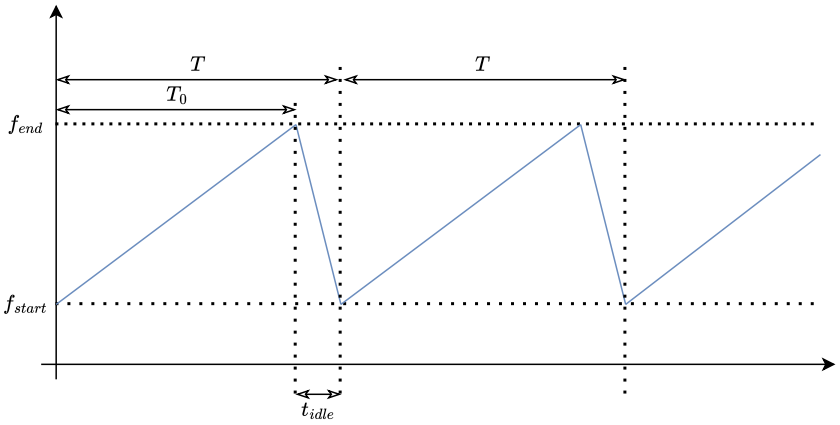


Figure 2.1: FMCW chirp model. T_0 is the chirp time, i.e., the useful part of the signal that can be sampled, t_{idle} is the needed time for the hardware to reset to the initial frequency, and T is the chirp period.

During one coherent processing interval (CPI), a sequence of FMCW chirps is transmitted with a pulse repetition interval (PRI) equal to T. A transmitted chirp can be modeled as:

$$s_0(t) = \begin{cases} e^{2\pi j(f_0 t + 0.5\mu t^2)} & t \in [0, T_0], \\ 0 & t \in [T_0, T], \end{cases} \tag{2.1}$$

where f_0 is the start frequency, $\mu = \frac{B}{T_0}$ is the frequency modulation rate (or chirp slope), and B is the transmitted bandwidth. A sequence of transmitted chirps can be decomposed in the fast-time domain $t' \in [0, T_0]$ and the slow-time domain $l = 0, 1, 2, ..., L-1$, where $l = \left\lfloor \frac{t}{T} \right\rfloor$, $t' = t - lT$, and L is the total number of chirps in a CPI (being $\lfloor . \rfloor$ the floor function). Then, the periodic transmitted signal can be modeled as:

$$s(t) = s(t' + lT) = s(l, t'). \tag{2.2}$$

The received signal, backscattered by K targets with unknown range and radial velocity, will be a superposition of time delayed versions of the transmitted signal. Considering the $k-$th target with R_k range and v_k radial velocity, the round trip delay is:

$$\tau_k(l,t') = \frac{2R_k}{c} + \frac{2v_k}{c}(t' + lT), \tag{2.3}$$

where c is the speed of light and $\gamma_k = \frac{2R_k}{c} \ll T_0$. The received signal is correlated with the conjugate of the transmitted signal in the analog implementation of the matched filter (this process is also called 'de-chirping') to generate the beat frequency signal. This signal for the $k-$th target can be defined as:

$$y_k(l,t') = \alpha_k e^{-2\pi j(f_0 \frac{2v_k}{c} Tl + \mu\gamma_k t')} e^{-2\pi j\mu \frac{2v_k}{c} Tlt'}, \tag{2.4}$$

where α_k is the complex amplitude of the received signal related to the target properties. This model assumes that the Doppler frequency is negligible compared to the beat frequency, which is typically the case in automotive radar. Then, the signal is sampled with the analog-to-digital converter (ADC) at a sampling frequency f_s. The discretized version of y_k, omitting the range migration coupling term, can be expressed as:

$$\mathrm{y}_k(l,n) = \alpha_k e^{2\pi j(f_D(v_k)l)} e^{2\pi j(f_r(r_k)n)}, \tag{2.5}$$

where $n = \lfloor t' f_s \rfloor \in [0,1,2,...,N]$ is the fast-time index, N is the number of fast-time samples in a chirp, $f_r(r_k) = -\frac{\mu}{f_s}\gamma_k$ and $f_D(v_k) = -f_0\frac{2v_k}{c}$. Directly in equation (2.5), it is possible to compute the ambiguities of the system according to the Nyquist criterion, i.e., $f_r(r_k) \in [-0.5, 0, 5)$ and $f_D(v_k) \in [-0.5, 0, 5)$, resulting in:

$$R_{max} = \frac{cN}{2B} = N\Delta r, \qquad v_{max} = \frac{c}{4Tf_0}, \tag{2.6}$$

being $\Delta r = \frac{c}{2B}$ the theoretical range resolution. These ambiguities are the maximum range and velocity the system can measure without frequency folding, meaning larger values beyond these limits will be estimated incorrectly.

However, the described signal model only accounts for one receiving antenna. In the following subsection, the concept of MIMO is explained, and the signal model is expanded to take this into account.

2.1.2. MIMO

As shown in Figure 2.2, the basic principle of angle estimation in an antenna array relies upon the extra distance a signal must travel to reach different elements. For a uniform linear array (ULA) with $m \in [0,1,2,...,M-1]$ number of elements, it is possible to expand (2.3) to take into account this new additional delay given the target angle θ_k as:

$$\tau_k(l,t',m) = \frac{2R_k}{c} + \frac{2v_k}{c}(t' + lT) + \frac{f_0 dm}{c}\sin\theta_k. \tag{2.7}$$

This extra time delay is translated to a phase shift in the de-chirped signal, and after digitization, the received signal can be expressed as:

$$y_k(l,n,m) = \alpha_k e^{2\pi j(f_D(v_k)l)} e^{2\pi j(f_r(r_k)n)} e^{2\pi j(\frac{md}{\lambda}\sin\theta_k)}, \tag{2.8}$$

where d is the distance between the antenna elements. The wide-band coupling term, $e^{2\pi j(\mu\frac{d}{cf_s})nm}$, can be neglected in automotive radar since the fractional bandwidth is small. Looking at the newly added term in (2.8) compared to the single receiver expression, it can be seen that the angular resolution (i.e., the distance between the first two nulls of the beam pattern) is inversely proportional to the product Md, which is also known as the antenna aperture. The angular resolution can be expressed as $\Delta\theta \approx \frac{\lambda}{Md\cos\theta}$.

As it can be seen, the straightforward way to improve the angular resolution is to increase the number of elements in the array. However, the angular resolution rises linearly with the number of new physical antennas, and the system's cost and complexity limit this approach. Therefore, another approach must be used to overcome this limitation, which is where MIMO technology comes into play.

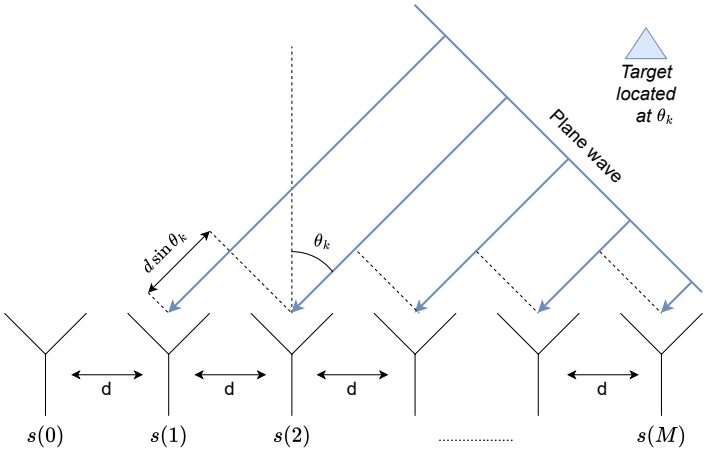


Figure 2.2: A wave scattered from a target located at θ_k, arriving at an antenna array. The time delay between each antenna element depends on the target angle θ_k.

MIMO radars use a transmit array with widely-spaced elements to add spatial diversity to the received signals. Figure 2.3a shows an example of a MIMO system with $M_{Tx} = 2$ transmit elements and $M_{Rx} = 4$ receive elements. As it can be seen, assuming the target is in the far field, the phase shifts induced by the difference in position of the two Tx elements are equivalent to having one transmitter and $M_{Tx}M_{Rx}$ receiver elements, as shown in Figure 2.3b. Therefore, the antenna aperture is being virtually increased, with its corresponding gain in the angular resolution.

In a general formulation for ULA topologies, the baseband signal model in (2.8) can be modified to take into account the phase shifts due to the transmit and receive arrays as:

$$y_k(l,n,m_{Tx},m_{Rx}) = \alpha_k e^{2\pi j(f_D(v_k)l)} e^{2\pi j(f_r(r_k)n)} e^{2\pi j(\frac{m_{Tx}d_{Tx}+m_{Rx}d_{Rx}}{\lambda}\sin\theta_k)}, \tag{2.9}$$

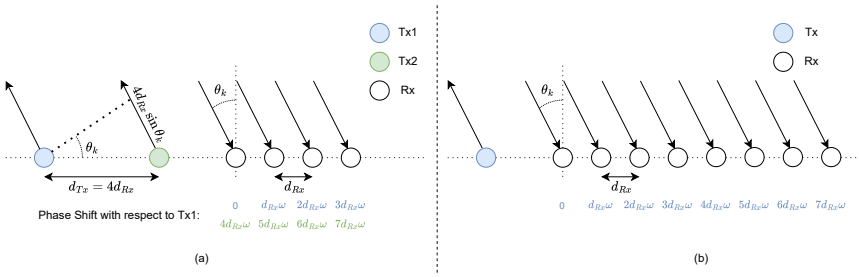


Figure 2.3: On the left in (a), a MIMO system with two transmitters and four receivers. On the right in (b), the equivalent virtual ULA with eight elements.

where $m_{Tx} \in [0,1,2,...,M_{Tx}-1]$ is the transmit element number, $m_{Rx} \in [0,1,2,...,M_{Rx}-1]$ is the receive element number, d_{Tx} is the distance between transmit elements and d_{Rx} is the distance between receive elements. In this case, both arrays are considered to be ULA, but generalization to non-uniform arrays can be easily made by modifying the $m_i d$ products to the positions of the elements. The improvement in angular resolution can also be seen from the point of view of the array pattern. In a MIMO system, the two-way array pattern will be the product of the transmit and receive array patterns. Using as an example the same topology shown in Figure 2.3a with $d = \lambda/2$, the generated array patterns can be seen in Figure 2.4.

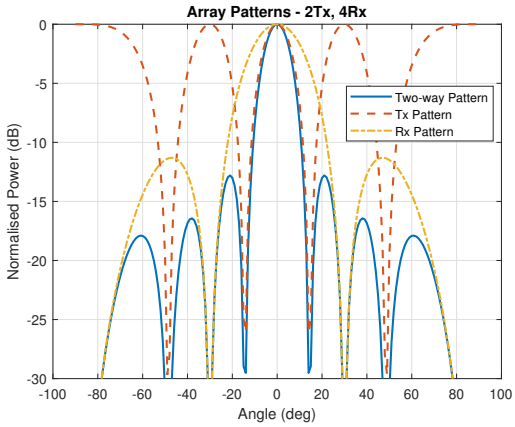


Figure 2.4: In red, the Tx array pattern, with grating lobes due to the wide space between the two elements. In yellow, the 4-elements ULA Rx array pattern. In blue, the combination of both patterns with MIMO approach, equivalent to the one of an 8-elements ULA.

As it can be seen, the array pattern of the transmit array presents grating lobes since the spacing between elements is one wavelength. On the other hand, the array pattern of the receiver does not have grating lobes since the elements are separated by half-wavelength. However, with the MIMO approach, it is possible to get an array pattern equivalent to an 8-element (virtual) ULA when combining both patterns. Therefore,

writing the equations for the virtual array is more convenient than keeping the distinction between transmit and receive elements. The baseband signal model can be simplified from (2.9) back to (2.8) just using m as the virtual element number and d as the distance between virtual elements. Each receiver will receive each transmitted signal, and thus, the resulting number of virtual elements is $M = M_{Tx}M_{Rx}$. Therefore, adding a new receive element will increase the angular resolution by a factor of M_{Tx} (and vice-versa) unless the elements are placed so that the virtual elements overlap.

The signal model in (2.8) can now be organized in matrix form, which is convenient for many of the following chapters. First, the DoA steering vector $\mathbf{a}(\theta_k) \in \mathbb{C}^{M\times 1}$, the fast-time sinusoidal vector $\mathbf{f}_r(r_k) \in \mathbb{C}^{N\times 1}$, and the slow-time sinusoidal vector $\mathbf{f}_D(v_k) \in \mathbb{C}^{L\times 1}$ are defined as:

$$\mathbf{a}(\theta_k) = [1, e^{-2\pi j(\frac{d}{\lambda}\sin\theta_k)}, ..., e^{-2\pi j((M-1)\frac{d}{\lambda}\sin\theta_k)}]^T, \tag{2.10}$$

$$\mathbf{f}_r(r_k) = [1, e^{-2\pi j(\frac{\mu}{f_s}\gamma_k)}, ..., e^{-2\pi j((N-1)\frac{\mu}{f_s}\gamma_k)}]^T, \tag{2.11}$$

$$\mathbf{f}_D(v_k) = [1, e^{-2\pi j(\frac{2v_k}{c}f_0)}, ..., e^{-2\pi j((L-1)\frac{2v_k}{c}f_0)}]^T. \tag{2.12}$$

Then, the three-dimensional baseband received signal $\mathbf{Y} \in \mathbb{C}^{N\times L\times M}$ in the presence of K point targets in matrix form is expressed as:

$$\mathbf{Y} = \sum_{k=1}^{K} \alpha_k \mathbf{a}(\theta_k) \circ \mathbf{f}_r(r_k) \circ \mathbf{f}_D(v_k) + \mathcal{N} \tag{2.13}$$

where, $\circ$ is the outer product and $\mathcal{N} \in \mathbb{C}^{N\times L\times M}$ is a complex Gaussian noise tensor. As it can be seen, with the few assumptions made to build this signal model, each of the variables to be estimated (θ_k, r_k, v_k) are independent. Finally, some of the chapters of this thesis will use a 2D antenna array; therefore, the signal model must be expanded to consider the targets' elevation angle as well. This can be easily done by modifying the steering vectors to take into account the new array topology, as:

$$\mathbf{a}(\theta_k, \phi_k) = \begin{pmatrix} 1 \\ e^{2\pi j\frac{d}{\lambda}(\cos\theta_k \sin\phi_k + \cos\phi_k)} \\ \vdots \\ e^{2\pi j\frac{d}{\lambda}((M_x-1)\cos\theta_k \sin\phi_k + (M_z-1)\cos\phi_k)} \end{pmatrix}, \tag{2.14}$$

where ϕ_k is the elevation angle of the $k-$th target, and M_x and M_z are the number of antenna elements of the 2D uniform rectangular array (URA).

In a MIMO system, the transmitted signals must be orthogonal to separate them on the receiver side. While many MIMO waveforms have different trade-offs [15], this thesis uses time division multiple access (TDMA) for its simple implementation. The detrimental effects of this approach and how to compensate for them will be treated in Chapter 6.

2.2. TERMINOLOGY

In recent years, automotive radar has become part of a wider multidisciplinary field in autonomous vehicles where scientists from rather different backgrounds cooperate. As different research communities might use different terms [13, 14], a list of definitions used in this work is provided before reviewing the classical FMCW signal processing chain.

- *Raw radar data* or *ADC data* refers to the complex-valued baseband samples the ADC provides at each receiver channel.
- *Virtual channel* or *channel* refers to one of the multiple unique combinations of Tx-Rx antenna in a MIMO radar, meaning the signal transmitted from a Tx is received, down-converted and sampled at the Rx.
- *Radar frame* refers to the set of ADC samples from each virtual channel's CPI. It has dimensions of $N \times L \times M$, where these are the number of samples in fast time, number of samples in slow time, and number of virtual channels, respectively. This *radar frame* corresponds with the signal model presented in (2.13).
- *Radar cube* refers to the spherical coordinate, discretized representation of the radar data, meaning the range, azimuth, elevation, and Doppler estimation have already been performed. When the elevation is also estimated, this is a 4D hypercube, but for simplicity, the term *radar cube* will be used to refer to the 3D and 4D versions. Each cell in the *radar cube* contains a scalar value indicating the reflected power in that cell. The size of each cell is related to the characteristics of the radar, such as the transmitted bandwidth or the antenna array topology. Hence, the cells are generally not the same size over the whole grid.
- An *extended target* is a target occupying multiple cells in one or several dimensions, in contrast to a *point target*, which occupies a single cell. Point targets present a clear peak in the estimation space (range-Doppler-angle), while extended targets do not.
- *Detection* is the binary decision problem determining whether a *radar cube* cell contains only noise or noise plus target. On the other hand, *classification* aims to associate a class to each detected cell, such as 'pedestrian', 'vehicle', or 'light pole', and so on. These two tasks are generally treated as two blocks in a conventional radar processing pipeline.
- *3D occupancy grid* refers to a binary cube, also in spherical coordinates, which contains ones in voxels that are occupied by detected targets and zeros otherwise. Such a *3D occupancy grid* could be generated directly from a lidar point cloud, but also from a *radar cube* through a detector. In the latter case, the resulting *3D occupancy grid* and the *radar cube* share the same grid.
- *Point cloud* refers to a set of P points, each containing P_L features that result from selecting only those cells containing ones in a *3D occupancy grid* and (typically)

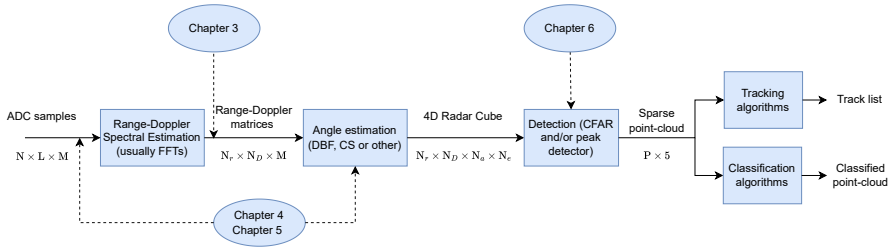


Figure 2.5: Typical processing pipeline in FMCW automotive radar, from the raw ADC samples to the output of classification & tracking steps. N and L are the number of samples in a chirp and chirps in a CPI, respectively. M are the number of virtual channels in a MIMO system given by the product of the number of Tx and Rx channels. N_r,N_D, N_a, and N_e are the number of range, Doppler, azimuth and elevation cells. Finally, P is the number of points after the detector, with the three spatial coordinates plus Doppler and power.
The ovals with dashed lines indicate at which point in the pipeline the chapters of this thesis contribute to.

converting them to Cartesian coordinates. For radar point clouds, it is usually assumed that $P_L = 5$, adding Doppler and power information to the three spatial dimensions, while for lidar point clouds, $P_L = 4$ since Doppler is not provided.

2.3. BRIEF RECAP OF FMCW SIGNAL PROCESSING

Figure 2.5 illustrates the conventional FMCW radar processing pipeline and to which step each of the chapters of this thesis is contributing. The steps are as follows:

1. Range and Doppler spectral estimation is performed from the baseband or ADC samples organized in fast-time, slow-time, and channel dimensions. Usually, this is achieved by applying a window with the fast Fourier transform (FFT) algorithm independently in fast time-and slow-time. However, this step may be enhanced by compensating the range/Doppler migration due to ego-vehicle and target motion [16].

2. Once a range-Doppler matrix is computed per channel, the angle estimation is performed (1D in azimuth or 2D in both azimuth and elevation, depending on the antenna array topology). Direction of arrival (DoA) estimation is a current area of widespread interest, with much active research. Usually, digital beamforming (DBF) is used for simplicity, employing FFT-based implementations. Still, many research works explore alternatives such as compressive sensing approaches [17, 18], Doppler beam sharpening [19, 20], or machine learning [21]. Sometimes, especially in real-time embedded systems, the detection stage is performed before the angle estimation to reduce the computational load [22], sacrificing the increase in signal-to-noise ratio (SNR) due to spatial coherent integration before detection. This process outputs a 4D radar (hyper)cube.

3. The detection stage then identifies the cells that contain the targets. Usually, a combination of a CFAR detector in some dimensions and peak finding in the rest is used. However, some works have also explored using machine learning algorithms

[10, 23–28]. In this stage, the data is often sparse since most of the space in the field of view does not reflect sufficient power or is simply empty. The detector outputs a 3D occupancy grid, but a conversion to a point cloud is usually performed since it is a convenient format for visualizations or dataset storage.

4. After the detection process and generating a point cloud, additional steps can be implemented to extract more task-relevant information. For instance, in the automotive context, it is critical to know the nature of each detected point to make the appropriate decisions, meaning if this originated from a pedestrian, a vehicle, or some road infrastructure, amongst others. Therefore, applying a classifier on the point cloud is common, usually based on DL techniques [13, 29, 30].

   If needed for the application, tracking algorithms can also be implemented on the point cloud using past information to reduce the estimation noise, eliminate false detections, and predict future target positions based on the trajectory. In the automotive radar domain, tracking algorithms have to deal with the problem of the extended nature of targets over the angular domain [31, 32].

3

Angular Resolution Enhancement Using Self-Supervised Learning

Upscaling images from lower to higher resolution using artificial intelligence (AI) is already happening in real-time in many consumer devices without us noticing. With such approaches, a lower-cost camera can be used to capture an image or a lower-cost GPU can be used to render a frame. After that, the captured/generated image is enhanced with AI. In automotive radar, angular resolution is one of the main challenges, since increasing it natively is not trivial for many reasons. In this chapter, I propose a way to enhance this resolution following a similar upscaling concept: capture the data with a lower angular resolution automotive radar and try to augment it using AI. However, unlike in images from cameras, upscaling after the image is formed is not the best strategy. Therefore, the enhancement is introduced in the time domain radar signal before the 'image' is formed (i.e., before the angle estimation).

Parts of this chapter have been published in:

I. Roldan, F. Fioranelli and A. Yarovoy, "Enhancing Angular Resolution Using Neural Networks in Automotive Radars," 2021 18th European Radar Conference (EuRAD), London, United Kingdom, 2022, pp. 58-61.

I. Roldan, F. Fioranelli and A. Yarovoy, "Self-Supervised Learning for Enhancing Angular Resolution in Automotive MIMO Radars," in IEEE Transactions on Vehicular Technology, vol. 72, no. 9, pp. 11505-11514, Sept. 2023.

3.1. INTRODUCTION

As explained in Chapter 1, high angular resolution is crucial for fully autonomous driving. Firstly, automotive radars need to have high azimuth discrimination capability to separate targets on the road located at the same distance and moving at the same speed. Secondly, high elevation resolution is needed to discriminate which objects can be driven over, such as small debris on the road or speed bumps, which objects can be driven under, such as bridges or tunnel entrances, and which objects should be safely avoided as significant obstacles on the road for the vehicles. Finally, typical automotive targets have multiple scattering points due to their different curvatures and corners [6–8]. A high-angular resolution system makes it possible to capture them, leading to denser point clouds. These richer representations can help for subsequent processing steps in the signal processing pipeline, such as target classification [13, 33–35], road mapping [36, 37], or scene semantic segmentation [29].

The use of larger antenna arrays would increase the angular resolution, but it would be unfeasible to integrate them into conventional vehicles due to their size and cost. For this reason, commercial automotive radars use MIMO [38–40] radar principles to achieve higher angular resolution without increasing the size of the system. A MIMO radar can synthesize virtual arrays with large apertures with only a few transmit and receive antennas.

The basic principle for angle estimation in MIMO radars relies on the extra distance a signal reflected from the same target must travel to reach different antennas in the system. The easiest way to exploit this, known as digital beamforming (DBF), applies a Fourier transform to translate the time delay of received signals at each antenna into a phase shift, which is proportional to the signal's arrival angle. There are more advanced algorithms such as MVDR [41], MIMO-Monopulse [42], subspace methods such as MUSIC [43] or ESPRIT [44], or methods based on compressive sensing [45–47]. However, these methods usually require higher computational costs and longer integration times; hence, they are not always easily applicable in automotive scenarios. Moreover, the angular resolution achieved with all of them is still proportional to the number of virtual antennas of the system [4]. Therefore, a direct way of improving angular resolution would be to increase the number of transmitters and receivers of the system, which is the trend the automotive radar industry is following [22, 48, 49]. For example, the prototype presented in [49] has 4 transmitting elements and 16 receiving elements, or the Texas Instruments MIMO radar [50] used in this work with 12 transmitters and 16 receivers. However, this increase also raises the price and the size of the radar, both critical constraints in the automotive industry.

A recent approach to address this problem is to exploit the relationship between the responses of each virtual antenna to extrapolate new responses of artificial antennas not physically present in the system. Correctly estimated, these new responses can be combined with the aforementioned algorithms, yielding a higher angular resolution. An example of this approach has been studied in [51], where piecewise cubic extrapolation is used, and in [52], by using autoregressive (AR) models in a front-looking SAR geometry. Also, in [53], the authors propose applying a guided generative adversarial network (GGAN) to range-angle radar matrices to produce a higher resolution version. However, this method is used after the angle estimation algorithm; therefore, the data is treated

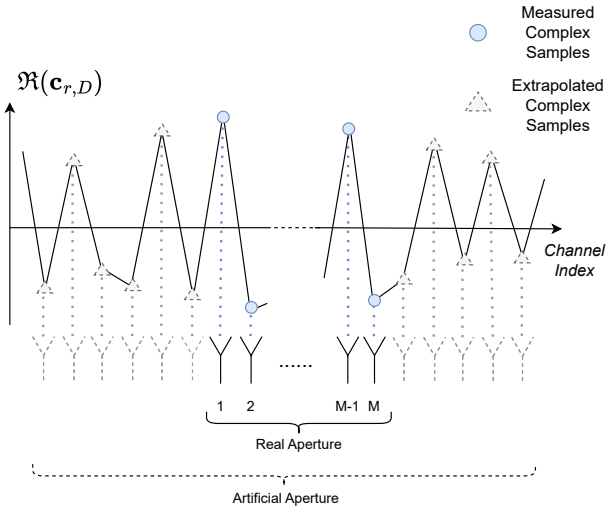


Figure 3.1: Scheme of the extrapolation of the channel vector for a fixed range-Doppler-time cell. For simplicity, this is shown only for the real component of $\mathbf{c}_{r,D}$, but the same procedure is applied to the imaginary part.

as real-valued images. The related results in the state-of-the-art are mostly verified with simulated or point-like targets in very simple scenarios. Moreover, the undesirable effects induced by these extrapolation methods (i.e., the creation of ghost targets or artifacts or the loss of real targets) are not studied in detail despite being critical in an automotive scenario.

This chapter presents a novel framework to enhance the angular resolution of MIMO radar systems without increasing the physical number of antennas. The proposed framework uses the recent advances in machine learning techniques to train a neural network (NN) with data from a large aperture radar (i.e., with high angular resolution) in a self-supervised scheme. Afterward, the trained network can be employed to enhance the angular resolution of a smaller aperture radar in the operational stage without using the larger radar used to gather training data.

3.2. PROPOSED METHOD

As explained in Section 2.1, the signals measured by each antenna of a MIMO radar receiver are the sum of time-delayed versions of the original signal. Thus, the response of each element is highly correlated with the neighbor elements. However, this relationship can be very complex if many targets at different DoA values are present in the scene.

After down-converting, filtering, sampling, and FFT processing, each signal can be arranged in a range-Doppler matrix. These matrices can be stacked to form a tensor with dimensions range-Doppler-channel. For a fixed range-Doppler tuple, the resulting channel vector $\mathbf{c}_{r,D}$ can be used to compute an angular profile using any of the angle

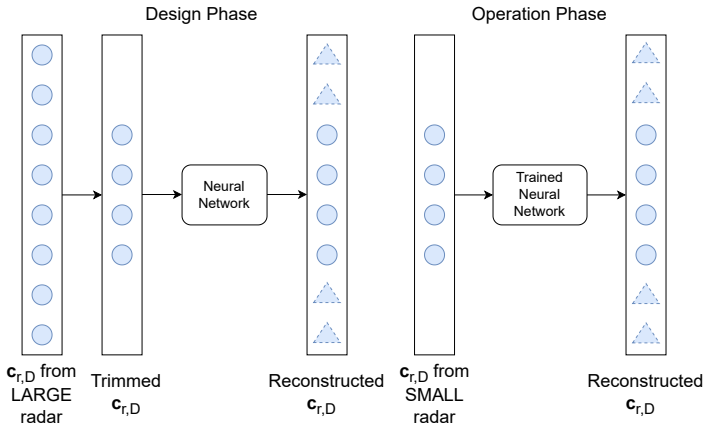


Figure 3.2: A high-level diagram of the proposed approach. In the design phase, data from a high-resolution MIMO radar ('*large*') is used to train an NN that can extrapolate full channel vectors from subsampled channel vectors. In the operational phase, the trained NN can be applied directly to data from a low-resolution MIMO radar ('*small*').

estimation methods mentioned in Section 3.1. However, before estimating the angle, the aforementioned correlation of this vector can be exploited by the proposed method to extrapolate additional complex-valued samples, as illustrated in Figure 3.1 for the real-part component.

Time series extrapolation has been extensively studied in the literature [54]. Many statistical models, such as AR models, are available to describe the likely outcome of a time series in the immediate future. However, these statistical models have been recently outperformed by the rise of neural networks (NNs) if enough diverse data is available. In this work, an NN is used to enlarge the antenna aperture by extrapolating the time series formed by the response in each antenna element. The proposed approach consists of two phases, namely, the design phase and the operational phase. In the design phase, a high angular resolution radar (i.e., a MIMO radar with an extra number of transmit channels, which results in a larger aperture of the virtual array than the operational radar) is used to collect data. Then, this data are used to train an NN in a self-supervised scheme to forecast the response of antenna elements additional with respect to the ones in the operational radar (i.e., a MIMO radar with less number of transmit antennas and, as a result, a smaller aperture of the virtual array). In the operational phase, the NN is used to enhance the angular resolution of an operational radar by forecasting the response of extra antenna elements. It is important to note that a radar with a large virtual aperture is needed only during the design phase and that radar will not be used during the operational phase. In the operational phase, only the operational radar on board the vehicles will be used to enhance its performance in conjunction with the pre-trained NN. A visual representation of this approach can be seen in Figure 3.2.

The approach proposed in this work is independent of the radar's waveform, but the algorithm's start point must be a 3D range-Doppler-channel data cube. In most cases in

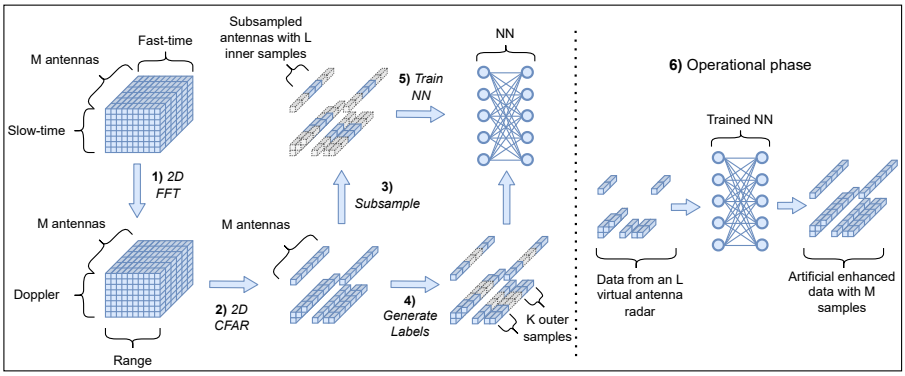


Figure 3.3: Block diagram of the proposed method. First, a 2D FFT (1) is applied to the data, followed by a detection stage (2). The remaining channel vectors (i.e., only those containing targets) are subsampled to generate the training data (3) and the corresponding labels (4). Afterwards, the NN is trained (5). Finally, the trained NN can be used to extrapolate the aperture of a compact radar at the operational phase (6).

automotive, an FMCW radar system is used, and therefore, a 2D FFT can be applied to obtain range-Doppler data [2, 55]. However, given the sparse nature of the data, most of the cells will contain only noise and must be filtered before applying any extrapolation algorithm. A cell averaging constant false alarm rate (CA-CFAR) [9] detector is applied for this, and only the range-Doppler cells containing at least one target will be used. Finally, the remaining channel vectors must be subsampled to match the number of virtual antennas of the low-resolution radar envisaged to be used in the operational phase after training. Then, an NN can be trained using the subsampled channel vectors as input and the extreme samples of the channel vectors as labels (i.e., the samples that must be extrapolated). Therefore, the NN is trained in a self-supervised manner, where the labels are generated automatically from the data. In summary, the steps of the proposed pipeline are:

1. Apply classical radar signal processing to the data obtained with a large high angular resolution radar to get range-Doppler-channel data cubes. In this work, a Hamming window and a 2D FFT have been used.

2. Apply a detector to filter the channel vectors that only contain noise. A CA-CFAR detector was used in this work.

3. Subsample the resulting channel vectors, taking the inner L samples corresponding to the number of antenna elements of the low angular resolution radar that will be used in the operational phase.

4. Generate the labels for the design phase by taking the K outer samples of the channel vectors, where K+L = M, being M the number of virtual antennas of the high-resolution radar.

5. Train the NN in a self-supervised scheme, using the subsampled channel vectors as input and the extreme samples of the channel vectors as labels.

6. Finally, in the operational/testing phase, data captured with a small aperture radar can be processed with the NN to enlarge its aperture, thus enhancing its angular resolution.

Figure 3.3 shows a block diagram of the method. It is important to notice that the proposed method is performed before the DoA estimation algorithm, essentially in order to generate additional artificial antennas from an extrapolation of the available raw radar signals. Thus, any DoA estimation technique, such as MUSIC, MVDR, or a simpler FFT-based beamformer, can be applied after the proposed method. Because of its easy implementation, the simple Fourier beamformer [56] has been used as DoA estimation technique for most of the results of this work when evaluating the enhanced performance provided by the proposed method. However, the single snapshot MUSIC algorithm presented in [57] is also implemented to verify that the proposed approach is estimator agnostic.

The most common NN model for dealing with sequential data is the so-called long short-term memory (LSTM) architecture. The main idea is to incorporate a memory cell that can maintain its state over time, capturing long-term dependencies. LSTMs have been proven state-of-the-art in many different domains [58]. In this case, each new element will be predicted by conditioning on the joint probability of previous values, including the past predictions. For this reason, the network learns to make conservative predictions and avoid successive errors accumulating rapidly, causing the signal to diverge. The proposed model has two stacked LSTM layers with 128 units each, followed by a dense layer. A block diagram of the network can be seen in Figure 3.4. Each $\mathbf{c}_{r,D}$ has been normalized only in its absolute value within the range of 0 to 1, while at the same time taking care that the phase information relevant for subsequent DoA estimation is maintained. This normalization, a common practice in machine learning, is performed to speed up the convergence of the NN. For the extrapolation of the initial samples $\mathbf{c}^{i}_{r,D}$ with $i \in [1, \frac{K}{2}]$ the complex conjugate of the 'flipped' signal has been used. Essentially, the flip$(\cdot)$ operator reverses the order of samples and can be interpreted as the mirrored sequence with respect to the middle element. For a vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ the flip operator is defined as flip$(\mathbf{x}) = (x_n, x_{n-1}, ..., x_1)$. The real and imaginary parts have been concatenated in a new dimension, treating the problem as a multivariate time series extrapolation.

3.3. EVALUATION ON SIMULATED RESULTS

A MIMO radar has been simulated to evaluate the performance of the presented method. The transmitter and receiver arrays have been placed closely located in a monostatic arrangement so that targets in the far-field are at the same distance from both arrays. The transmitted waveforms are assumed to be mutually orthogonal, and the individual elements are isotropic antennas for the $y > 0$ half-plane. The resulting virtual array is a ULA with half-wavelength separation between elements. One million scenes have been generated in a Monte Carlo fashion, where different numbers of point targets have been

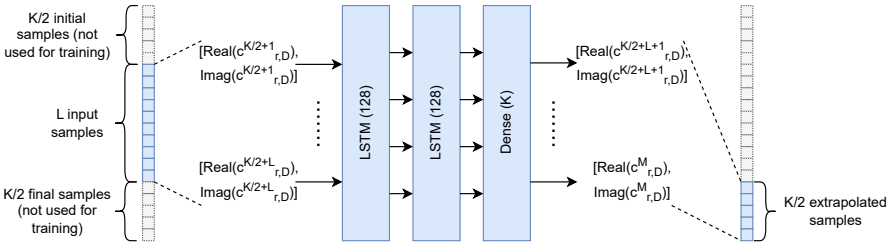


Figure 3.4: The neural network architecture used with two LSTM layers stacked and a dense layer. The input is a matrix with dimensions L×2 composed of the inner samples of the channel vector. The output is a matrix with dimensions $\frac{K}{2}\times 2$, in this case, the extrapolated last $\frac{K}{2}$ samples of the channel vector. Notice that the first $\frac{K}{2}$ can also be extrapolated by using instead the complex conjugate of the input vector after the flip operator.

Table 3.1: Simulation Parameters

Number of scenes	10^6
Number of targets per scene	$X \sim \mathcal{U}(1,10)$
Position of the targets (deg)	$X \sim \mathcal{U}(-70,70)$
RCS of each target (dB)	$X \sim \mathcal{U}(0,10)$
SNR in the scene (dB)	$X \sim (-5,0,5,10,15,20,25)$
M = Num virtual antennas large array	86
L = Num virtual antennas small array	44

placed in the angular space, assuming they are at the same range bin. The RCS and location of the targets have been sampled from a uniform distribution. Table 3.1 summarises the simulation parameters.

The LSTM network has been trained using 900.000 scenes, leaving 50.000 for validation and 50.000 for testing. Adam optimizer has been employed using the default hyperparameters (η=0.001, β_1=0.9, β_2=0.999, ϵ=1e-7), and the mean squared error (MSE) as loss function. The results shown in the rest of the section have been computed using only the test data set.

Since the angle information is encoded in the signal phase, the accurate extrapolation of the phase is a requirement for the method to work. As an example of this phase reconstruction, Figure 3.5 shows the signal phase for a test scene with only one point target and 5 dB SNR, for the array with the large aperture and the array with a small aperture processed with the proposed method. It can be seen how the phase is well extrapolated, which will lead to a higher angular resolution estimation. However, it can also be seen that the phase is not perfectly reconstructed, which could lead to higher side-lobe levels or false alarms in the worst case. For this reason, a statistical analysis of the 50.000 test cases has been done, addressing the accuracy in the angle estimation, the minimum angular separation, and the probability of false alarms (P_{fa}).

First, a receiver operating characteristic (ROC) curve has been computed using a fixed threshold detector, where the threshold is varied by reducing its value from the maximum. For simplicity, a conventional beamformer, also known as Fourier beam-

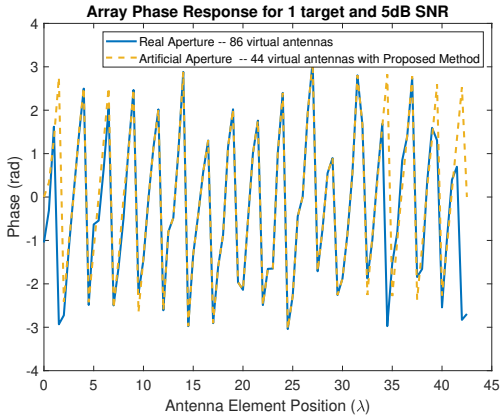


Figure 3.5: Signal phase for the large aperture radar (blue) and for the enhanced version of the small aperture radar using the proposed method (yellow).

former, has been used for DoA estimation. Since the positions of the targets are generated randomly, in many cases, the angular separation between them will be too small to trigger two detections, even for the large aperture array. Moreover, in some scenes, the SNR is set to -5 dB, and weak targets will be present. For these two reasons, the probability of detection (P_d) never reaches 1. Figure 3.6 shows the ROC curves computed for the 50.000 test cases. Two observations can be made from this plot. First, the P_{fa} is not increased in the artificial aperture array; therefore, the proposed method introduces no significant undesirable effects. Second, the P_d is higher in the artificial aperture array than in the small array, even though they use the same number of physical antennas. This effect can be explained that due to the increase in angular resolution, more targets are correctly detected in situations where the targets are closely spaced.

However, the increase of the P_d for a fixed P_{fa} is also induced by the increase of the SNR when processing more coherent antennas, as the target's power will add coherently while the noise will not. For this reason, it is important to analyze the minimum angular separation between detections independently. For each scene, the minimum angular separation has been computed as:

$$\min(|\hat{\theta}_i - \hat{\theta}_j|) \quad \forall i, j \in [1, N_{targets}] \mid i \neq j, \tag{3.1}$$

where $\hat{\theta}_i$ is the estimated DoA of the target i and $N_{targets}$ is the number of targets in the scene. The histograms of the minimum angular separation of the detected targets are presented in Figure 3.7. In Figure 3.7a, it can be seen how the large aperture array is able to resolve targets that are more closely spaced than the small aperture array. This is the expected behavior since it is known that the (boresight) angular resolution of an 86-element ULA and a 44-element ULA are 1.33° and 2.6°, respectively. On the other hand, Figure 3.7b shows the minimum angular separation when applying the proposed method to the small aperture array. It can be seen how the two distributions are more similar, indicating that the proposed method can improve the performances of the small

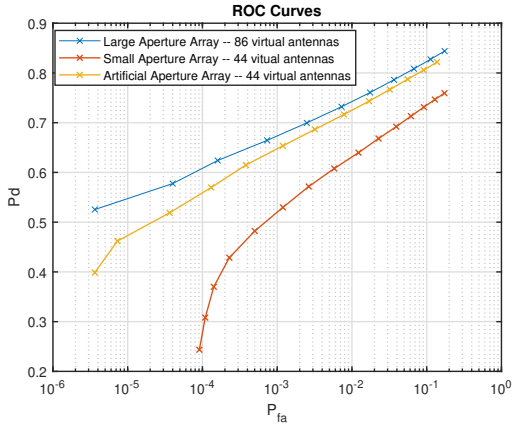


Figure 3.6: ROC curves using a fixed threshold detector after a Fourier beamformer. In blue, the detection capabilities of the large aperture array. In red, the detection capabilities of the small aperture array. As expected, the performance of the small array is lower, both because of the resolution degradation and the decrease in the SNR due to coherent integration. In yellow, the enhanced version with the proposed method using only 44 channel.

physical array, making it more similar to those of the physically larger array. Nevertheless, the distribution for the artificial aperture array is more skewed to a higher angular separation than the original large array but still performs better than the small aperture array in most cases.

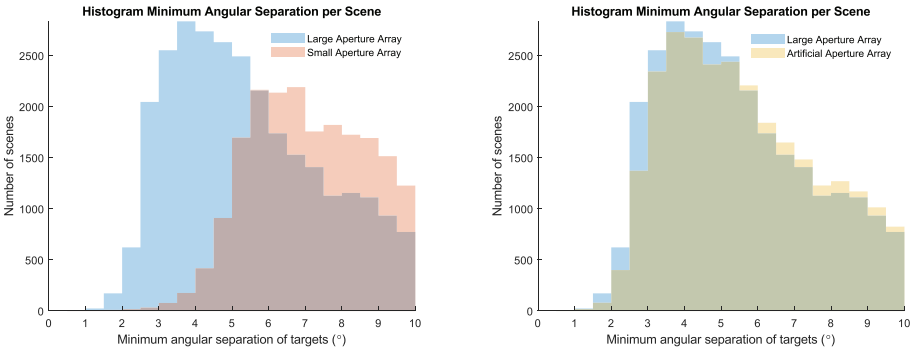
Finally, it is important to analyze if the method induces a degradation in the accuracy of the DoA estimation. In this case, two DoA estimation methods have been evaluated: the same Fourier beamformer used for the previous results in this section and the single snapshot MUSIC proposed in [57]. The MSE of the angular estimation for each method has been computed, and the results have been averaged over the test scenes. Moreover, the Cramer-Rao bound (CRB) has been included for comparison, calculated as [59]:

$$\text{CRB} = \frac{\sigma_n^2}{2}\left|\Re\left[\mathbf{S}^H\mathbf{D}^H(\mathbf{I}-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H)\mathbf{D}\mathbf{S}\right]\right|^{-1}, \tag{3.2}$$

where:

- $\mathbf{A} = [\mathbf{v}(\theta_1),...,\mathbf{v}(\theta_p)]$ is the array steering matrix, formed by p steeting vectors.
- $\mathbf{D} = [\mathbf{d}(\theta_1),...,\mathbf{d}(\theta_p)]$ contains the first derivative of steering vectors $\mathbf{d}(\theta) = \frac{\delta\mathbf{v}(\theta_1)}{\delta\theta}$
- $\mathbf{S} = \text{diag}(s_1,...,s_p)$ formed by the p targets complex response.
- σ_n^2 is the noise variance.

The results are shown in Figure 3.8a for the Fourier beamformer and in Figure 3.8b for the single snapshot MUSIC. In both cases, it can be seen how the artificial aperture generated with the proposed method has improved the MSE with respect to the small aperture array. However, the improvement in the single snapshot MUSIC is lower. This



(a) (b)

Figure 3.7: Histograms of the minimum detected angular separation per scene. (a) compares the large and small aperture array, and as expected, the small aperture array cannot resolve very close-spaced targets. (b) compares the large aperture array and the proposed method applied to the small array. It can be seen how the two distributions are more similar, meaning that the proposed method enhances the angular resolution.

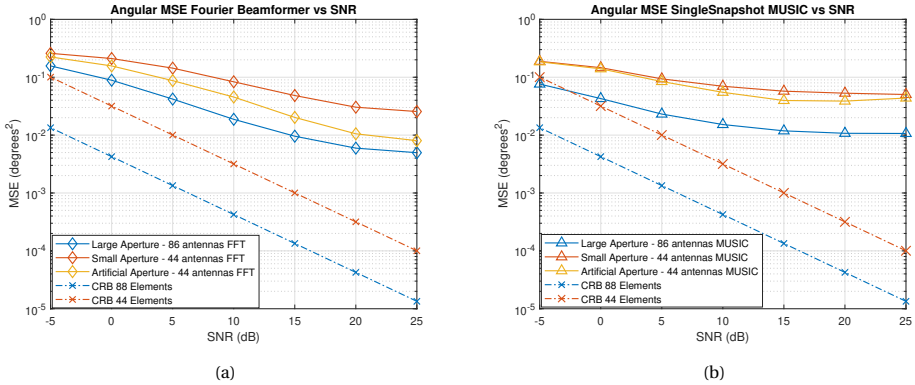
Table 3.2: Radar Parameters for Experimental Validation

Parameters	**Value)**
Center Frequency	78.58 GHz
Effective Bandwidth	320 MHz
Chirp slope	5 MHz / μs
Samples per chirp	256
Chirp period	80 μs
Chirps per frame	128
Transmitted power	+13 dBm
Number of effective azimuthal antennas	86

is because the MSE is only computed in those cases where the targets are resolved correctly. Since the single snapshot MUSIC requires the exact number of targets in the scene as input, it resolves more challenging cases than the Fourier beamformer (i.e., cases that in a real unknown scene would not be resolved). Thus, cases where the DoA angles to be estimated are very close are included for the single snapshot MUSIC evaluation, degrading the method's performance.

3.4. EVALUATION ON EXPERIMENTAL DATA

To evaluate the performance of the presented framework, experimental data from different scenarios have been collected with a commercial FMCW MIMO radar (Texas Instrument MMWCAS-RF [50]), working at 79GHz and capable of synthesizing 86 virtual antennas by cascading four conventional radar chips [50]. The details of the parameters used to collect the data can be seen in Table 3.2. The data collection was designed to be heterogeneous in terms of the type of targets and the relative position between them



(a) (b)

Figure 3.8: In (a), the results for the FFT beamformer. In (b), the results for the single snapshot MUSIC algorithm. The results for the 86 antenna array with its corresponding CRB are shown in blue. The results for the 44 antenna array and CRB are shown in red. Yellow shows the performance when applying the proposed method to the 44 antenna array before the angle estimation.

with respect to the radar. Also, different scenarios with different background clutter have been included.

First, an analysis of the performance using corner reflectors has been done. To this end, 25 scenes were recorded in two scenarios, with corner reflectors with different RCS at the same range and different angles. The detections obtained with the full array have been used as ground truth in this case to avoid errors induced by the lack of equipment for measuring the exact positioning of the targets. Similar to the simulated scenes in the previous section, the ROC curve averaging the results for the 25 recordings has been computed. Figure 3.9 shows the result when performing the DoA estimation using only 44 virtual antennas and the result when the trained NN is used beforehand to augment the resulting aperture artificially. The same data was used to generate both curves, which consist of 11 points each. It is important to mention that the NN has only been trained with simulated data to push its generalization capabilities to unseen, in this case experimental, data. As it can be seen, the artificial aperture, enhanced with the NN, can have better detection probability while keeping the probability of false alarm low. This is because the small aperture is merging targets together due to the limited angular resolution; on the other hand, the proposed method can help separate such targets. Moreover, an example of the angular domain estimation is shown in Figure 3.10a, where two corner reflectors are placed at -30 degrees in the same range. As can be seen when the angular domain is estimated using the whole array ('full aperture'), two main peaks are distinguishable. However, when only 44 virtual elements are used, emulating a smaller aperture array, these two peaks are merged due to poor angular resolution. On the other hand, if the proposed method is applied before the angle estimation, the two peaks can be clearly seen again ('artificial aperture'). It is also possible to observe that the peaks in the artificial aperture estimation are slightly shifted from the full version. The average estimation accuracy using the full aperture estimation as ground truth has been computed for the small and artificial apertures. The mean MSE over the 25 scenarios is 0.0268 deg^2

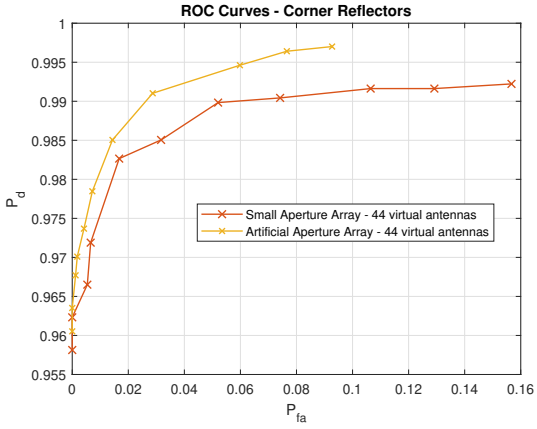
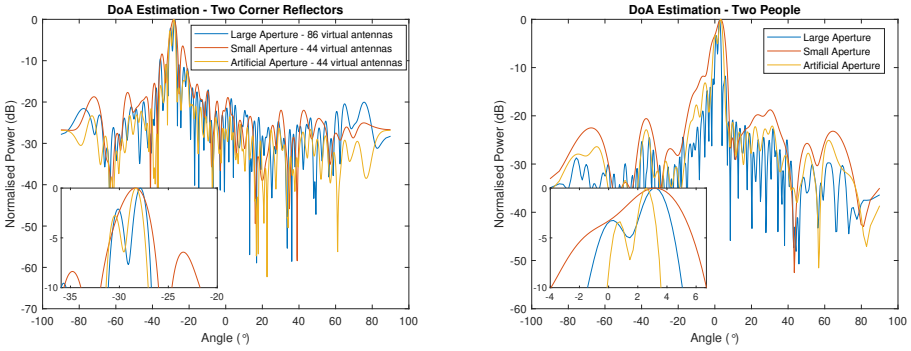


Figure 3.9: ROC curves for 25 scenes with a different number of corner reflectors using the estimation computed with the large aperture array as ground truth. In red, the performance when only 44 virtual antennas are used. In yellow, the performance when the proposed method is applied before the angle estimation.

for the small aperture and 0.0075 deg^2 for the artificial aperture. These numbers confirm the simulation results, where the proposed method does not degrade the accuracy performance but improves it. Finally, the proposed method has been tested with more realistic automotive targets. To this end, a new set of measurements has been collected where two people are present in the scene. Figure 3.10b shows one angle estimation example where the small aperture radar cannot generate two peaks corresponding to the two pedestrians. This is a crucial result since the radar will report a single large object instead of two independent targets, which may lead to critical errors in the subsequent decision-making system. To illustrate this phenomenon better, a dynamic experiment was carried out, where two pedestrians walked towards the radar at the same speed and range, maintaining a constant separation of approximately 1 meter between them in the angular direction. After applying standard radar processing and filtering out the static clutter, the angular domain has been estimated with an FFT for the large, the small, and the artificial aperture. A simple peak detector has been applied in each snapshot, and the results have been plotted in Figure 3.11. If two detections are triggered in the snapshot (i.e., the two pedestrians have been detected), the detections are plotted in black and considered correct detections. On the other hand, if only one peak is found, the two pedestrians have been merged due to poor angular resolution; thus, the detection is plotted in red and counted as an incorrect detection. As seen in the left-hand side of Figure 3.11, when using the 86 virtual antennas, the trajectory of both pedestrians can be clearly followed since in more than 94% of the snapshots, both people are detected. However, when the angular domain is estimated with only 44 virtual antennas, both pedestrians started to merge into a single detection after approximately 15 meters and became completely indistinguishable after 25 meters, as seen in the middle Figure 3.11. In this case, only in 30% of the snapshots can both people be detected. This decrease in resolving capabilities is only due to the lower angular resolution since the targets' SNR is always high in this recording. Finally, if the method presented in this paper is applied before estimat-



(a) (b)

Figure 3.10: Angle estimation of a scene with two corner reflectors (a) and with two people (b) with a small angular separation. In blue, the estimation with the original large aperture array; in red, the estimation with the small aperture array; in yellow, the estimation with the artificial aperture enhanced via the proposed method.

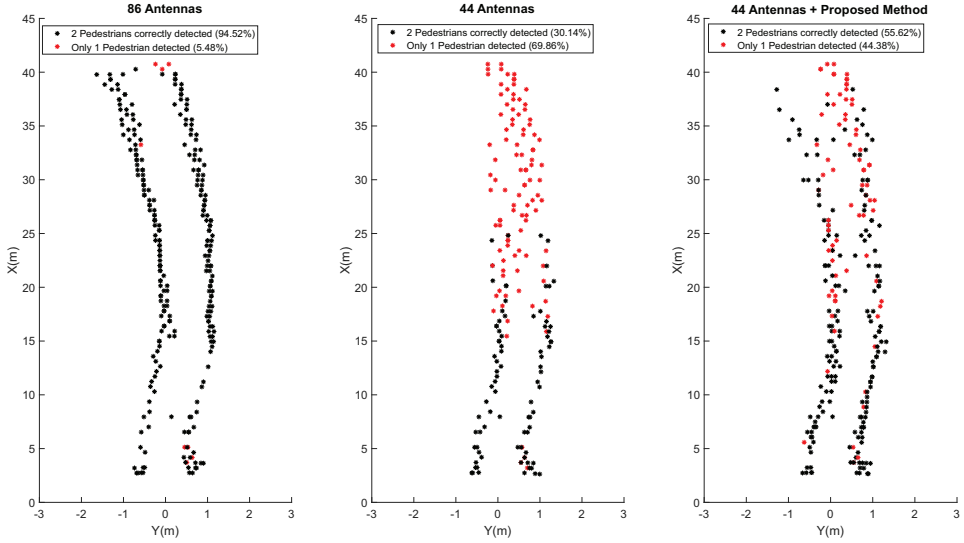


Figure 3.11: Detections triggered in a scene where two people walk toward the radar. If two detections are triggered, meaning that both people have been detected, the detections are plotted in black. On the other hand, if only one peak is detected, the detection is plotted in red. More than 94% of the snapshots are correctly resolved with the full aperture array. With the small aperture array, only 30% are correctly detected, while this number is boosted to 55% with the proposed method.

ing the angular domain, the percentage of correct snapshots is boosted to 55.62%. This effect can be seen in the right-hand side of Figure 3.11, where both trajectories are again discernible. These results show a clear benefit of applying the proposed method since the enhanced radar system can distinguish the two targets at a greater distance. Thus, the decision-making system of the autonomous vehicle will have more time to react.

3.5. CONCLUSIONS

This chapter presents a novel framework to enhance the angular resolution in MIMO radars without increasing the number of physical antennas. A neural network with a self-supervised learning scheme is used to extrapolate the antenna array's response, increasing its aperture artificially. Data from a high angular resolution radar (i.e., a radar with a large aperture) is used to train an LSTM Network that is later used to increase the angular resolution of a low-resolution radar (i.e., a radar with a small aperture). One million simulated scenes have been used to train the system, validating its performance in terms of detection capabilities and estimation accuracy. Two angular estimators, a simple Fourier beamformer and the single snapshot MUSIC, were used, and their performances were compared with and without applying the proposed method. This demonstrated that the proposed method can enhance the angular resolution of MIMO radars without introducing false alarms or degradation in the estimation accuracy.

Moreover, results in experimental data show that this method can help in automotive scenarios to resolve closely spaced targets, such as pedestrians walking closely together, providing accurate information earlier to the decision-making system. To the best of the author's knowledge, this is the first time that neural networks have been used to increase the angular resolution of MIMO radars. Finally, it is important to mention that the presented work aims to propose a framework in which different neural network architectures could be used.

4

BAYESIAN COMPRESSIVE SENSING APPLIED TO AUTOMOTIVE RADAR

When working in engineering research, we often focus on our particular task and develop an algorithm to solve a specific problem. This is sometimes reinforced by the academic and publication system itself, where individualism is often rewarded. In automotive radar, like probably in most engineering fields, this is, in my opinion, the wrong approach, and the big picture should always be kept in mind. Suppose you are researching angular estimation algorithms. In that case, you may want to talk with the person researching array topologies or MIMO waveform design. No part of the automotive radar will be alone in the vacuum, and awareness of the other components is necessary.

During my second year of PhD, I started to supervise the master thesis of Lucas Lamberti. Initially, the idea was to use compressive sensing to improve the angular resolution of an automotive radar. However, Lucas realized that since we knew which algorithm we would use for angle estimation, we could also try to find an array topology specifically designed for that particular algorithm. From there comes the first part of the chapter, trying to connect angular estimation and array topology design together. The second part started in one of our many meetings about compressive sensing. The angular estimation had to run fast, given the constraints of the automotive radar. Could we exploit the Bayesian Compressive Sensing theory to speed up the angular estimation?

Parts of this chapter have been published in:

L. L. Lamberti, I. Roldan, A. Yarovoy and F. Fioranelli, "Sparse Array Placement for Bayesian Compressive Sensing Based Direction of Arrival Estimation," 2024 IEEE Radar Conference (RadarConf24), Denver, CO, USA, 2024, pp. 1-6.

I. Roldan, L. Lamberti, F. Fioranelli and A. Yarovoy, "Low Complexity Single-Snapshot DoA Estimation via Bayesian Compressive Sensing," 2023 IEEE Radar Conference (RadarConf23), San Antonio, TX, USA, 2023, pp. 1-6.

4.1. INTRODUCTION

In Chapter 3, a method to enhance the angular resolution of MIMO radars before performing the actual angular estimation has been presented. In this and the next chapter, the angular estimation problem is explored. Many DoA algorithms have been developed in the last decades, where the most commonly used ones are the multiple signal classification (MUSIC) or the signal estimation parameter via rotational invariance technique (ESPRIT), and their many variants. However, two of the main drawbacks of these algorithms are the need for the a priori knowledge of the number of signals impinging on the array, and the need for multiple snapshots of the same scene for an accurate covariance matrix estimation. From an automotive radar perspective, both constraints present a problem. First, the number of targets (i.e., signals) present on the scene is unknown and stochastic. Second, targets can move at high velocities; thus, coherently processing multiple snapshots is not always possible. In this single-snapshot scenario, the estimation of the inverse of the noise covariance matrix cannot be computed since the estimate is rank deficient. Several variations of the MUSIC algorithm have been developed to increase the performance or to deal with broadband signals, such as Root-MUSIC [60] or polynomial MUSIC (P-MUSIC) [61], but they generally present the same two mentioned problems.

In recent years, a new family of approaches in DoA estimation has been explored, which exploits the fact that the signals are intrinsically sparse in the spatial domain, i.e., only a few signals or targets are present in the scene. In automotive scenes, this is the case when the radar used has a very high range and Doppler resolution, and thus, only a few targets per range-Doppler cell are present. These methods are based on the compressive sensing (CS) theory [17, 62–64].

The CS problem aims to invert or reconstruct a K-sparse signal $\mathbf{x}$ from a set of linear measurements $\mathbf{y}$ in the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, being $\mathbf{A}$ the 'sensing matrix'. The sparse reconstruction approach enables us to solve this linear problem even in the underdetermined case where the number of samples in space, time, and frequency is less than demanded by the Nyquist-Shannon criteria [65]. A unique solution can only be found by imposing the assumed sparsity property in the estimation problem, yielding the sparsest solution. The straightforward formulation utilizes an l_0-pseudo norm optimization problem, which can recover the sparse $\mathbf{x}$ exactly. However, as this combinatorial search problem is computationally expensive, the most popular relaxation is to use the ℓ_1-norm convex optimization problem as:

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad s.t. \quad \mathbf{A}\mathbf{x} = \mathbf{y}, \tag{4.1}$$

also known as the basis pursuit (BP) problem. However, in practice, measurements will always contain noise, and therefore, the least absolute shrinkage and selection operator (LASSO) [66] is commonly used:

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad s.t. \quad ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2 < \epsilon, \tag{4.2}$$

where ϵ is a parameter to bound the noise in the data [67]. It has been proved that CS algorithms provide a guaranteed stable solution of the sparse signal $\mathbf{x}$ for a given sensing matrix if the product $\mathbf{A}\mathbf{x}$ satisfies the restricted isometry property (RIP) defined as:

$$(1-\delta_K)||\mathbf{x}||_2^2 \leq ||\mathbf{A}\mathbf{x}||_2^2 \leq (1+\delta_K)||\mathbf{x}||_2^2, \tag{4.3}$$

for all K-sparse vectors $\mathbf{x}$. However, verifying the RIP for practical sensing matrices is computationally unfeasible, and therefore, this condition is relaxed by employing the mutual coherence property $\mu(\mathbf{A})$ instead:

$$\mu(\mathbf{A}) = \max_{1\leq i\neq j\leq M} \frac{|\langle \mathbf{a}_i^*, \mathbf{a}_j\rangle|}{||\mathbf{a}_i||_2||\mathbf{a}_j||_2} \tag{4.4}$$

where $|\langle , \rangle|$ is the inner product between any two columns $\mathbf{a}_i, \mathbf{a}_j$ of the sensing matrix $\mathbf{A}$. For a good convergence of any recovery algorithm, the mutual coherence $\mu(\mathbf{A})$ must be < 1 [65, 68].

In the DoA estimation problem, assuming the range and Doppler estimation operations have already been performed, equation (2.13) can be simplified to:

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{a}(\theta_k) + n \tag{4.5}$$

where $\mathbf{a}(\theta_k)$ is the steering vector pointing to the kth target, as defined in equation (2.10). Grouping all the steering vectors into the so-called steering matrix as in $\hat{\mathbf{A}} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$, the signal model can be expressed in a compact form as:

$$\mathbf{y} = \hat{\mathbf{A}}\boldsymbol{\alpha} + \mathbf{n} \tag{4.6}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^K$ is the complex amplitude of the K targets. However, this problem formulation is unsuitable for the CS framework since it is not linear (i.e., the parameters to be estimated are in the complex exponential function) and is not sparse. To linearise the problem of estimating θ_k and make it sparse, the angular domain is discretized into a grid of G equally-spaced angles. The steering matrix can then be generated as an over-complete dictionary matrix $\mathbf{A} \in \mathbb{C}^{M\times G}$ where each column is a steering vector $\mathbf{a}_g = \mathbf{a}(\theta_g)$ with $g = 1, \ldots, G$ corresponding to each possible target direction. Moreover, since the number of targets K is not known but $G \gg K$, the vector of signal coefficients $\boldsymbol{\alpha}$ is now expanded into a sparse vector $\mathbf{x} \in \mathbb{C}^G$ with unknown support, corresponding to the targets present at those angles. The resulting signal model is denoted as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \tag{4.7}$$

This over-complete dictionary matrix $\mathbf{A} \in \mathbb{C}^{M\times G}$ can now be used directly in the CS or BCS framework. It should be noted that with the chosen on-grid formulation, targets with DoAs falling not exactly onto one of the G angles will lead to some losses similar to straddle losses in the application of DFTs [9].

4.2. BAYESIAN COMPRESSIVE SENSING

The Bayesian compressive sensing (BCS) framework, as it is typical in Bayesian statistics, introduces a prior belief formulated as a probability density function (PDF) [69–72] to the weights $\mathbf{x}$. In this case, the prior enforces sparsity in the vector $\mathbf{x}$. The BCS problem

then aims to estimate a posterior PDF of this vector **x**. Due to this full posterior PDF, confidence levels about the estimates are also derived.

Using the assumption that **x** in (4.7) is compressible, a new vector $\mathbf{x}_s \in \mathbb{C}^G$ can be defined identically to **x** for the K elements with the largest magnitude, and zero values in the remaining G-K elements, resulting in a K-sparse vector. Similarly, a vector $\mathbf{x}_e \in \mathbb{C}^G$ can be defined identically to **x** for the lowest G-K elements and zeroes in the rest, yielding a (G-K)-sparse vector. Then, (4.7) can be reformulated as follows [69]:

$$\mathbf{y} = \mathbf{A}\mathbf{x}_s + \mathbf{A}\mathbf{x}_e + \mathbf{n} = \mathbf{A}\mathbf{x}_s + \mathbf{n}_e + \mathbf{n} = \mathbf{A}\mathbf{x}_s + \mathbf{n} \tag{4.8}$$

where the influence of $\mathbf{A}\mathbf{x}_e$ may be approximated as zero-mean Gaussian noise $\mathbf{n}_e$ and merged into the measurement noise **n** for simplicity [1]. Therefore, using the Gaussian assumption for the noise, the model likelihood can be modeled as [69]:

$$p(\mathbf{y}|\mathbf{x},\sigma^2) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2\right\}. \tag{4.9}$$

Now, the problem is to estimate the full posterior PDF for **x** and σ^2 given the sensing matrix **A** and the measurements **y**. Different alternatives are presented in the literature for choosing the sparsity-promoting prior, and a common choice is the Laplace PDF. Then, the point estimates of **x** can be found using the maximum a posteriori (MAP) estimate. However, the Laplace PDF is not conjugate to the Gaussian likelihood model, and therefore, the posterior cannot be evaluated in closed form [69]. Another alternative is to use a hierarchical prior with similar sparsity-promoting characteristics [69]. First, a zero-mean Gaussian PDF is defined as the prior on each element in **x** as:

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \prod_{i=1}^{G} \mathcal{N}(x_i|0,\alpha_i^{-1}), \tag{4.10}$$

where $\boldsymbol{\alpha} = [\alpha_1,...,\alpha_G]$ is the precision of each Gaussian distribution and $\alpha_i^{-1} = \sigma_i^2$ the variance of each distribution of x_i. However, this Gaussian prior alone does not enforce sparsity over **x**. A hyperprior in the form of Gamma distributions is then imposed in $\boldsymbol{\alpha}$ and the noise variance σ^2 with precision $\beta = 1/\sigma^2$ as:

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{G} \Gamma(\alpha_i|a,b), \tag{4.11}$$

$$p(\beta) = \Gamma(\beta|c,d), \tag{4.12}$$

where a, b, c and d are usually set to zero. When the posterior is evaluated analytically [69], the result shows that the distribution will strongly peak around $x_i = 0$, and therefore, the hierarchical structure will favor most x_i being equal to zero, thus promoting the sparseness.

The posterior over the weights can be evaluated as a multivariate Gaussian distribution using Bayes' equation, as follows:

[1] Since we are only interested in $\mathbf{x}_s$, the subscript s will be omitted in the rest of the chapter, considering **x** the sparse solution.

$$p(\mathbf{x}|\mathbf{y},\alpha,\sigma^2) = \frac{p(\mathbf{y}|\mathbf{x},\sigma^2)\,p(\mathbf{x}|\boldsymbol{\alpha})}{p(\mathbf{y}|\boldsymbol{\alpha},\sigma^2)} = \frac{1}{(2\pi)^{(G+1)/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}||\mathbf{x}-\mu||^2_{\boldsymbol{\Sigma}^{-1}}} \tag{4.13}$$

with mean and covariance defined as:

$$\boldsymbol{\Sigma} = (\sigma^{-2}\mathbf{A}^T\mathbf{A} + \mathbf{D})^{-1} \tag{4.14}$$

$$\mu = (\sigma^{-2}\boldsymbol{\Sigma}\mathbf{A}^T\mathbf{y}) \tag{4.15}$$

where $\mathbf{D} = \text{diag}(\alpha_0, \alpha_1, ..., \alpha_G)$. The closed-form solution can be obtained by marginalizing over $\mathbf{x}$ as:

$$p(\mathbf{y}|\alpha,\sigma^2) = \int p(\mathbf{y}|\mathbf{x},\sigma^2)\,p(\mathbf{x}|\alpha)\,d\mathbf{x} \tag{4.16}$$

$$= \frac{1}{(2\pi)^{G/2}|\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y}\right\} \tag{4.17}$$

with $\mathbf{C} = \sigma^2\mathbf{I} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}^T$. The solution can be found by a type-II maximization (or evidence maximization) analytically as:

$$\mathscr{L}(\boldsymbol{\alpha}) = \log p(\mathbf{y}|\boldsymbol{\alpha},\sigma^2) \tag{4.18}$$

$$= -\frac{1}{2}\left[G\log(2\pi) + \log(|\mathbf{C}|) + \mathbf{y}^T\mathbf{C}^{-1}\mathbf{y}\right] \tag{4.19}$$

Considering some known initial values for $\boldsymbol{\alpha}$ and β, expression (4.14) and (4.15) can be evaluated. Following this, $\boldsymbol{\alpha}$ and β can be re-estimated using 4.18 in an iterative procedure such as the expectation maximization (EM) algorithm. An efficient algorithm to solve this is the relevance vector machine (RVM) inversion derived in [71], which is used in the next sections. However, this formulation is only valid for real-valued measurements and sensing matrices. In this case, both the measurement vector $\mathbf{y}$ and the steering matrix $\mathbf{A}$ are complex-valued. Therefore, an extra step is needed to adapt the signal model in (4.7) to the BCS framework. Following [73], the signal model is expanded into a real-valued formulation as:

$$\begin{bmatrix}\Re(\mathbf{y})\\ \Im(\mathbf{y})\end{bmatrix} = \begin{bmatrix}\Re(\mathbf{A}) & -\Im(\mathbf{A})\\ \Im(\mathbf{A}) & \Re(\mathbf{A})\end{bmatrix}\begin{bmatrix}\Re(\mathbf{x})\\ \Im(\mathbf{x})\end{bmatrix} + \begin{bmatrix}\Re(\mathbf{n})\\ \Im(\mathbf{n})\end{bmatrix} \tag{4.20}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of the complex numbers. By using this expansion, the dimensions of the problem are doubled.

4.3. ARRAY PLACEMENT FOR BCS DOA ESTIMATION

In Section 4.2, the sensing matrix $\mathbf{A}$ is assumed to be known. In the DoA estimation problem, this matrix is defined by the array topology, and it has been proved that not all sensing matrices will yield a stable, sparse solution [74]. The question this section aims to answer is how to find an appropriate array topology with recovery guarantees,

knowing that BCS will be used as the DoA estimation algorithm. This problem has been theoretically addressed in [75] and inspires the formulation of the proposed approach for designing sparse arrays for radar-based DoA estimation.

Each iteration of the RVM inversion will estimate the covariance matrix (4.14) and mean values (4.15) of each x_i, and the estimated DoA can be obtained by the index of the non-zero mean values. With the covariance matrix $\mathbf{\Sigma}_x$ the differential entropy of the reconstructed signal can be defined as:

$$h(\mathbf{x}) = \frac{1}{2}\log|\mathbf{\Sigma}_x| + c = -\frac{1}{2}\log|\mathbf{D} + \alpha_0 \mathbf{A}^T\mathbf{A}| + c \tag{4.21}$$

The differential entropy can be understood as the overall uncertainty in the reconstruction [76], and therefore, the goal is to select the new sensor position that minimizes this uncertainty. In [75], it is proposed to add new measurement projections based on the Eigenvalue analysis of the covariance matrix $\mathbf{\Sigma}$. However, in our model for the radar-based DoA estimation, the sensing matrix is not a random or arbitrary matrix but rather the array manifold matrix with columns obtained from discretized steering directions, where each row is directly related to a physical antenna position [77]. Taking the eigenvector of $\mathbf{\Sigma}_x$ with maximum eigenvalue as a new measurement is not realizable in terms of antenna array geometry. The differential entropy, therefore, has to be minimized in another, different way. To this end, each new candidate antenna position within a search space has to be evaluated in terms of how including this measurement would benefit the minimization of the entropy. Inspired by [75], an update equation is proposed which allows testing a new candidate row for how much it reduces the entropy if it is included in the sensing matrix [77]:

$$h_{new} = h_{old} - \frac{1}{2}\log(1 + \alpha_0 \mathbf{r}_{new}\hat{\mathbf{\Sigma}}_{\boldsymbol{x}}\mathbf{r}_{new}^H) \tag{4.22}$$

Here, α_0 denotes the noise variance, $\hat{\mathbf{\Sigma}}_{\boldsymbol{x}}$ is the estimate of the covariance matrix obtained in the previous run of the BCS algorithm, and $\mathbf{r}_{new}$ denotes a new candidate row of the sensing matrix $\mathbf{A}$, which in the DoA case is equivalent to a row of the steering matrix. Therefore, adding a new candidate row amounts to including a new candidate antenna position in the sparse array. The important term is the second one in (4.22). This term has to be maximized for each new candidate in order to minimize the entropy as much as possible with each new antenna. This term is defined as inspired by [78]:

$$\delta h(\mathbf{r}_{new}) = \log(1 + \alpha_0 \mathbf{r}_{new}\hat{\mathbf{\Sigma}}_{\boldsymbol{x}}\mathbf{r}_{new}^H) \tag{4.23}$$

This general approach based on entropy estimation via BCS and its minimization is formulated and adapted in the rest of this chapter to the problem of designing effective sparse arrays for radar-based DoA. Figure 4.1 shows a visual representation of the inclusion of new antenna positions based on a uniform linear grid of candidate positions. To guarantee the largest possible aperture and, thus, the best angular resolution, the two outer positions (marked in grey) are always included. Two different versions of the approach have been developed and presented here for the case of a physical linear array and MIMO architecture, respectively [77].

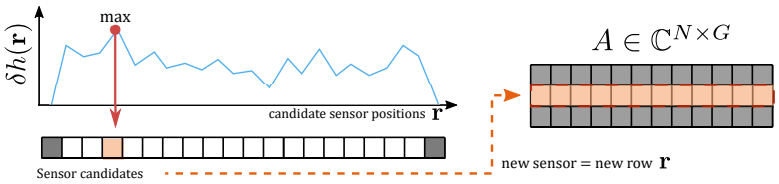


Figure 4.1: Concept visualization for sparse array generation for DoA estimation: the effect of adding each candidate antenna position is checked via differential entropy, and the one which achieves max(δh) in (4.23) is selected.

4.3.1. CASE OF A PHYSICAL ULA

In this version, the algorithm assumes an array of candidate positions based on a physical ULA, thus giving maximum freedom for antenna placement on the underlying grid. This case applies to a system with an array of receiving antennas, such as the ones used in communications, or to a radar system with a single transmitter and multiple receivers. Starting with only the two outer antennas shown in grey in Figure 4.1, a new antenna is included at each iteration until the sparse array is filled. At its core, the proposed method runs the BCS algorithm to estimate the covariance matrix $\mathbf{\Sigma}$ based on a set of training scenes. This estimate of $\mathbf{\Sigma}$, which models the uncertainty of the recovered sparse vector of coefficients directly related to the DoA estimates, is then used to compute (4.23). This is done to select the next row $\mathbf{r}$ that should be added to the steering matrix based on entropy minimization. The algorithm repeats these steps in an iterative manner, as shown in Figure 4.2.

The algorithm generates a reduced (in terms of the number of rows) steering matrix, which serves as the dictionary matrix for the BCS algorithm. The columns of the steering matrix, corresponding to possible target locations (basis functions), are not altered during the process and depend on the angular grid resolution that has been chosen. It should be noted that as the complex-valued steering matrix is expanded into the real-valued matrix described in (4.20), each sensor corresponds to two rows in the expanded steering matrix. Therefore, when the algorithm selects a row from the expanded steering matrix that achieves the highest score for δh, also its twin-row resulting from the complex to real values expansion will be included to represent the new sensor location correctly. This can be graphically seen in Figure 4.3.

4.3.2. CASE OF A MIMO ARRAY

The MIMO array architecture is more challenging since the resulting virtual receiver array used in the DoA estimation results from two separate arrays, the physical transmitter array, and the physical receiver array. Using the algorithm explained in the previous section for the virtual array will lead to the problem of factorizing this array into a physical transmitter and receiver array afterward. This is not trivial, and it might be the case that a feasible factorization cannot be found (especially when overlapping virtual elements are considered). To address this, the iterative approach from the previous section is modified, and new antennas are added to either the transmitter or receiver arrays. This decision is again made based on the BCS entropy, but evaluated on the resulting virtual array [77]. This ensures that no virtual arrays are obtained that cannot be realized

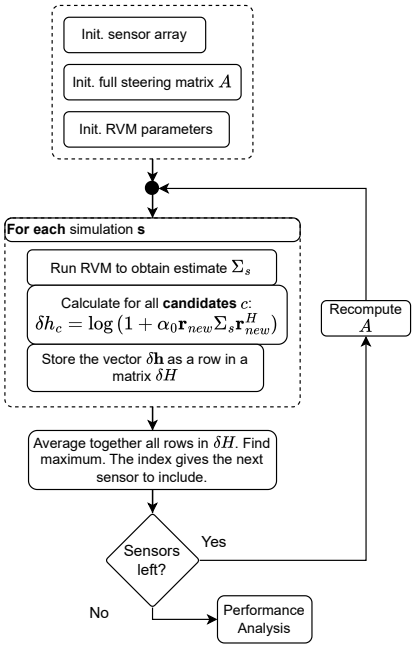


Figure 4.2: Block diagram of the proposed BCS entropy-based sparse array design algorithm for the physical ULA case.

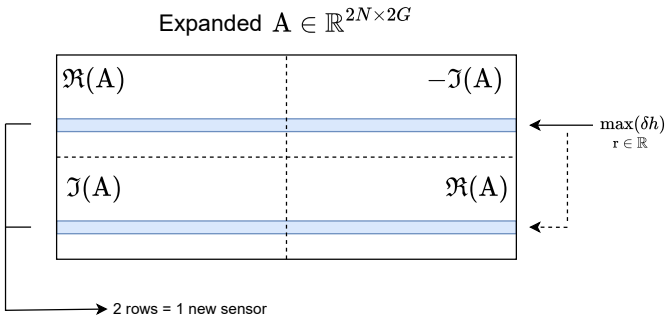


Figure 4.3: Conceptual visualization of the relationship between sensor elements and rows in the expanded steering matrix for the physical ULA case. The upper row is the one for which the entropy update equation maximizes. However, adding the corresponding sensor will add the lower row as well due to the matrix expansion from a complex to a real-valued one.

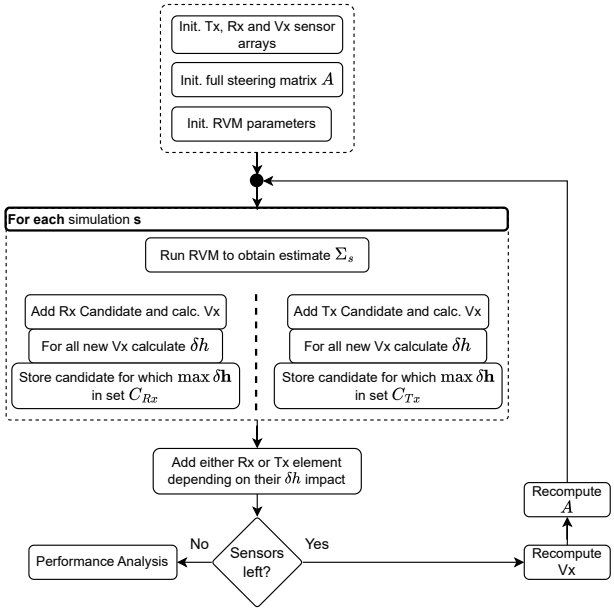


Figure 4.4: Block diagram of the proposed BCS entropy-based sparse array design algorithm for the MIMO case.

in terms of physical transmitter and receiver arrays. Figure 4.4 shows a block diagram of the proposed algorithm.

In the MIMO architecture, adding a new antenna element to either of the two physical arrays usually leads to more than one new element in the virtual array. Unlike in the physical ULA case, this, in a sense, reduces the degree of freedom to place new antennas to maintain the physical feasibility of the resulting sparse array. Moreover, due to the expansion of the complex steering matrix to real values, each of the new virtual sensors leads to two new rows in the expanded matrix. In total, this can cause the addition of many rows to the steering matrix during just one iteration of the algorithm for the MIMO case, illustrated in Figure 4.5.

4.3.3. EVALUATION WITH SIMULATED DATA

Once an array topology has been selected for a desired number of elements, the performance of the DoA estimation using BCS can be evaluated. As explained in the previous section, the antenna elements' possible positions have to be defined to generate the topology. In this case, the array layout of the Texas Instruments MIMO board [50] used in previous chapters has been used as a reference since it will be used in the next section for evaluation with experimental results. On this board, several virtual antenna positions are overlapped, leading to additional constraints to be considered in the sparse array design with the proposed BCS approach with respect to a generic MIMO architecture. The

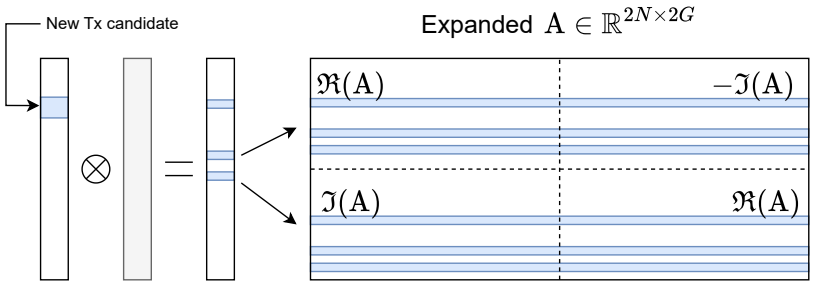


Figure 4.5: Illustration of the relation between new sensors in the Tx MIMO array, resulting in new sensors in the virtual array and new rows in the expanded steering matrix. The Rx MIMO array is considered unchanged in this specific example.

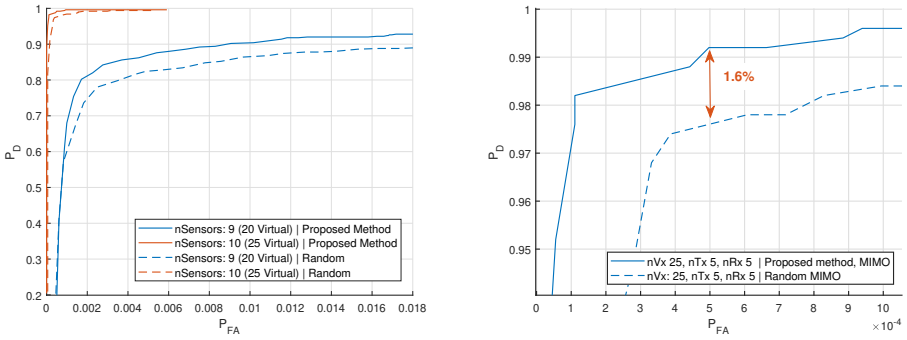
total number of non-overlapped channels for DoA estimation in azimuth is 86.

A case study is presented with 5 simulated point targets placed in a field of view of $\pm 40°$. The minimal distance between two targets is never closer than the Rayleigh resolution limit of the underlying array aperture, which is $\Delta\theta \approx 1.33°$ in this case of an aperture of $86 \cdot \frac{\lambda}{2} \approx 16.96cm$. The steering matrix that is used as input to the BCS method and determines the angular resolution capabilities is therefore discretized to 1°, which will be sufficient for each simulation in this case. The BCS framework is implemented as a single-task method using only one data snapshot. Noise has been added to achieve an SNR ranging from -20dB to 15dB.

The ROC curves are plotted in Figure 4.6 for two different sparse array topologies with different numbers of elements. For both cases, the proposed method performs better than the average random array of the same number of virtual antenna elements, as in line with the results in [78] from the sonar literature. It should be noted that the number of antennas/sensors indicated in the legend by 9-10 refers to the physical Tx and Rx antennas that would constitute resulting virtual arrays with 20 and 25 sensors, respectively. A zoomed-in version can be seen in Figure 4.6b for better visibility. For an array of 25 virtual antennas, the improvement in terms of the ROC curves is indicated for a constant false alarm probability and is approximately 1.6%. In a conventional MIMO setup, these 25 virtual antennas could be realized with 5 Tx and 5 Rx antennas. Translating this to the TI board hardware with 3 Tx antennas and 4 Rx antennas per radar chip, only two such chips would be necessary to realize the sparse array. This would directly translate to a reduction of 50% of the utilized chips, as two of the four implemented chips could be, in principle, removed. While this assessment in terms of required radar chips for given DoA performances is crude and does not consider the full implications of altering the hardware [79], it shows an initial quantitative assessment of the possible benefits of the proposed approach to realize sparse arrays for DoA.

4.3.4. EVALUATION WITH EXPERIMENTAL DATA

The generated sparse arrays have been tested with experimental data collected with the Texas Instruments Cascade board [50] used in previous chapters, which has a total of



(a) (b)

Figure 4.6: ROC curves and zoomed version (b) for two different sparse arrays where the solid line represents the sparse array generated by the proposed method, compared to the average random one with the same number of elements shown in dashed line

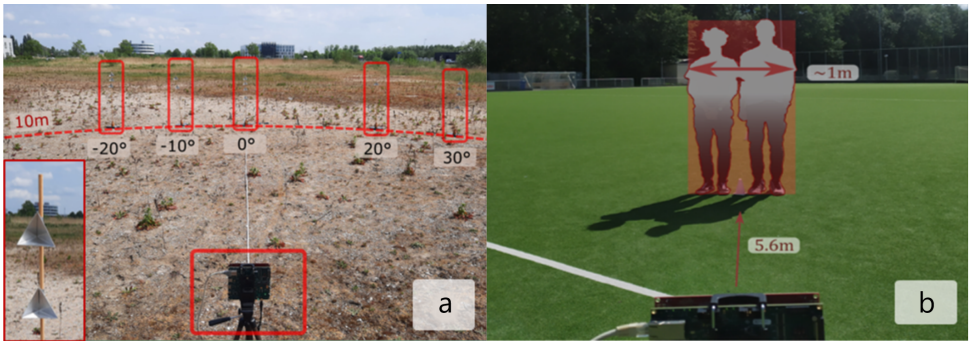


Figure 4.7: Measured scenarios: separated corner reflectors (a); two persons forming an extended target (b).

86 unique virtual antenna positions for DoA azimuth estimation. The data has been collected with all 86 virtual antennas, but only the antenna elements selected by the proposed BCS sparse array design approach will be used for DoA processing. Two scenes have been measured and shown for reference in Figure 4.7, namely 5 corner reflectors with good separation between them (Figure 4.7a) and 2 people together to generate an extended target (Figure 4.7b).

The five corner reflectors have been placed at a fixed distance of 5 or 10 m, with varying angles to the radar's line of sight. A transmitted bandwidth that yields a range resolution of 37.5cm has been set so that the range bin where the targets are placed can be easily identified in post-processing. Using all the 86 unique virtual antennas, the theoretically resulting angular resolution is:

$$\Delta\theta = \frac{\lambda}{d \cdot M \cdot \cos\theta} \cdot \frac{180^\circ}{\pi} \tag{4.24}$$

and for a zero-degree azimuth angle equal to $\Delta\theta_0 = 1.33^\circ$. Good detection performance is

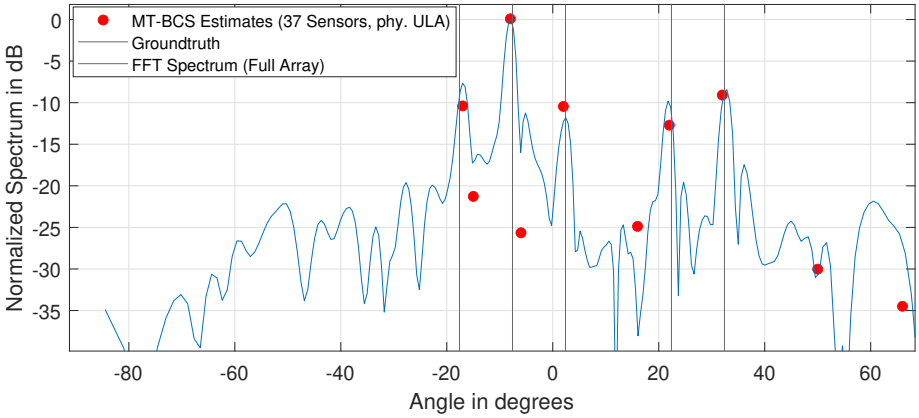


Figure 4.8: Example of BCS-based DoA estimation (red dots) for 5 corner reflectors in the scene.

reached when up to 37 out of 86 possible virtual antennas are included, which is shown in Figure 4.8 compared to a classical FFT-based spectrum using the full 86 virtual sensors.

Also, the trade-off between P_d and P_{fa} has been analyzed with a fixed threshold detector to compare the performance with the average random array with the same number of elements. Figure 4.9 shows the ROC curve for the two generated topologies and the average random array. As expected, the algorithm assuming total freedom in the choice of the positions of the elements (the physical ULA) has the best performance. However, this may not be realizable in a MIMO array, and therefore the red curve is more representative. In any case, both approaches outperform the average random array, and therefore, it would be beneficial to design the array topology using the proposed algorithm.

Finally, it is interesting to test the generated topologies with extended targets. To this end, two people have stood shoulder to shoulder to form an extended target spanning about 10°. Figure 4.10 shows the results of an FFT beamformer (blue) compared with DoA values (red dots and green crosses) estimated by the BCS method using a proposed sparse antenna array. The BCS approach returns high-valued detections in the region where the extended target is located, and also, the highest peaks of the FFT are present. These peaks are most likely related to the scattering centers of the extended target. For the physical ULA-based array and this data capture, around 30 channels are to be sufficient so that the BCS algorithm obtains non-zero coefficients that cover the angular span of the extended target. Similarly, the MIMO-based generated sparse array with 30 virtual elements is able to recover those coefficients as well, which is displayed by the green cross marks in Figure 4.10. In both cases, fewer than 50% of the sensors from the original, full (virtual) ULA of 86 sensors are needed.

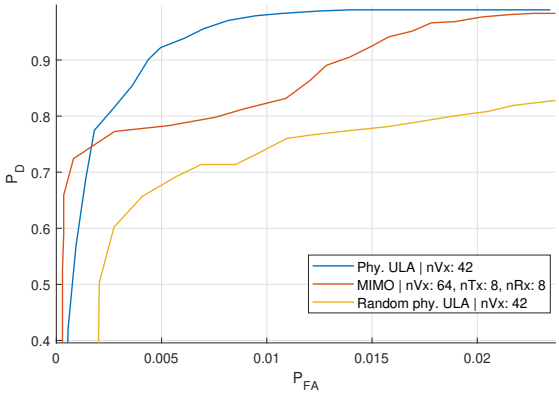


Figure 4.9: Comparison of the performance of three different array topologies. In blue, the performance of the array generate with the physical ULA algorithm. In red, the performance of the MIMO method. In yelow, the average random array.

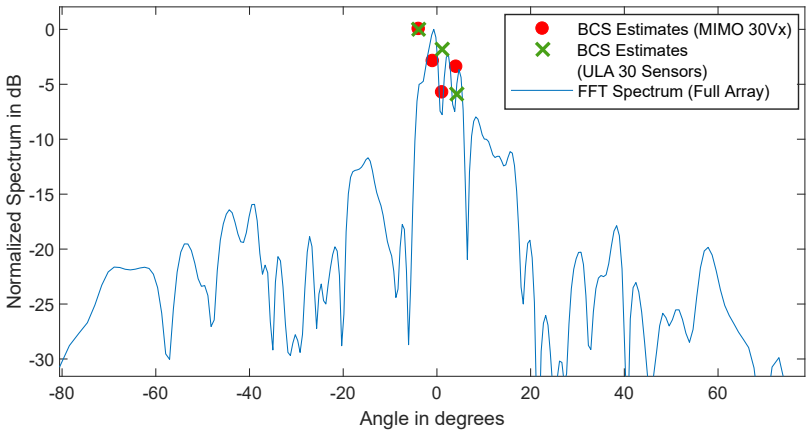


Figure 4.10: Comparison of the performance of three different array topologies. In blue, the performance of the array generate with the physical ULA algorithm. In red, the performance of the MIMO method. In yelow, the average random array.

4.4. EFFICIENT SINGLE-SNAPSHOT BCS DOA ESTIMATION

The previous section has introduced a method for finding a suitable array topology for BCS DoA estimation. With this, the hardware complexity of a MIMO radar can be reduced by using fewer antenna elements at the cost of using an iterative algorithm to estimate the DoA. However, many applications require a very fast DoA estimation; for example, current automotive radars provide a whole radar cube every 50ms. Moreover, the DoA estimation must be implemented in embedded hardware with low computational resources. For this reason, the computational complexity of the DoA algorithms should not be overlooked, and new low-complexity methods are needed.

As explained in Section 4.2, BCS is usually solved efficiently with the RVM inversion and has been proven effective in DoA estimation, using both the standard BCS and the multi-task BCS (MT-BCS) [72] by processing multiple snapshots. However, in any CS-based DoA estimation algorithm, the estimation performance depends on how fine the space discretization is done: the finer the grid, the better the performance, until a point where the sensing matrix starts to be coherent, and the performance degrades. However, increasing the search space also greatly impacts the computational cost of the algorithms since most of them have an exponential computational complexity with respect to the grid size. Thus, different strategies have been proposed to overcome this limitation. For example, in [80], the authors propose a multi-scaling approach where a coarse grid is defined first. Then, by using the estimation uncertainty provided by BCS, a refined grid is formed in the regions of interest.

This section introduces a low-complexity DoA estimation method based on BCS that uses multiple non-overlapping small grids. The proposed method is verified using simulated and measured data, analyzing the estimation accuracy, the resolution capabilities, and the computational cost. Moreover, the performance is compared with different state-of-the-art methods, namely Fourier beamformer, single snapshot MUSIC algorithm [57], and the BCS implementation from [73].

In [72], a sequential method is derived from the fast RVM algorithm [71]. This algorithm works in a constructive manner, adding, deleting, or re-estimating a relevant vector in each iteration. Thus, the complexity of the algorithm is related to the number of relevant vectors k (i.e., the number of targets present in the scene), and it is proved to be $\mathcal{O}(Gk^2)$, where G is the number of angular grid cells. The proposed method adapts this algorithm by dividing the estimation space into smaller sections and performing independent BCS estimations. Then, a common stage is used to combine the estimations per sector into a global estimate, where signal leakage between sectors is resolved.

The first step of the proposed approach is to build W new sensing matrices $\mathbf{A}_w$, being W the number of sectors covering the same angular space. Figure 4.11 shows a visual representation of these sensing matrices. Then, W BCS algorithms are run in parallel, and the output of each of them is concatenated to form a single output. The reduction in computation cost in this step comes from two main reasons. First, because of the quadratic dependency of the complexity with the number of targets and knowing that the sum of squares is less than the square of sums. Thus:

$$\mathcal{O}(k_1^2 + k_2^2 + \ldots + k_W^2) < \mathcal{O}((k_1 + k_2 + \ldots + k_W)^2), \tag{4.25}$$

where k_i is the number of targets in sector i. Second, due to the sparse nature of the

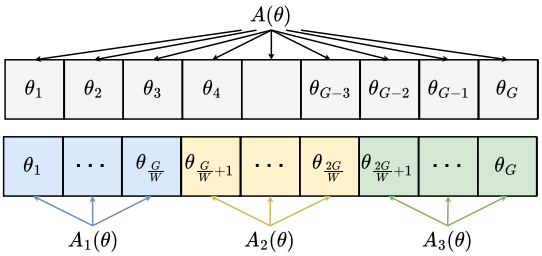


Figure 4.11: Schematic of the DoA angle discretization. In the upper part, a large grid covers the G possible DoAs in a conventional approach. In the lower part, three non-overlapping smaller grids cover the same G possible DoAs with the proposed approach.

scene (i.e., there are only a few targets in each range-Doppler cell), many sectors will contain no targets, leading to a speed-up in the processing. However, signals cannot be physically separated in sectors, and thus, leakage between sectors may happen when the DoA to be estimated is close to the edge between sectors. As an example of this phenomenon, Figure 4.12a and Figure 4.12b are presented. In this case, the discretization grid is 0.5°, and the sectors are defined such as the last cell of one of them is located at -0.5°, and the first cell of the adjacent one is at 0°. Two cases are analyzed, where the targets are placed on-grid and off-grid (i.e., at 0° and at -0.2°). As can be seen in the upper plot of both figures, many targets (i.e., relevance vectors) are generated due to the signal leaking. Thus, a correction stage is needed to compensate for this effect.

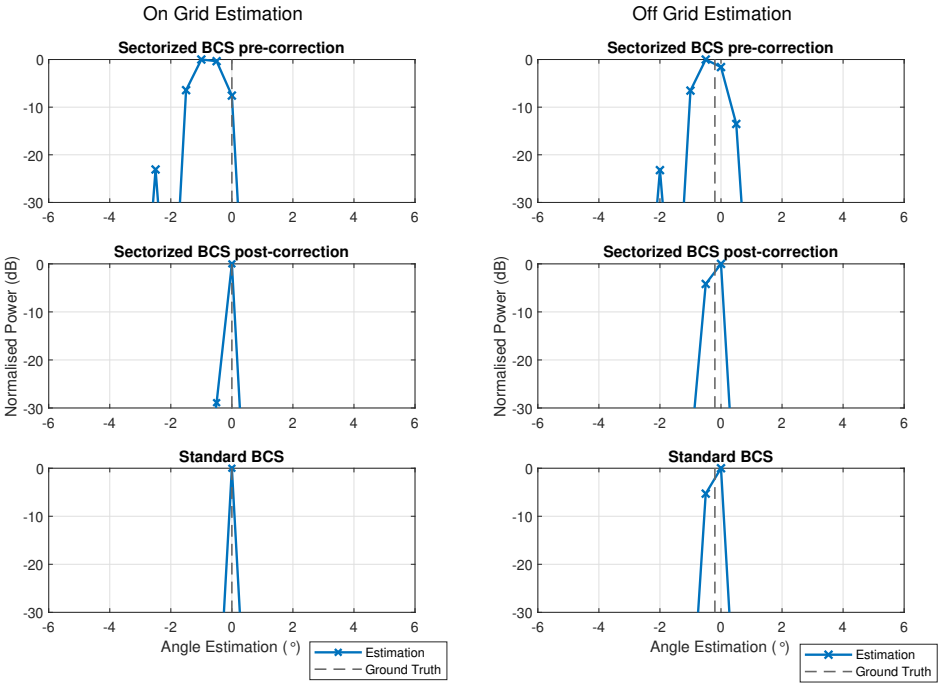
After the W BCS estimations have been performed, a new BCS algorithm is run, but in this case, using the full $\mathbf{A}$. However, only those relevance vectors generated from the previous stage are used as candidate basis vectors. This implies that only the conditions in steps 7 and 8 of the algorithm in [71] need to be evaluated to asses if a relevance vector should be removed. Therefore, the computational cost of this step is very small. The middle plots of Figure 4.12a and Figure 4.12b show the DoA estimation after the proposed correction step, while the lower plots show the estimation with the standard BCS for comparison. As can be seen, both estimations are very close, but the proposed method is much faster.

It is important to note that, unlike for BCS, this approach cannot be applied to most of the DoA estimation methods since their computational complexity is related to the grid size (i.e., the complexity of an FFT is $\mathcal{O}(G\log G)$ while the complexity of the MUSIC algorithm is $\mathcal{O}(G^3)$). The next section presents an analysis to quantify the reduction in computational complexity, as well as some performance analysis.

4.4.1. RESULTS

SIMULATION RESULTS

To evaluate the performance of the proposed approach, a ULA with half-wavelenght separation between elements has been simulated, with isotropic radiation patterns for the $y > 0$ half-plane. Seven thousand scenes have been generated in a Monte Carlo fashion, where a different number of point targets have been placed at different azimuth angles.



(a) (b)

Figure 4.12: DoA estimation for the on-grid (a) and off-grid (b) cases at the boundary between sectors. On the upper plot, the estimation before the correction stage. The middle plot shows the result after the correction stage, and the lower plot shows the standard BCS for comparison

The RCS and location of these targets have been sampled from a uniform distribution. Table 4.1 summarizes the parameters of the Monte Carlo simulation.

First, an evaluation of the estimation accuracy has been performed. To this end, several DoA estimators are compared:

1. The single-snapshot MUSIC (SS-MUSIC) estimator from [57] with a grid size of 0.5°.

2. The conventional Fourier beamforming with a 0.5° grid.

3. The standard BCS with a grid size of 0.5° [72].

4. The proposed approach, sectorized BCS with $W = 10$ sectors covering 18° each, with a grid size of 0.5°.

Figure 4.13 shows the mean squared error (MSE) for the four different estimators. As can be seen, the sectorized BCS slightly outperforms the classical BCS, while the SS-MUSIC and the Fourier beamformer have the worst performance.

Table 4.1: Monte Carlo Simulation Parameters

Parameter	**Value**
Number of scenes	7000
Number of targets per scene	$X \sim \mathcal{U}(1,10)$
Position of the targets [◦]	$X \sim \mathcal{U}(-70,70)$
RCS of each target [dBsm]	$X \sim \mathcal{U}(0,10)$
SNR in the scene [dB]	$X \sim (-5,0,5,10,15,20,25)$
Number elements in array	86

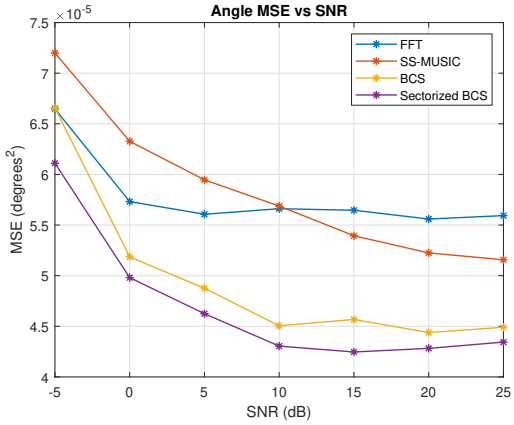


Figure 4.13: MSE for the different DoA estimators as a function of the SNR.

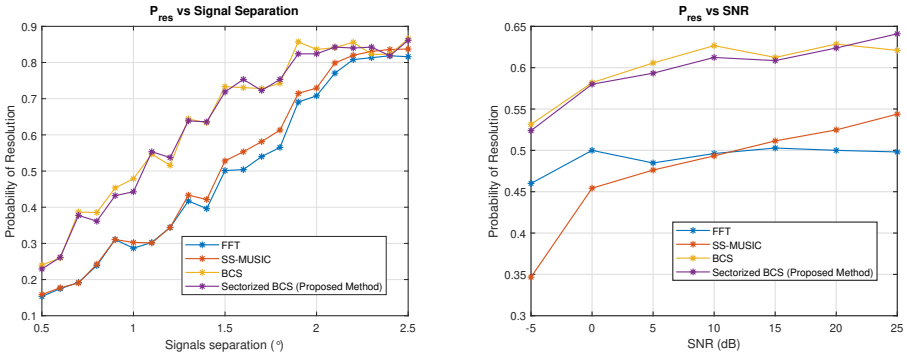
In addition to the MSE, it is important to analyze the resolution capabilities of the proposed method. To this end, the same approach as in [81] is used, where the following random inequality is defined as:

$$\gamma(\theta_1,\theta_2) = \frac{1}{2}[\hat{x}(\theta_1) + \hat{x}(\theta_2)] - \hat{x}(\frac{\theta_1+\theta_2}{2}) > 0. \tag{4.26}$$

With this, two signals are said to be resolvable if the inequality holds and to be irresolvable otherwise. Therefore, the probability of resolution can be written as a binary decision problem as:

$$P_{res} = Pr\{\gamma > 0\}. \tag{4.27}$$

Figure 4.14a shows the probability of resolution for the DoA estimators mentioned above as a function of the true angular separation of the targets. As can be seen, the proposed approach follows the same trend as the standard BCS, with a very similar performance. Moreover, it can be seen how the BCS-based methods have much higher P_{res} than the SS-MUSIC and the Fourier beamformer. Furthermore, in Figure 4.14b, the probability of resolution for different SNRs can be seen, where the methods behave similarly to the previous result.



(a) (b)

Figure 4.14: Probability of resolution of the DoA estimators. In (a) as a function of the true angular separation of targets. In (b) as a function of the SNR

Table 4.2: Computational Time for Different DoA Estimators

Method	**Time (ms)**
SS-MUSIC [57]	16.3
FFT	0.035
BCS [73]	101.5
Sectorized BCS (Proposed Approach)	5.0

From the results presented in this section, it is clear that the proposed sectorized BCS has estimation accuracy and resolution capabilities similar to the standard BCS, overperforming the Fourier beamformer and the Single Snapshot MUSIC algorithm. However, as explained in the previous section, the more considerable improvement is in the computational cost of the method. To analyze this improvement, the average computational time over the 7000 scenes for each DoA estimator has been calculated, and results are shown in Table 4.2. As it can be seen, the proposed method can perform the DoA estimation in only 5ms, which is approximately 20 times faster than the standard BCS (68% of the time is spent in the first step of the method, and 28% on the following correction stage).

EXPERIMENTAL RESULTS

A field experiment has been carried out with a 77GHz FMCW radar with an 86 virtual ULA to evaluate the performance of the proposed method with experimental data. Two corner reflectors have been placed at a range of 5.6m with a distance of 13cm (equivalent to approximately 1.3°) between the centers, as shown in Figure 4.15. A single snapshot has been captured, and the same estimators mentioned in the previous section have been applied. As seen in Fig. 4.16, both BCS-based methods can resolve two peaks, while the FFT and SS-MUSIC fail to do so. Notice that the targets are placed in the boundary of two sectors, so the example shown is the most challenging case for the proposed method. The estimation is as good as the standard BCS, but with significantly less computational

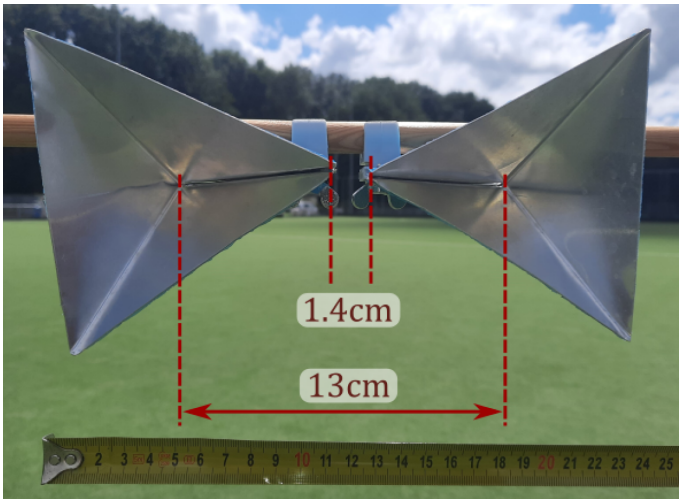


Figure 4.15: Placement of the corner reflectors with 13cm between their centers, leading to 1.3° separation at 5.6m distance from the radar.

complexity.

Finally, it is important to mention that the estimation accuracy cannot be evaluated due to the lack of an accurate ground truth. The experimental results are shown only as an example, and future work will perform a statistical analysis with additional data.

4.5. CONCLUSIONS

This chapter introduces two different topics related to BCS. The first one is how to generate an array topology given that BCS will be used as the DoA estimation method. The second one tackles how to reduce the computational load of BCS when using the RVM implementation. Leveraging both studies, it may be possible to design an automotive radar that uses BCS as the angle estimation method.

The array topology generation works based on the intrinsic measurement of uncertainty provided by BCS. This approach is versatile and applicable to both physical ULA and MIMO architectures. The approach works by selecting a suitable number of antenna elements based on an entropy minimization operation. It is shown by means of simulations and experimental results with corner reflectors and an extended target that the proposed approach for antenna element selection results in faster uncertainty reduction compared to random selection, confirming the results of [78] in sonar. This decreased uncertainty leads to a noticeable performance improvement in the detection of targets while maintaining accuracy in their DoA estimation. From a practical perspective, the results in this work suggest the feasibility of utilizing sparse arrays when combined with a BCS DoA estimation algorithm, offering the possibility to reduce hardware complexity, cost, and energy consumption by utilizing fewer antenna elements.

The second part of the chapter has shown a low-complexity single-snapshot DoA estimation method that uses a sectorized approach to exploit the advantages of BCS. The

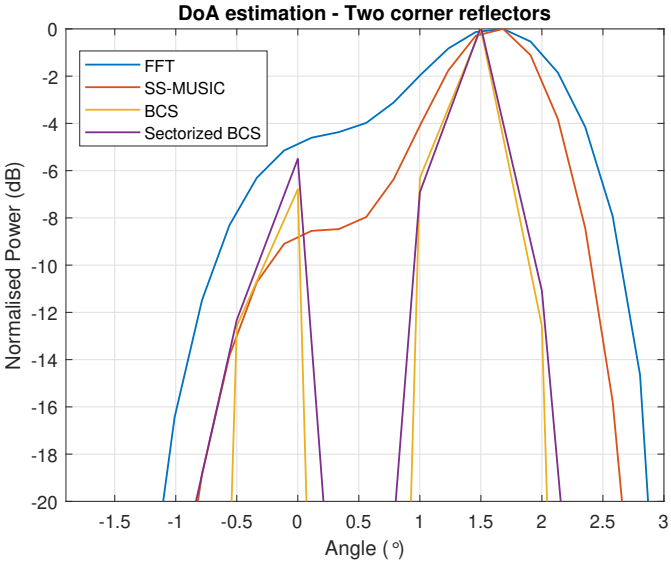


Figure 4.16: DoA estimation results of two corner reflectors with 1.3° separation between them. The proposed sectorized BCS as well as the standard BCS can resolve the two targets. The FFT beamformer and SS-MUSIC only show one peak.

performance of the proposed method has been analyzed in terms of estimation accuracy, resolution capabilities, and computational cost. To this end, a Monte Carlo simulation has been performed with 7000 runs, simulating different scenarios with a varying number of targets with different characteristics. Results prove that the proposed method can perform as well as the standard BCS in terms of accuracy and resolution but with a significant reduction of 20 times the computational cost. Moreover, the method's performance has been verified experimentally using commercially available radar, proving that the proposed method can work well with measured data.

5

TOTAL VARIATION COMPRESSIVE SENSING APPLIED TO AUTOMOTIVE RADAR

This chapter may look like a contradiction to the previous one. I mentioned a few pages before that in automotive radar the angular estimation space is sparse due to the high range and Doppler resolution; therefore, we can apply compressive sensing. However, I realized this may not always be true at close ranges, especially with the rapid increase in the number of antenna elements in automotive radars. Luckily, if you can find a domain where the signal you want to estimate is sparse, you can still apply compressive sensing. This chapter is about transforming the 'maybe-not-so-sparse' angular domain into a sparse domain using the discrete gradient. The advantage is that even if the original space was sparse, this transformation would also yield a sparse domain, so nothing would be broken or lost. The disadvantage is that the algorithm is computationally heavy and not (yet) implementable in real time without further optimization.

Parts of this chapter have been published in:

I. Roldan, F. Fioranelli and A. Yarovoy, "Total Variation Compressive Sensing for Extended Targets in MIMO Radar," 2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM), Trondheim, Norway, 2022, pp. 61-65.

I. Roldan, F. Fioranelli and A. Yarovoy, "Total Variation Compressive Sensing for 3D Shape Estimation in Short-Range Imaging Radars," in IEEE Transactions on Radar Systems, vol. 1, pp. 583-592, 2023.

5.1. INTRODUCTION

In the previous chapter, a CS-based method was presented, and like every CS problem, the estimated signal needed to be sparse in some domain. It was argued that because of the high resolution in range and Doppler in automotive radars, the angular estimation space will be sparse (i.e., only a few targets will be present per range-Doppler cell). This is the case in many applications, and the previous method can be applied directly. However, it may happen that with the increase in the number of antennas in automotive radars, targets start behaving like extended objects in short and mid ranges, resulting in new challenges that must be addressed [2]. From this, two main problems can be derived for conventional direction of arrival (DoA) algorithms [4]. First, targets will be composed of many coherent scatterers; thus, subspace methods such as MUSIC or ESPRIT are not directly applicable (even in multiple snapshots). These methods assume that the covariance matrix of the measurements is not singular, meaning that the sources are totally uncorrelated concerning each other. Several algorithms have been proposed to deal with coherent sources, spatial smoothing being the most common one [82, 83]. The spatial smoothing technique divides the array into smaller overlapping subarrays and aims to build a smoothed covariance matrix by averaging the subarrays' responses. However, this algorithm was developed for the case of only a few correlated sources present, for example, with multi-path components or jamming situations. In [84, 85], it is proved that a necessary condition for the covariance matrix to have full rank is to apply the smoothing operator at least the same number of times as the largest coherent sources group. Therefore, in the situation of extended targets with many coherent sources, this smoothing approach is not applicable. Furthermore, these algorithms rely on multiple snapshots to estimate the covariance matrix, which may also not be available due to the fast-changing scenes in automotive scenarios. For these reasons, the DoA estimation problem using large arrays in automotive radar does not fulfill the typical requirements to apply subspace methods.

The second problem of dealing with extended targets with high-resolution radars is that the estimation space may not be sparse anymore. For a fixed distance (i.e., a fixed radar range cell), several extended targets may be present, especially in dense urban environments, such as the facade of a building and a few vehicles. Each target will return detected power within several azimuth and elevation cells, making the scene for DoA estimation not sparse. Therefore, all methods based on conventional compressive sensing (CS) will not be reliable. A valid alternative is to apply data-independent approaches such as delay-and-sum (DaS) beamforming [86] (also known as Fourier beamformer for narrow band signals). However, they suffer from low resolution and high sidelobe levels, which is undesirable in automotive applications.

For the above reasons, it is necessary to design algorithms for the new generation of 4D imaging radars. This chapter presents two methods to overcome the problem of angular estimation for extended targets using the compressive sensing theory but with a total variation (TV) regularization. With this proposed formulation, there is no need for the estimation space to be sparse, but sparsity is required only for the estimated signal's discrete gradient. The discrete gradient of the signal will be high when sharp changes in the signal occur and low, theoretically zero, when the signal is essentially constant. Specifically, the discrete gradient will only contain high values when the power of adja-

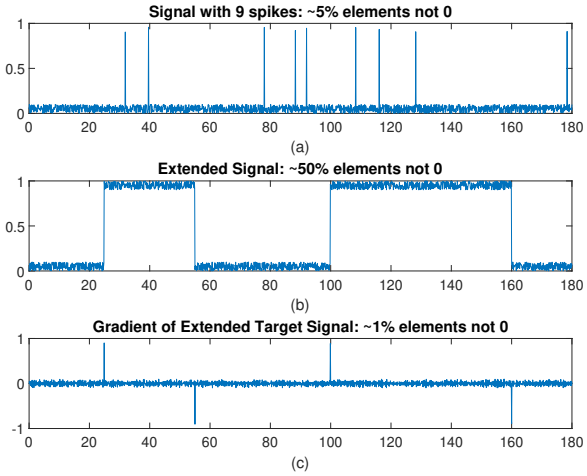


Figure 5.1: Sketch of a simplified generic signal to illustrate the concept of sparsity related to the proposed methods. In (a), an example of a sparse signal, where only 9 spikes are present. In (b), a signal with two relatively large, extended areas of high values is shown, making this signal no longer sparse. Finally, (c) shows a signal generated from the computation of the discrete gradient of the signal in (b), which makes the resulting signal sparse again.

cent cells is very different, for example, in the transitions between a DoA cell occupied by a target and one with empty space. For this reason, even large extended targets occupying many cells will only generate two spikes in the gradient, at the beginning and at the end of the target's shape in an angular profile. Figure 5.1 shows a simplified sketch illustrating the concept of sparsity and discrete gradient. As seen, only 5% of the values are occupied by useful signal, i.e., they have higher values, easily detected compared to the noise floor of the other values. Thus, the signal can be considered sparse. However, if two large, extended areas of high values are present as in the other generic signal shown in Figure 5.1(b), the number of cells/bins with high detectable values will be much higher (e.g., approximately 50% of all the cells in this example). This means that this second signal is not sparse. However, if the discrete gradient of the non-sparse signal is taken, the resulting signal will have only a few high values in the transitions between target and noise. The discrete gradient output can be seen in Figure 5.1(c), where only about 1% of the cells are occupied.

Previous research has used the concept of TV regularization and CS applied to radar data. In [87] TVCS is applied to ultrashort range synthetic aperture radar (SAR) imaging with non-uniform sampling. While the performance reported is impressive, the proposed approach is challenging to apply to automotive radar for several reasons. The imaging method presented, based on SAR, needs observation times in the order of tens of minutes while the target is static. Moreover, the observation space is in the order of hundreds of millimeters. In automotive radar, only one snapshot is usually available, which is in the order of milliseconds, while the observation space can be in the order of hundreds of meters. In the work presented in [88], the authors applied TVCS using an antenna array, and the results were evaluated using uncorrelated point-like sources.

While the presented scenario is more similar to those relevant to automotive radar, the hypothesis of uncorrelated point-like sources is hardly applicable to realistic automotive radar.

In this chapter, two different TVCS methods are presented. The first one independently estimates the azimuth and elevation space for each range cell, whereas the second one performs a joint range-azimuth-elevation estimation. It is important to note that these methods promote a sharper reconstruction of the target's edges, which makes the data more suitable for estimating the extent of targets in the radar field of view and classifying them in subsequent processing steps.

5.2. INDEPENDENT 2D TVCS

In this section, the beat frequency contained in $f_r(r_k)$ in equation (2.11) is estimated using a Fourier transform, and N sub-problems are solved to estimate the azimuth-elevation space, where N is the number of range cells. To cast the problem in a CS form, the azimuth-elevation space is discretized in a G points grid, with G being the product of the cells in azimuth G_θ and the cells in elevation G_ϕ. Targets are assumed to lie on this grid. Then, a dictionary matrix $\mathbf{A} \in \mathbb{C}^{M_x M_z \times G}$ can be constructed with steering vectors pointing to the defined grid, as explained in Section 4.1, expressed as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + e, \tag{5.1}$$

where $\mathbf{y} \in \mathbb{C}^{M_x M_z \times 1}$ is the sample vector for a fixed range cell, and $\mathbf{x} \in \mathbb{C}^{G \times 1}$ is an unknown complex-valued vector containing the reflected power of the targets and encoding their angular position. Note that the $\mathbf{A}$ matrix is known, pre-computed a priori by stacking $\mathbf{a}(\theta_g, \phi_g)$ steering vectors pointing to each $g = 1, 2, ..., G$ grid cell. The definition of the steering vectors can be found in (2.10).

In the standard CS framework, the undetermined system of equations in (5.1) is solved by enforcing sparsity in the signal $\mathbf{x}$, i.e., $G \gg K$, and thus $\mathbf{x}$ contains only a few non-zero elements. The following ℓ_1-norm minimization problem is then solved:

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 + \frac{\mu}{2}||\mathbf{A}\mathbf{x} - \mathbf{y}||_2, \tag{5.2}$$

where μ is the quadratic penalty parameter that balances the data fidelity to account for the noise and $||.||_2$ is the ℓ_2*-norm*.

Some studies have solved the optimization problem in (5.2) for radar-related applications [89, 90], but with the assumption that the signal of interest $\mathbf{x}$ is sparse. As discussed in the introduction, this may not always be the case in the problem of DoA estimation in automotive radar discussed in this work, where many multiple extended targets can be present in the scene of interest. For this reason, an alternative approach is proposed by using the TVCS framework. In this case, the assumption is that the discrete gradient of the signal is sparse, which is a more appropriate model for large extended targets (i.e., there are only a few transitions between empty cells and cells with returned power from targets). Moreover, TVCS enforces a sharp estimation of the targets' boundaries, which is essential for object classification and estimation of their extent. The TVCS optimization problem is formulated as follows:

$$\min_{\mathbf{x}} ||\nabla\mathbf{x}||_1 + \frac{\mu}{2}||\mathbf{A}\mathbf{x} - \mathbf{y}||_2, \tag{5.3}$$

However, for MIMO radars with a large number of channels, a discretization of the angular space at the sub-degree level is desired. Therefore, the above problem becomes unsolvable in a reasonable time (e.g., assuming a 48×48 URA and 1° cell in both azimuth and elevation dimensions, $\mathbf{A}$ would be a 2304×32400 complex matrix). In this work, the $\mathbf{A}$ matrix is substituted by a partial 2D inverse discrete Fourier transform. To better understand this implementation, let us assume a generic time signal $\mathbf{z} \in \mathbb{C}^{G\times 1}$ and its Fourier transform in matrix form as:

$$\zeta = \mathbf{W}\mathbf{z} \tag{5.4}$$

where $\zeta \in \mathbb{C}^{G\times 1}$ is the transformed frequency domain and $\mathbf{W}$ is the standard DFT matrix. This formulation is still valid if, instead of a single signal, we have many signals organized in a matrix form. However, in our case, the time signal has fewer samples than the desired frequency domain one, and is not necessarily uniformly sampled. Therefore, the DFT matrix $\mathbf{W}$ has to be modified accordingly by selecting only those columns in which we have samples in the $\mathbf{z}$ vector. Since the frequency domain is in a generic case also non-uniform, this transformation is at times also known as non-uniform discrete Fourier transform type III (NUDFT-III) [91]. Moreover, knowing that the DFT matrix $\mathbf{W}$ is unitary, we can define the IDFT matrix as $\mathbf{W}^{\mathrm{H}}$. Again, in our case, only the appropriate rows of the IDFT matrix will be used due to the sparse sampling in the time domain. Instead of computing the matrix multiplication, the fast Fourier transform algorithm is solved, using only some of the coefficients (due to the sparse sampling), and therefore, the operator pIFFT(·) is defined as the application of the partial Fast Inverse Fourier transform algorithm, which yields the same results as multiplying the matrix by the IDFT matrix $\mathbf{W}^{\mathrm{H}}$ with only the appropriate rows. In our case, we apply the pIFFT(·) operator in two dimensions, so a new operator, pIFFT2(·) is defined as $\text{pIFFT2}(\cdot) = \text{pIFFT}(\text{pIFFT}(\cdot)^{\mathrm{T}})^{\mathrm{T}}$.

Then, by reshaping $\mathbf{x}$ into a matrix $\mathbf{X} \in \mathbb{C}^{G_\theta \times G_\phi}$ and $\mathbf{y}$ into a matrix $\mathbf{Y} \in \mathbb{C}^{M_x \times M_z}$, the new formulation is given by:

$$\min_{\mathbf{X}} ||\nabla \mathbf{X}||_{1,1} + \frac{\mu}{2}||\text{pIFFT2}(\mathbf{X}) - \mathbf{Y}||_{\mathrm{F}}. \tag{5.5}$$

where $||\cdot||_{\mathrm{F}}$ is the Frobenius norm, and $\nabla(\cdot)$ is the isotropic gradient. The optimization problem in (5.5) can now be solved efficiently using the total variation minimization by augmented Lagrangian and alternating directions algorithm (TVAL3) [92, 93]. This algorithm casts the problem into a sequence of unconstrained optimization problems, and the well-known augmented Lagrangian is used. TVAL3 has been proven to be one of the most efficient total variation regularization solvers in many fields [94, 95], and in combination with the fast partial Fourier transforms, it allows the estimation of the complete 2D angular space in azimuth and elevation. The first step of TVAL3 is to reformulate (5.5) into an equivalent form, introducing a set of auxiliary variables as:

$$\min_{\mathbf{W},\mathbf{X}} ||\mathbf{W}||_{1,1} + \frac{\mu}{2}||\text{pIFFT2}(\mathbf{X}) - \mathbf{Y}||_{\mathrm{F}} \; s.t. \; \nabla\mathbf{X} = \mathbf{W}. \tag{5.6}$$

Then, to cast the problem into a sequence of unconstrained optimization problems, the well-known augmented Lagrangian is used:

$$\mathscr{L}(\mathbf{W}, \mathbf{C}) = ||\mathbf{W}||_{1,1} - \nu^T(\nabla\mathbf{X} - \mathbf{W}) + \frac{\beta}{2}||\nabla\mathbf{X} - \mathbf{W}||_{\mathrm{F}}^2 + \frac{\mu}{2}||\text{pIFFT2}(\mathbf{X}) - \mathbf{Y}||_{\mathrm{F}}^2, \tag{5.7}$$

where $.^T$ is the transpose operator, β is the penalty term, and v is the Lagrangian multiplier. Therefore, (5.6) can be reformulated as:

$$\min_{\mathbf{W},\mathbf{X}} \mathscr{L}(\mathbf{W},\mathbf{X}). \tag{5.8}$$

Finally, this unconstrained optimization problem series is solved using the alternating direction method proposed in [92], where the pseudo-code of the algorithm can be found. Following the authors' recommendations [92], the default hyperparameters are used, and the only one tuned for the problem at hand is the hyperparameter μ, which has to be related to the noise level and sparsity of the specific scene. However, since these values are not known *a priori*, μ has been selected using a data-driven approach based on the simulations described in section 5.4. In this case, μ has been selected as 2^8. Moreover, the isotropic operator has been used. Additionally, it should be noted that the authors also provided the necessary adaptations of the algorithm to deal with complex-valued measurements [92].

5.3. 3D TVCS

Automotive radars can also provide high-resolution range estimation. Therefore, targets will occupy several cells in the range dimension as well. Thus, the extension of the method explained in the previous section to the 3D joint range-angle estimation is a logical step. This 3D estimation problem can be modeled as follows:

$$\mathbf{Y}_{3D} = \mathbf{F}_r \mathbf{X} \mathbf{A}^T + e, \tag{5.9}$$

where $\mathbf{Y}_{3D} \in \mathbb{C}^{N \times M_x M_z}$ is the sample matrix with the N fast-time samples stacked per receiver, $\mathbf{X} \in \mathbb{C}^{G_r \times G_\theta G_\phi}$ is an unknown matrix containing the reflected power of the targets and encoding their angular and range position, and $\mathbf{F}_r$ is the inverse Fourier time shift matrix of size G_r with N selected Fourier coefficients. $\mathbf{F}_r$ is formed with the fast-time vectors $\mathbf{f}_r(r_k)$ defined in (2.11). A common approach to cast this two-matrix problem into a suitable form for CS is by the use of the Kronecker product $\otimes$ as:

$$\text{vec}(\mathbf{Y}_{3D}) = (\mathbf{A} \otimes \mathbf{F}_r)\text{vec}(\mathbf{X}) + e. \tag{5.10}$$

where vec$(\cdot)$ is a linear transformation of a matrix which converts the matrix into a vector. However, this problem is again unfeasible in both computational and memory complexity (e.g., following the same example of a 48×48 URA and 1° cell in both angular dimensions, 128 fast-time samples, and 256 range cells, $\mathbf{A} \otimes \mathbf{F}_r$ would be a 294912×8294400 complex matrix).

To solve this problem, the same approach of using partial inverse Fourier transform discussed in the previous sub-section is used, by reshaping $\mathbf{Y}_{3D}$ into a $N \times M_x \times M_z$ tensor and $\mathbf{X}$ into a $L_r \times L_\theta \times L_\phi$ tensor, and solving the optimization problem given by:

$$\min_{\mathbf{X}} ||\nabla \mathbf{X}||_1 + \frac{\mu}{2} ||\text{pIFFT3}(\mathbf{X}) - \mathbf{Y}_{3D}||_2. \tag{5.11}$$

where pIFFT3$(\cdot)$ is the partial inverse FFT operator, equivalent to applying the pIFFT$(\cdot)$ operator in each of the dimensions independently. Then, (5.11) can be solved efficiently

Table 5.1: Performance metrics derived from the PSF

		-3dB MLW ($\theta = 0°, \phi = 0°$)	-3dB MLW ($\theta = 60°, \phi = 60°$)	SLL
Angle	FFT	5.76°	9.39°	−13.18dB
	2D TVCS	2.75°	3.93°	−21.37dB
	3D TVCS	3.72°	5.60°	−18.50dB
Range	FFT	0.468m	0.468m	−13.15dB
	2D TVCS	0.468m	0.468m	−13.15dB
	3D TVCS	0.256m	0.256m	−21.9dB

using the TVAL3 algorithm, adapting it for the 3D case by computing the discrete gradient in the three dimensions and using the Lagrangian expansion defined in (5.7). Similarly to the 2D case, the default hyperparameters of TVAL3 indicated in [92] have been used, except for the hyperparameter μ, which has been tuned following a data-driven approach, resulting in a value of 2^{13}. Moreover, this work has adapted the algorithm by computing the 3D isotropic TV operator.

Before exploring the performance of the proposed method in complex extended targets, it is important to analyze the response for single scatter points, also because targets far from the radar may occupy a single angular cell. To this end, the point spread function (PSF) has been computed by placing a single scatter point at boresight with a distance of 10 meters and using a 20×20 URA. Figure 5.2 shows in the middle column two cuts of the PSF computed with the proposed 2D method, and in the left column, the PSF computed with a conventional Fourier beamformer. As can be seen, the main lobe width (MLW) in the azimuth-elevation plane is thinner in comparison with the FFT-based method, which will result in higher angular resolution. Moreover, the sidelobe level (SLL) of the proposed approach is lower than that of the FFT-based method without any windowing, which is desirable not to mask weak targets. On the other hand, since the range dimension is computed with a conventional Fourier transform, both approaches have similar performance in the range domain. When looking at the performance of the 3D version on the right column of Figure 5.2, it can be seen how the main lobe in the range dimension is now reduced while at the same time maintaining the suppression of the range sidelobes.

A comparison of the PSF performance metrics of the two proposed TVCS methods and conventional Fourier beamformer can be seen in Table 5.1. For the 2D version, the MLW is reduced by 3° and the SLL is reduced by 4dB, for targets located at boresight as well as at $\theta = 60°, \phi = 60°$. For the 3D case, the MLW is reduced, and the SLL is suppressed in both angle and range with respect to the Fourier estimation, maintaining the scanning capabilities. However, the angular MLW is larger than in the 2D case at the expense of reducing the range MLW.

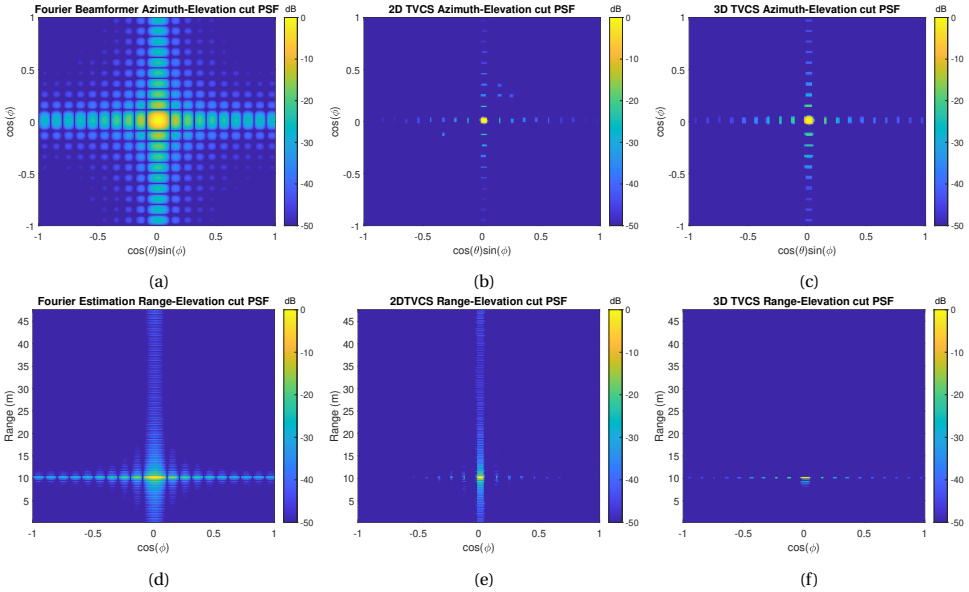


Figure 5.2: Point Spread Function of the conventional Fourier beamformer and of the two TVCS proposed methods using a 20×20 array. As can be seen in (a) and (d), the Fourier beamformer presents higher sidelobes and a wider main beam compared to the other plots. In (b) and (e) the PSF of the 2D TVCS method is shown, where the angular mainlobe is reduced and the sidelobes are suppressed, maintaining the same properties in the range domain. Finally, (c) and (f) show the PSF when the proposed 3D TVCS is applied. In this case, the range is also improved at the cost of minor degradation in the angular domain.

5.4. DATA GENERATION

As explained in the previous section, the aim of the proposed methods is the DoA estimation of extended targets in the context of automotive radar; therefore, a target simulator has been developed to evaluate the algorithms' performance for complex targets. In the literature, many researchers have studied the properties of automotive targets at the W band [96–98], and some have proposed models to simulate them. However, this thesis does not aim to design a realistic and precise extended target simulator but to develop a tool to validate the proposed TVCS algorithms. Therefore, the simulator has been developed accounting for a trade-off between computational complexity and electromagnetic fidelity, and for this reason, full electromagnetic simulations have been avoided. Moreover, multi-path, second-order scattering mechanisms, and different material properties have been ignored. Two different simulation versions have been developed to evaluate the proposed method's performance. The first one only simulates simple shapes in the azimuth-elevation plane, suitable for the 2D version of the algorithm, while the second one simulates more complex targets with realistic shapes.

5.4.1. 2D SIMULATION WITH SIMPLE SHAPES

For the first method, several extended targets with different shapes were simulated as a cloud of scatter points with the same radar cross-section (RCS). Examples of these tar-

gets can be seen in Figure 5.3. The scatter points have been spaced 1° in both azimuth and elevation and have been randomly shifted by ±0.25° to avoid the so-called 'inversion crime' (i.e., the points lie precisely in the discretized grid). Also, the convex haul, which will be used as the target boundary, can be seen in the black dotted line. Then, a 20Tx-20Rx MIMO radar system that yields a 20×20 virtual URA with half-wavelength separation between elements is simulated. The signal reflected by each scatter point is computed and aggregated in each virtual element to generate the $\mathbf{y}$ sample vector.

5.4.2. 3D SIMULATION WITH CAD MODELS

Computer-aided design (CAD) models from the 3D shape categorization benchmark (3DSCB) [99–103] dataset have been used to generate the target models, considering vehicles, pedestrians, and bicycles. Figure 5.4 shows some examples of the shapes from the dataset.

Once the appropriate CAD models have been selected, a scatterer point is placed at each of the vertices, setting the same RCS for every point. Therefore, after superposition of the contributions from each scatterer, the RCS of the whole object is not uniformly distributed in space, but will be higher where more vertices are present. Higher vertex density means essentially more complex shapes where the EM waves can backscatter. However, once the target is placed in a scene with a given position and orientation with respect to the line of sight of the radar, not all its scatter points will be visible due to self-occlusion. For this reason, the hidden point removal operator [104] is used to compute which scatter points should be included in the simulation and contribute to the received radar signal. An example of the visibility of a car model located at 8 meters from the radar with two different orientations can be seen in Figure 5.5. As shown, different numbers of scatter points are seen by the radar depending on the aspect angle of the target and its orientation. With this approach, the perceived RCS will also depend on the perspective.

An automotive radar with a virtual URA of 100×20 elements spaced half-wavelength has been simulated, with isotropic antennas for the $y > 0$ half-space, and the complex response from the targets has been computed as the superposition of the scattering from the visible scatter points on the target model. To this end, a Monte Carlo simulation was carried out where the response from 1680 targets was calculated, varying their position and aspect angle, and different CAD models were used for each target type. Moreover, the SNR of each simulation was chosen between -10dB and 20 dB, and it was computed

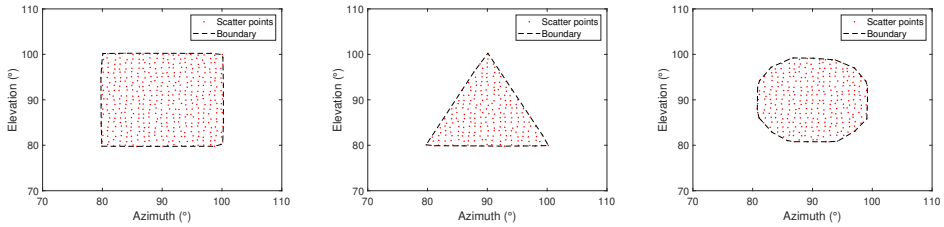


Figure 5.3: Examples of simulated extended targets as a cloud of scatter points. In black dotted line the convex haul, which will be used as the target boundary

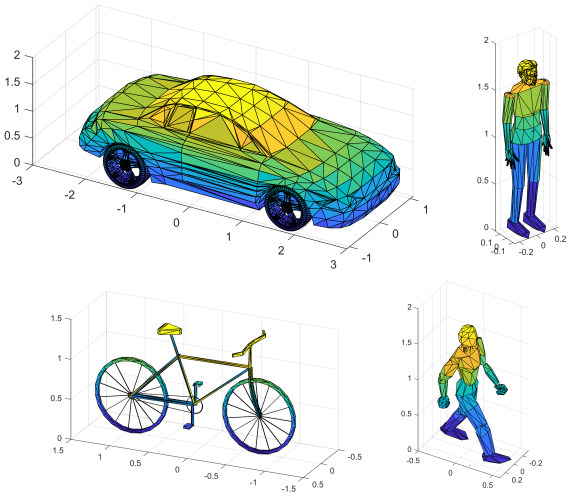


Figure 5.4: Examples of different CAD models used for the generation of synthetic data. Three different categories relevant to automotive have been selected: cars, pedestrians, and bicycles, and two types of pedestrians are shown as examples (standing and walking).

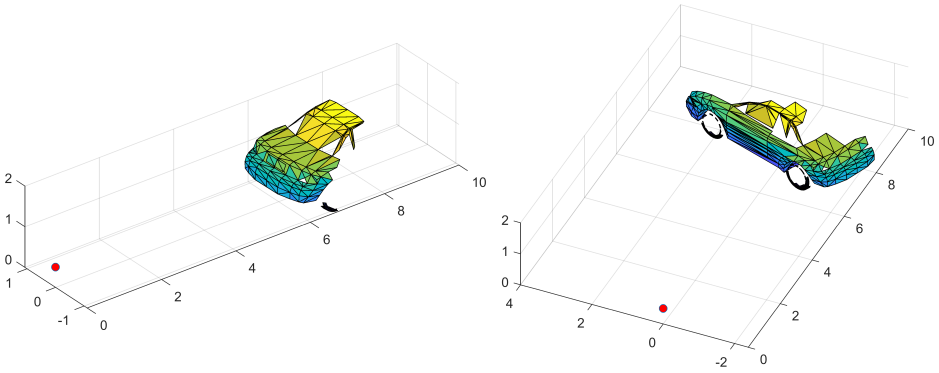


Figure 5.5: The same CAD model of the car is seen from two different points of view, where the red point represents the position of the radar. As can be seen, the number of visible vertices and, therefore, the related perceived RCS depends on the perspective.

Table 5.2: Simulation Parameters

Radar Parameters	**Value**
Center Frequency (GHz)	77
Bandwidth (MHz)	200
Array Topology	100×20 URA (Azimuth×Elevation)
Element separation	λ / 2
Height from ground (cm)	30
Scene Parameters	**Value**
Number of scenes	1680
Range of targets (m)	$X \sim \mathcal{U}(5,100)$
Azimuth of targets (deg)	$X \sim \mathcal{U}(-70,70)$
Orientation of targets (deg)	$X \sim \mathcal{U}(0,360)$
Scene SNR (dB)	$X \sim (-10,-5,0,5,10,15,20)$
Type of target	Car, Pedestrian, Bicycle
Models per target type	8

in the time domain before any processing gain. The parameters of the simulation can be seen in Table 4.1.

5.5. RESULTS

In this section, results using both simulation strategies are presented. First, the 2D method will be applied to the simple shapes, and then, the results using the CAD models will be shown.

5.5.1. 2D RESULTS

After simulating the baseband signal **y** received by a MIMO radar described in (2.13) using the point scatterers given by the extended targets, two methods are computed: a conventional delay and sum (DaS) beamformer and the proposed 2D TVCS estimation. A visual example of both angle estimations, with a zoom in the region of interest, can be seen in Figure 5.6. The result yielded by the TVAL3 algorithm can be seen in the rightmost column of Figure 5.6. The default TVAL3 parameters proposed by the authors have been used, except μ, which has been lowered to 2^4 to guarantee a good performance in the cases with a high noise level. By visual inspection, it is clear that the reconstruction using TVAL3 represents better the shape of the targets, as well as having lower sidelobes. These two aspects are the key metrics to assess the algorithm's performance.

In order to evaluate the similarity between the real shape and the reconstructed shape of the target, a shape must be extracted from the azimuth-elevation matrix. In this paper, the boundary at the -10dB drop from the maximum value in each image has been chosen to define the reconstructed shape. It is not in the scope of the article to discuss an appropriate value for this definition of the shape, and changing this -10 dB threshold does not change the results from a qualitative point of view. An example of these

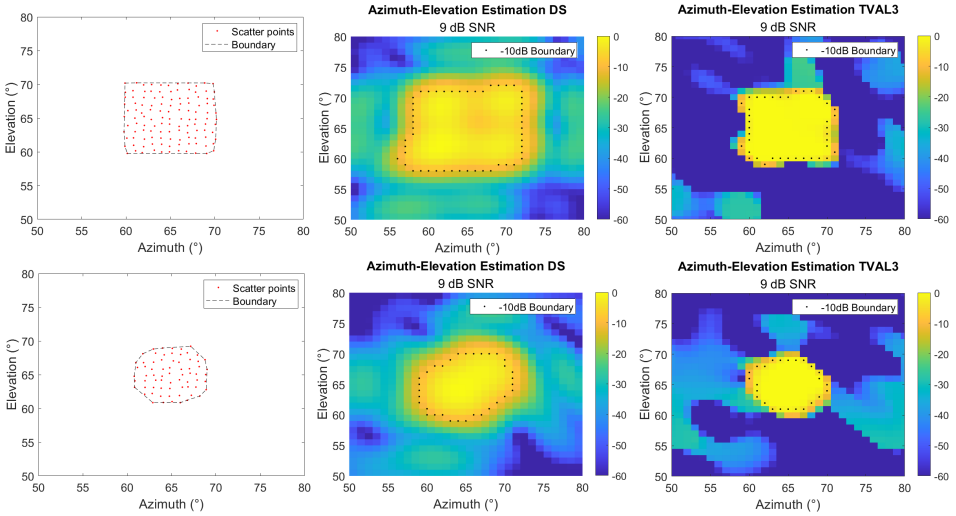


Figure 5.6: On the leftmost column, the cloud scatter points simulate distributed targets. In the middle, the reconstructed azimuth-elevation matrix using a conventional DaS beamformer. On the rightmost column the output of the TVAL3 algorithm is shown. The simulations assumed 9dB of SNR.

boundaries can be seen as the dotted black line in the middle and rightmost columns of Figure 5.6. Moreover, Figure 5.7 shows an example of the real and reconstructed shapes as colored areas for one realization of a rectangular target. Once the target shape has been determined given the aforementioned threshold, the Jaccard index, also named intersection over union (IoU), is computed between the real and estimated shape. The IoU metric is equal to 1 when two areas perfectly overlap and goes to 0 when there is no union between them. A Monte Carlo simulation has been implemented, modifying the target shape, position, and SNR. 100 trials were simulated using three shapes: rectangular, triangular, and circular. The position of the targets has been uniformly sampled in ±60°for both angles. The results aggregated for each SNR are presented in Figure 5.8a. It is important to note that the SNR refers to the ratio before the angular processing, i.e., before the DS gain or TVAL3 gain. It can be seen that the IoU is higher in the TVAL3 case for every SNR and only starts decaying for rather low values. This appears to indicate that it is a robust method against noise.

However, the IoU metric only provides an indication of how well the shape is reconstructed but does not consider how the energy is spread into secondary lobes. This is important not only because of its influence on the reconstruction quality but also because it could prevent weak target detection. For this reason, a pseudo integrated side lobe level ratio (P-ISLR) is computed, where all the energy inside the real target boundary is considered as the main lobe, and all the energy outside is considered as sidelobes. Again, this test is repeated in a Monte Carlo fashion for different target shapes and locations. The results for different SNRs are shown in Figure 5.8b. It can be seen how the TVAL3 reconstruction outperforms significantly in terms of P-ISLR.

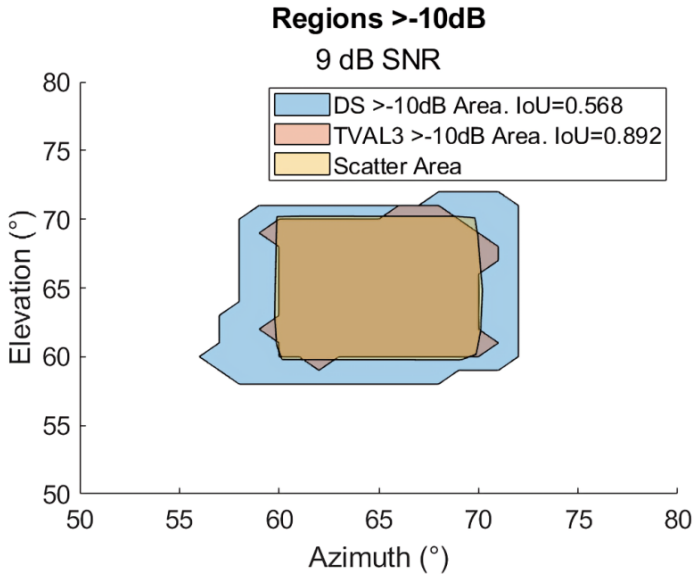


Figure 5.7: Original and reconstructed shapes using a DS beamformer and TVAL3. In this case, a rectangular target has been simulated. It can be seen how the TVAL3 reconstructs the target's shape better, yielding an IoU of 0.892.

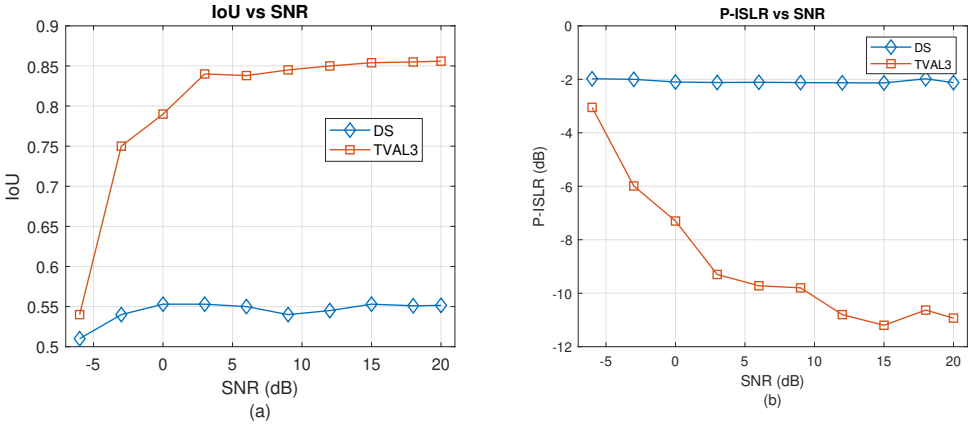


Figure 5.8: In (a) the IoU vs. SNR for different targets' shape and locations. In (b) the P-ISLR vs. SNR also for different targets' shape and locations.

5.5.2. 3D RESULTS

As explained in the section 5.3, the proposed TVCS methods aim to better estimate the 3D shape of extended targets in range, azimuth, and elevation. Moreover, the edges of objects are reconstructed in a sharper manner, and thus, it is expected that classification methods based on neural networks such as the one proposed in [35] will have higher accuracy when used in combination with the proposed methods. After processing the simulated baseband signals to obtain the range-azimuth-elevation radar cube, the target's shape is estimated to compare it with the ground truth shape. A fixed -20 dB threshold detector is applied to the result of all three methods, i.e., proposed 2D TVCS, proposed 3D TVCS, and a Fourier-based estimation. This allows us to filter out the noise and obtain only those cells where the target is present. Moreover, a fourth method based on ℓ_1 minimization with quadratic constraints is used for comparison. The ℓ_1-*magic* package available at [105] is used for the implementation, which uses a log-barrier algorithm in the optimization algorithm. It should be noted that a Hann window is applied before the Fourier-based estimation so that the sidelobe level is below -20dB. An example of the estimated point clouds with each method can be seen in Figure 5.9a for a car and in Figure 5.10a for a walking pedestrian. The subplots (a) show where the scatter points are placed, as explained in the previous section. As can be seen, they are extended targets composed of multiple scatter points, occupying more than one resolution cell. The subplots (b) show the result after conventional Fourier processing, and it can be seen how almost no detailed shape information is preserved, with all the points clustered in a single large shape. However, both proposed methods reconstruct more details. For example, in Figure 5.9c and Figure 5.9d, it can be seen that the middle part of the car has a higher height, and the two wheels are visible. On the other hand, Figure 5.10c and Figure 5.10d show how the head and some limbs of the walking pedestrian are separated from the main body.

However, it is important to provide a quantitative metric to assess the shape estimation performance of each method. In a 2D case, this can be easily done by generating the bounding box of the 2D point cloud and then computing the IoU between the estimated and the ground truth surfaces. However, two challenges are presented when the problem is formulated in three dimensions. First, automatically generating a tight bounding box around the point cloud is not trivial since most available algorithms generate a cuboid that does not capture the realistic volumetric details of objects of interest. Second, computing the IoU on a 3D continuous space is only solved for oriented cuboids, and calculating it for complex, arbitrary 3D volumes is still an open problem.

For this reason, this work uses an alternative, simplified approach where the 3D range-azimuth-elevation space is discretized into voxels. A voxel is then occupied if at least one point of the point cloud generated by the different algorithms under evaluation is inside the voxel. This process is known as *voxelization* and allows for a more straightforward computation of the IoU, comparing the occupied voxels between the estimation and the ground truth. It is also the approach routinely followed by the latest research in autonomous driving using computer vision and Lidar data [106, 107]. The IoU will have a value of 1 if both volumes perfectly overlap and 0 when the two volumes are completely disjoint.

In this work, the voxel size has been selected as half the resolution cell in each di-

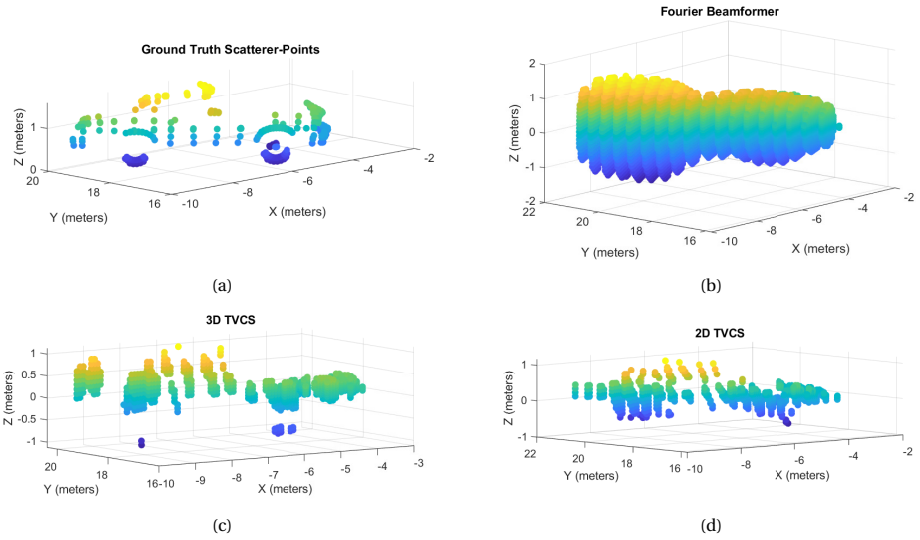


Figure 5.9: In (a) the scatterer-points of a car; in (b) the estimated point cloud with a Fourier beamformer after a fixed threshold; in (c) the estimated point cloud with the 3D TVCS method; in (d) the estimated point cloud with the 2D TVCS. As it can be seen, the Fourier beamformer generates a single large shape without details, whereas in the 3D TVCS and 2D TVCS results, more details are preserved: for instance, it can be seen that the middle part of the vehicle is higher and the wheels are visible

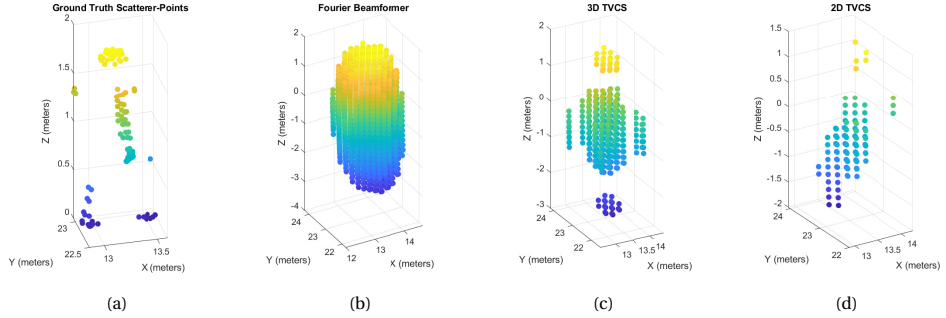


Figure 5.10: In (a) the scatterer-points of a pedestrian; in (b) the estimated point cloud with a Fourier beamformer after a fixed threshold; in (c) the estimated point cloud with the 3D TVCS method; in (d) the estimated point cloud with the 2D TVCS. As it can be seen, the Fourier beamformer generates a single large shape without details, whereas in the 3D TVCS and 2D TVCS more details can be appreciated.

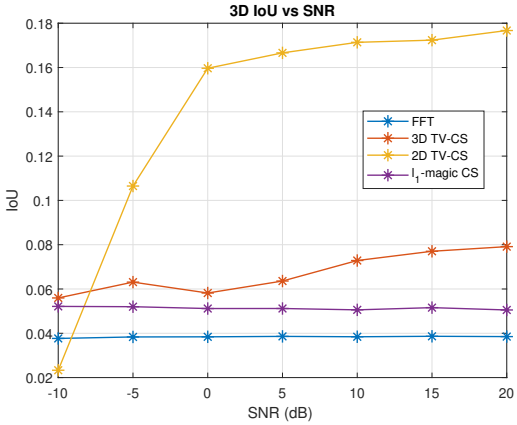
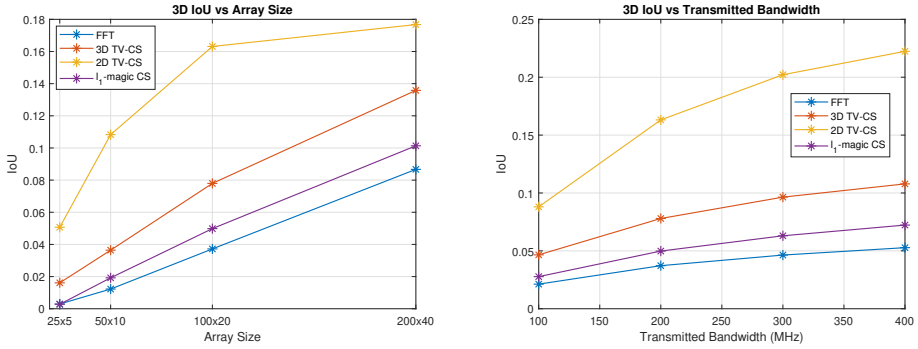


Figure 5.11: Average 3D IoU for each compared method as a function of SNR. As can be seen, both proposed methods outperform the standard CS and FFT-based methods. The 2D version of the proposed algorithm has a better performance under higher SNR conditions.

mension to minimize the quantization errors. The average IoU over the whole dataset has been computed for the four methods, and the results can be seen in Figure 5.11 as a function of the SNR. As can be seen, both proposed methods outperform the Fourier-based and the standard ℓ_1 minimization approach. It is important to note that in the simulation, single extended targets are considered in the scene, and therefore, the standard ℓ_1-norm method is stable. This may not be the case in a realistic automotive scene with additional targets. On the other hand, it is also possible to observe that the 2D version of the proposed algorithm yields a higher IoU than the joint 3D version. From the PSF analysis, it was noted that the angular spread was lower in the 2D case than in the 3D case. Since the range resolution is already high, improving the angular resolution appears to have a higher impact on the overall IoU. Moreover, it is possible to observe that the impact of the noise is higher in the 2D version. A possible explanation for this phenomenon can be found in the high difference in the dimensionality of the estimation space between the 2D and 3D problem formulations. Since the estimation space of the 3D problem is two orders of magnitude larger, also the sparsity level of the estimation space is two orders of magnitude higher. The hyperparameter μ of the TVAL3 algorithm is supposed to be tuned using both the sparsity level and the noise level in the observations, but in practice, it is not possible to know these values. Therefore, μ has been kept constant for a fair comparison, but its impact on the 2D version is higher, making the algorithm under-perform in low SNR conditions.

It is also interesting to analyze the algorithms' performance when the resolution of the system changes. To this end, two independent analyses have been performed, where the angular and range resolutions are analyzed separately. A constant SNR of +10dB has been used for both tests. For the first analysis, the system's aperture has been modified by changing the number of elements in the array, from 25 to 200 elements on the x-axis and from 5 to 40 on the z-axis. Moreover, the voxel size has been kept constant, ensuring it is equal to half the smaller resolution cell. The results averaged over the whole dataset



(a) (b)

Figure 5.12: Average 3D IoU for each compared method at +10dB SNR as a function of array size (a) and as a function of signal bandwidth (b). In both cases, it can be seen that the proposed methods outperform the standard CS and FFT-based methods, and the performance increase is higher with larger apertures.

5

can be seen in Figure 5.12a. As expected, the results show that the IoU is higher when a larger array is used due to increased angular resolution and, thus, the capability to reconstruct more details of the objects. Moreover, it can be seen how the difference between both proposed methods remains quite constant, while the improvement against the other methods rises. With a larger array (i.e., smaller angular resolution cells), the targets will spread over more angular cells even at long ranges. Therefore, the proposed two methods specifically designed for extended targets will perform better.

Finally, the impact of the range resolution on the algorithms' performance has been analyzed. The transmitted bandwidth has been changed from 100 to 400 MHz (i.e., half and double the baseline bandwidth, respectively) while keeping the original array size. Figure 5.12b shows the results for the different bandwidth values.

5.6. CONCLUSIONS

Two compressive sensing based methods to estimate the 3D range-azimuth-elevation space of extended targets with MIMO radars have been developed. Based on the total cariation regularization to enforce sparsity in the signal's gradient, these methods are specifically designed for the next generation of automotive radars, where the targets will be extended into multiple resolution cells in the cross-range due to the expected improvements in angular resolution. The first proposed method performs an independent 2D azimuth-elevation estimation at each range cell, whereas the second proposed method performs a complete joint 3D range-azimuth-elevation estimation. Extensive radar simulations using CAD models of typical targets for automotive applications have been developed to validate the algorithms in terms of their shape reconstruction capabilities. Results show an improvement in the shape reconstruction for both methods with respect to the standard Fourier estimation and the standard CS based on ℓ_1 minimization with quadratic constraints. The 2D version of the proposed approach (2D TVCS), where the range is estimated with a Fourier transform, outperforms the 3D ver-

sion (3D TVCS) on higher SNR conditions, but it is more sensitive to the tuning of the algorithm hyperparameter μ. Moreover, it has been proved that the proposed methods significantly increase performance when using high-resolution radars with larger array apertures, due to targets' signatures spreading over more resolution cells. Future work will assess the impact on classification accuracy by using the presented methods.

6

DEEP LEARNING DETECTOR AND RADELFT DATASET

In the original plan, this chapter was supposed to be about classification (and in reality, the whole thesis was supposed to be just about classification). However, when we finally mounted the radar on the car, solved the many technical and non-technical issues involved, collected lots of data, and started to process it, I realized that I had to implement a detector (in the 'radar sense' of the word) before diving into classification. The problem was that designing a suitable detector for the heterogeneous types of targets and scenes was not easy. What worked in some situations did not work in others, and tuning the hyperparameters of the available detectors for all the situations was impossible. The content of this chapter is about designing a detector where there is no need to tune its parameters for each specific target or scene, leveraging on artificial intelligence.

Also, this chapter is about data. Unlike the computer vision community, where many public datasets are available, only some are available for radar practitioners (although this is rapidly changing). There are many reasons for this, from the costly data collection process in terms of time and finances, to the little reward of sharing such data. To do my bit in changing things, I have shared all the data we collected, including radar, camera, lidar, odometry, and the code to use it.

Parts of this chapter have been published in:

I. Roldan, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli and A. Yarovoy, "See Further Than CFAR: a Data-Driven Radar Detector Trained by Lidar," 2024 IEEE Radar Conference (RadarConf24), Denver, CO, USA, 2024, pp. 1-6.

I. Roldan, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli and A. Yarovoy, "A Deep Automotive Radar Detector using the RaDelft Dataset," in IEEE Transactions on Radar Systems.

6.1. INTRODUCTION

Previous chapters have dealt with the angular estimation problem and how to generate a high resolution radar cube. This cube contains not only returns from targets, but also noise. The question this chapter tries to answer is how to effectively select only those cells containing target and drop those containing noise, i.e., how to design a detector for high resolution automotive radar data. The most common detectors are the well-know window-based constant false alarm rate (CFAR) detectors. While CFAR detectors have proven optimal in other environments [9], their application in the dynamic and unpredictable conditions of road traffic scenarios suffers from poor performance [10, 11]. Namely, they are designed to maintain a constant rate of false alarms amidst varying clutter, but they struggle to adapt to the rapidly changing environments typical of roadways. Complications such as non-uniform clutter (or the lack of reliable clutter models for this task), target masking, and shadowing can significantly reduce the effectiveness of CFAR detectors in automotive radar settings. Additionally, CFAR detectors are constrained by a fundamental limitation: they typically assume a fixed, expected target size based on predefined guard and training cell hyperparameters. However, in an automotive context, this assumption is problematic as the size of potential targets can widely vary, ranging from medium-sized objects such as pedestrians to large vehicles like trucks or buses. Moreover, the perceived size of these targets in the radar's angular dimension changes with distance. Large objects occupying multiple cells at close range can appear as simpler point-like targets at further distances. This relationship between angular target size and distance adds another layer of complexity to using CFAR detectors in automotive radar, necessitating alternative solutions to accurately detect and classify objects under varying road conditions.

To address these limitations, a new data-driven radar target detector using a unique cross-sensor supervision pipeline has been developed. The proposed data-driven detector is initially trained with synchronized radar and lidar data together and can subsequently generate denser point clouds using only raw data from a high-resolution automotive radar. The proposed approach is extensively validated using the *RaDelft* dataset. This dataset is a large-scale, real-world multi-sensory dataset recorded in various driving scenarios in the city of Delft, collected as part of this thesis and shared publicly. This chapter then introduces two contributions: the recording & sharing of the *RaDelft* dataset and the proposed data-driven detector. Therefore, two related work subsections are included to review the state-of-the-art and highlight the need for new radar datasets and new detection algorithms.

6.1.1. RADAR DATASETS RELATED WORK

Several automotive radar datasets have recently been published for different tasks, covering many of the processing steps listed in Section 2.3. However, most of them are unsuitable or, at the very least, limited for radar practitioners since the data is already processed, often to the point cloud level. Thus, it is impossible to apply signal processing algorithms that operate on lower-level data. Some datasets also provide the radar cube data, but few give the raw ADC data needed to test advanced signal processing methods. Essentially, each already-performed processing step limits the scope of the research that can be performed with that data. On the other hand, this simplifies the steps needed to

make it suitable for other subsequent tasks.

A recent summary of the available automotive radar datasets can be found in [108]. Nevertheless, in this section, only those datasets that provide data before the point cloud level of processing are considered since they are the most useful for radar practitioners. Table 6.1 summarizes such datasets. As can be seen, most of these available datasets are recorded with automotive radars with linear antenna arrays, meaning that there is only azimuth resolution and no information about the elevation of targets. While useful for some tasks, this type of data is not representative of the data of next-generation 4D radars that are becoming the standard in the automotive field. On the other hand, some datasets already include a 4D imaging radar [27, 109, 110]. The RADial [109] dataset provides ADC data level suitable for radar practitioners, but the array topology used is not public, and thus, advanced array processing methods cannot be applied. The ColorRadar [110] dataset uses a commercially available radar, therefore its datasheet is public. However, most of the scenes are recorded indoors and without a vehicle. Moreover, camera information is not provided. Finally, the K-Radar [27] dataset is the most complete, providing range-azimuth-elevation-Doppler cubes, many auxiliary sensors, and useful code to parse the data. However, no ADC-level data is provided, which may limit the potential research scope of the dataset.

Considering the limitations of the aforementioned public datasets, this work presents a new dataset, *RaDelft*, aiming to close the gaps in the existing available datasets collected with a commercially available radar. Our dataset contains three different levels of data processing, namely ADC-level, radar cubes, and point clouds as defined in the previous sections, such that it can serve different future research directions. Additionally, the dataset contains synchronized data from camera, lidar, and odometry, recorded in real-world driving scenarios in the city of Delft. Additional details are provided in the following Section 6.2, specifically the sensors used and the developed radar signal processing pipeline.

Table 6.1: Available public datasets providing either ADC or pre-detection data. In the *DATA TYPE* column, R, D, A, E, and C stand for Range, Doppler, Azimuth, Elevation, and Channel, respectively, while PC means point cloud. In the *ARRAY TYPE* column, 'dense' means that all the half-wavelength spacing is filled with virtual elements. In the *OTHER SENSORS* column, C, L, and O stand for Camera, Lidar, and Odometry, respectively.

Name	**Data Type**	**Array Type**	**Virtual Aperture (x×z $\frac{\lambda}{2}$ spacing)**	**Other Sensors**	**Record Time (Radar Frames)**	**Potential Gaps**
Zendar[111]	RDC / PC	Dense ULA	4x1	CLO	478s (4780)	No elevation, no ADC data, small aperture.
Radiate [112]	RA	No array mechanical scanning	N/A	CLO	5h (44000)	No elevation, no Doppler, no ADC data
CARRADA [113]	RA / RD	Dense ULA	8x1	C	12.1m (12666)	No elevation, no ADC data, small aperture
RADet [114]	RAD	Dense ULA	8x1	C	1015s (10158)	No elevation, no ADC data, small aperture
CRUW [30]	RA	Dense ULA	8x1	C	3.5h (400000)	No elevation, no Doppler, no ADC data, small aperture
Radical [115]	ADC	Dense ULA	8x1	C	104m (189000)	No elevation, small aperture
SCORP [26]	ADC / RAD	Dense ULA	8x1	C	N/A (3913)	No elevation, small aperture
ColoRadar [110]	ADC / RAE PC	Sparse URA	86x7	LO	145m (43000)	No camera, mostly indoor
RADial [109]	ADC / RAD PC	N/A	N/A	CLO	2h (25000)	Unknow array topology
K-Radar [27]	RAED	NUA	N/A	CLO	N/A (35000)	No ADC data
RaDelft (Ours)	ADC / RAED PC	Sparse URA	86x7	CLO	35m (16975)	

6.1.2. RADAR DETECTORS RELATED WORK

The radar detection problem can be formulated as a binary decision task for each radar cube cell, whose objective is to determine whether there is a target or only noise in that specific cell. As mentioned in the previous sections, the automotive radar field has particular challenges when tackling the detection problem. First, the definition of clutter is not univocal in this application since targets of very different natures should be detected, including pedestrians, vehicles, bridges, potholes, road debris, and buildings, among others. Second, since modern automotive radars have high resolution in range, Doppler, and, to some extent, angle, targets occupy more than a single cell, behaving as extended targets. Finally, not only do the sizes of targets to be detected have a large variance, but also, for the same target, its perceived size can change over time. This is due to two different physical phenomena: the dependency of the angle estimation with its cosine with respect to the radar line of sight and the relationship of the Cartesian size of the cell with the range due to the angle. Due to all these reasons, conventional CFAR detectors are expected to perform poorly in automotive radar data [10, 11].

In the past years, several works have been published on detecting extended targets in radar data. Image-based detector techniques have been explored in the literature [116], but usually rely on high-contrast data where sharp transitions occur between noise and target. However, due to the finite length nature of signals, spectral leakage in the Fourier processing makes, in general, these transitions soft. Moreover, subspace detectors for extended targets in range and Doppler have been developed [117, 118], but still, an expected spread size of the target energy is needed, in addition to a high computational cost, making them unsuitable for real-time imaging automotive radars.

Also, DL techniques have been applied to the radar detection problem [10, 23–27]. In [23], a DL detector is proposed, outperforming several 2D CA-CFAR detectors, but only tested in simulated data. In [119] and [25], the authors propose a similar network structure using 3 autoencoders in three 2D projections (range-angle, range-Doppler, and angle-Doppler) using the annotated dataset in [30] by a camera, avoiding full 3D detection. However, using camera detections as ground truth may be limited due to the 2D nature of camera images. On the other hand, the works [27] and [10] propose two different NNs, but both use the lidar point cloud as ground truth. Since lidar provides high-resolution 3D point clouds, it seems a more reasonable choice to serve as ground truth. The proposed method in [10] uses a neural network to detect targets only in the range-Doppler dimensions, followed by the angle estimation and a spatio-temporal filter to enhance the resulting point cloud. On the other hand, in [27], a novel sparse approach to use an NN to detect in the range-azimuth-elevation space is presented. However, the Doppler information is collapsed into a single value, preventing the network from learning the possible angular estimation enhancement due to its relationship with Doppler [19, 20]. Moreover, only the top 10% power cells are used as input to the network, and therefore, a pre-detection step is used, which can potentially remove target cells. This may be critical in automotive scenarios, where the angular sidelobes of close-range targets maybe even 20dB higher than weakly-reflecting distant targets such as pedestrians.

Figure 6.1: Different frames of different scenes of the *RaDelft* dataset. As it can be seen, there are city center environments, suburban, and different road infrastructures such as large bridges.

6.2. RADELFT DATASET

The dataset was recorded with the demonstrator vehicle presented in [120] with an additional Texas Instrument MMWCAS-RF-EVM [50] imaging radar mounted on the roof at 1.5 meters from the ground. The details of the radar and the waveform used are provided in Section 6.2.1. The collection was performed driving in multiple real-life scenarios in the city of Delft with different scene characteristics, such as suburban, university campus, and Delft old-town locations. Figure 6.1 shows four different camera frames to illustrate the differences in the environments. The output of the following sensors was recorded: a RoboSense Ruby Plus Lidar (128 layers rotating lidar) and the imaging radar board installed on the roof, a video camera (1936 × 1216 px, ~30 Hz) mounted behind the windshield, and the ego vehicle's odometry (filtered combination of RTK GPS, IMU, and wheel odometry, ~100 Hz). The sensor setup can be seen in Figure 6.2. All sensors were jointly calibrated following [121] and time synchronized. With a 10 Hz frame rate, each scene contains around 2500 radar frames, adding to a total of 16975 frames.

Example code for loading and visualizing the data is provided in a repository[1] to facilitate the use of the dataset, which can be downloaded from [122]. Moreover, the radar data is specifically provided at different processing stages for researchers with different backgrounds and interests, including ADC data, radar cubes, and point clouds. The details of the radar processing applied to the data can be found in the next subsection.

6.2.1. RADAR CONFIGURATION AND PROCESSING

In terms of the specific details of the radar system, this is the MIMO FMCW evaluation board MMWCAS-RF-EVM from Texas Instruments, with 12 transmitters and 16 receivers [50]. The resulting virtual array is an 86-dense uniform linear array (ULA) in the X-direction (shown in Figure 6.2) with half-wavelength spacing, allowing azimuth estimation without grating lobes and a theoretical resolution of 1.33° looking at boresight. However, from the point of view of the 2D angular estimation problem in both azimuth & elevation, the resulting uniform rectangular array (URA) is very sparse, with only a few minimum redundancy arrays (MRA) in the Z-direction (shown in Figure 6.2). Thus, the elevation estimation is very poor in terms of both resolution and ambiguity. The details of the array topology can be found in [50] with graphical representations of the positions of all the elements. Moreover, some elements are overlapped, which can be used to address some of the problems introduced by using time division multiple access (TDMA) in transmission, as detailed later in this sub-section.

[1] https://github.com/RaDelft/RaDelft-Dataset

Figure 6.2: Vehicle used to collect the dataset presented in this section, equipped with a high-resolution radar, lidar, camera, and odometry. The radar is shown in the top-right inset, with the defined X and Z coordinate axes assumed in this work.

Table 6.2: Radar Waveform Parameters used in the Data Collection

Waveform Parameters	**Value**
Start Frequency (GHz)	76
Effective Bandwidth (MHz)	750
Chirp Slope (MHz/μs)	35
Chirps Length (μs)	28
Idle Time (μs)	5
Number ADC Samples per Chirp	256
Number of Chirps per Frame	128
Sampling Frequency (Msps)	12
Tx strategy	TDMA
Derived Quantities	**Value**
Range Resolution (m)	0.2
Maximum Unambiguous Range (m)	51.4
Velocity Resolution (m/s)	0.046
Maximum Unambiguous Velocity (without extension) (m/s)	2.48
Maximum Unambiguous Velocity (with extension) (m/s)	17.36

The radar waveform parameters used can be seen in Table 6.2, with the derived resolution and ambiguity values. The complex baseband samples are saved in the dataset using the same format provided by the radar manufacturer, but MATLAB code is provided to parse it, reshape it into a $N_{fast} \times N_{slow} \times N_{Vchan}$ 3D tensor, and process it to the radar cubes.

The first step of the processing is to apply a Hamming windowing and an FFT in the fast-time and slow-time dimensions to perform range and Doppler estimation. Then, the detrimental effects of the TDMA have to be compensated. The first effect is related to the extension of the pulse repetition interval (PRI) by a factor equal to the number of transmitters. Therefore, the maximum unambiguous Doppler and the corresponding maximum measurable velocity (without ambiguity) v_{max} is reduced, as can be seen in equation (6.1):

$$v_{max} = \frac{c}{4 f_c PRI}, \tag{6.1}$$

where c is the speed of light and f_c is the carrier frequency. This effect is especially problematic in the automotive context, where targets can have high relative speeds. Moreover, the phase difference between signals received from different transmitters will depend on both the angle of arrival of the signal and the velocity of the targets, due to the target's movement between transmission times of different transmitters operating in TDMA mode [123]. This resulting phase migration term is shown in equation (6.2):

$$\phi_{mig} = \frac{4\pi}{\lambda} v \Delta t, \tag{6.2}$$

where λ is the wavelength, v is the relative speed of the target, and Δt is the time difference between transmitters. This term must be compensated before performing angle estimation to avoid significant artifacts.

In this work, both undesirable effects of TDMA are solved using the overlapped virtual antennas present in the radar system with the algorithms provided in [124]. However, it is important to take into account that the maximum unambiguous velocity extension only works when a single target is present in a range-Doppler cell. Therefore, if multiple targets are folded into the same Doppler bin, or there are targets in different angles at the same range-Doppler bin, the algorithm will not be able to address the problem. Since this work does not aim to solve the Doppler ambiguity problem in TDMA, the aforementioned constraint is accepted as a limitation of the current commercial radar system. Nevertheless, it is assumed that making the ADC samples directly available in our dataset can be valuable for the research community, for example, applying more advanced approaches for Doppler/velocity ambiguity in TDMA in the future.

The angle estimation can be performed once the TDMA effects have been compensated. It is important to remember that the resulting virtual array is a very sparse URA with some structures. While other research works deal with this type of array, for instance by trying to fill/interpolate the missing elements or applying compressive sensing techniques [17, 125], the core of this work is not to improve the angular estimation with sparse arrays. Therefore, a very simple approach of zero-filling and FFT processing has been applied. However, due to the sparseness of the radar antenna array in the Z-direction (shown in Figure 6.2), grating lobes and high-side lobes appear in elevation. To

mitigate this problem, the field of view in elevation has been restricted to ±15° degrees, and the elevation value with the highest power has been selected and saved, discarding the rest. Also, the azimuth estimation has been restricted to ±70° for two reasons. First, the angular estimation performance outside this region is rather poor, as:

$$\Delta\theta \propto \frac{1}{\cos\theta}, \tag{6.3}$$

being $\Delta\theta$ the angular resolution and θ the estimated angle. Secondly, the radiation power is almost 10dB lower than at boresight outside this region, making target detection very challenging.

Subsequently, after zero-padding, FFT processing and field of view (FoV) cropping, the resulting radar cubes have dimensions $N_r \times N_D \times N_a \times 2$ ($500 \times 128 \times 240 \times 2$). This essentially means that for each range-Doppler-azimuth cell, there are two values: the elevation value with the highest detected power level, and the power level itself. Note that the 240 azimuth bins span the ±70° degrees of the FoV after cropping, but not uniformly, due to the non-linear relation in the equation (6.3). For simplicity and to save storage space in the shared dataset, the aforementioned values are saved as different cubes since the elevation can be stored as an integer number (i.e., denoted as elevation bin), while the power value is a float.

Finally, a detection stage is applied to the radar cubes to generate a point cloud. This lower dimensionality representation of the data is also provided within the shared dataset to ease the process for researchers who want to use this highly processed data straightforwardly without going into the details of radar signal processing.

6.3. PROPOSED DATA DRIVEN DETECTOR

To address the aforementioned shortcomings of current detectors in automotive radar, a novel data-driven detector is proposed to generate 3D occupancy grids only with radar data, using neural networks and lidar data as ground truth. A visual summary of the method can be seen in Figure 6.3.

The first step of the method is to adapt the lidar point cloud to serve as the ground truth. For each radar cube, the closest lidar point cloud in time is selected based on the timestamps for both radar and lidar data, assuming that a small error due to different start times may be present. Since the lidar system used in this work is mechanically rotating, it provides 360° coverage. Therefore, the first step is to crop this to the same FoV of the radar, i.e., ±15° in elevation, ±70° in azimuth, and a maximum range of 50m. To illustrate this difference in the FoV, Figure 6.4b shows the cropped lidar point cloud compared to the original point cloud in Figure 6.4a. Moreover, removing all the lidar points from the road surface is essential as the road surface is hardly visible to the radars and could lead to noisy ground truth for the training process. The Patchwork++ algorithm is used to this end [126]. After removing the road surface points, the resulting lidar point cloud can be seen in Figure 6.4c.

Finally, the processed lidar point cloud has to be converted into a 3D cube to serve as ground truth. This voxelization process can be understood as generating a 3D occupancy grid, where each voxel contains 'one' if at least one lidar point is inside and 'zero' otherwise. However, it is important to note that the radar cube grid is not uniform due

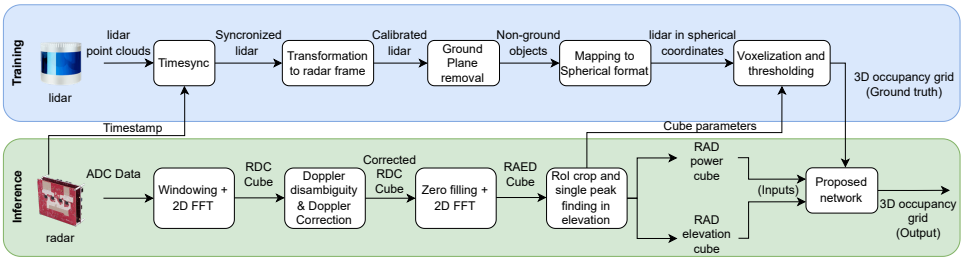
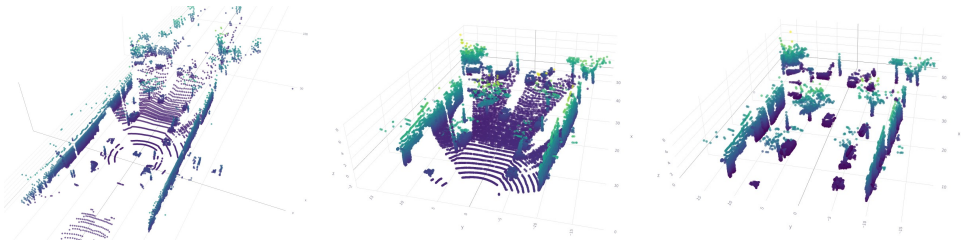


Figure 6.3: Overview of the proposed data-driven detector. The steps to generate the 3D lidar occupancy grid are shown on the top row, which will then be used as ground truth for training the neural network. The radar signal processing pipeline is shown at the bottom of the figure and is needed to generate the input data for the network. RDC stands for range-Doppler-channel (no angle estimation), and RAED stands for range-azimuth-elevation-Doppler [127].



(a) (b) (c)

Figure 6.4: In (a) the original point cloud as provided by the lidar sensor. In (b) the lidar point cloud after cropping to mimic the radar field of view (i.e., $\pm 70^\circ$ in azimuth and $\pm 15^\circ$ in elevation). In (c), the lidar point cloud after the road surface removal using PatchWork++ [126], which will be used as ground truth to train the proposed data-driven detector.

to the Fourier Transform processing for angular estimation and its relationship with the cosine of the estimated angle. This effect, which essentially makes the cells thinner at boresight and broader at the edge of the field of view, must be considered to generate the same non-uniform lidar 3D occupancy grid. This process of transforming the lidar point cloud into the 3D occupancy grid effectively reduces the resolution of the lidar since a grid cell may contain many lidar points. The resulting 3D occupancy grid matches the dimensions of the radar cube, and thus, have the same resolution. It is important to notice that all this process can be performed offline, outside the NN training loop

Once the ground truth has been appropriately generated, as described above, the NN can be trained. The proposed NN is an evolution of the previous model validated in [127]. Specifically, in this case, the network is modified to use three data frames as input to model temporal patterns, and the NN simultaneously predicts the 3D occupancy grid for the three frames. This modification has been implemented to reduce the 'flickering' usually present in the radar point clouds, where isolated points pass the detection threshold due to instantaneous high noise but disappear in consecutive frames. Therefore, the proposed NN tries to enforce some temporal consistency. A diagram of the complete network architecture is shown in Figure 6.5. As it can be seen, the input is

a T × 2 × R × A × D tensor, where in practice, T = 3 (frames), R = 500 (range bins), A = 240 (azimuth bins), and D = 128 (Doppler bins). As explained in the previous section, these values are higher than the initial number of fast-time samples, slow-time samples, and virtual channels due to zero padding applied before the FFT processing. Moreover, the number of frames T = 3 has been chosen as a trade-off between managing to capture temporal information and losing useful correlation between frames since the scene is often not static, and including too many frames will result in inconsistencies.

In terms of architecture, the first part of the proposed NN is the *DopplerEncoder* subnetwork. As the lidar cannot measure Doppler information, the detections on the Doppler dimension of the radar data cannot be directly utilized and compared to the ground truth. However, there is a known relationship between Doppler and angle in the case of moving platforms (or moving targets). Thus, the Doppler dimension is not simply removed from the radar data but rather encoded so that it can still be used in the overall detection process, as it may be beneficial for angular estimation. Specifically, here the *DopplerEncoder* subnetwork extracts all the Doppler information in each range-azimuth cell and encodes it into the channel dimension. This is achieved by using two 3D convolutional layers followed by a 3D max pool layer, transforming the 2 × R × A × D input tensor into a 64 × R × A tensor, where the 64 channel dimension contains the encoded information of Doppler and elevation.

The second part of the proposed network is an off-the-shelf 2D CNN *backbone*, applied to estimate the final R × A × E (500 × 240 × 44) 3D occupancy grid. The significant advantage of using such 2D CNN backbones is their compatibility with hardware accelerators (e.g., GPUs and TPUs) and major machine learning frameworks (e.g., TensorFlow, PyTorch), leading to enhanced computational efficiency. While the current proposed implementation employs a feature pyramidal network (FPN) [128] with a Resnet18 backbone [129], our modular design allows for different architectures to be used for this purpose, enabling the system to be tailored to the specific memory and computational requirements of the intended platform.

These two parts of the proposed network are applied to each of the three considered frames independently, as shown in Figure 6.5, but the weights of the layers are shared, and the output is concatenated into a T×R×A×E (3×500×240×44) tensor. Finally, to take into account the temporal relationship between the three frames, a third module composed of six 3D convolutional layers is included (referred to as *TemporalCoherenceNetwork* in Figure 6.5). It is important to notice that even if the output is a 3D occupancy grid for each frame, the power information on each cell is not lost, since the indices of the detected cells from such grid can be used to retrieve the corresponding intensity information from the original RAED cube.

One of the key characteristics of the radar data is the scene sparsity. Of all the voxels in the generated 3D occupancy grid, only around 1% contain targets. Therefore, this must be considered when selecting the loss function for training the neural network. In this work, the focal loss [130] is used for this purpose, which handles class imbalances in a similar way to the weighted cross-entropy loss, and adds an extra modulating factor to focus on the hard cases. The focal loss [130] is defined as:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t), \tag{6.4}$$

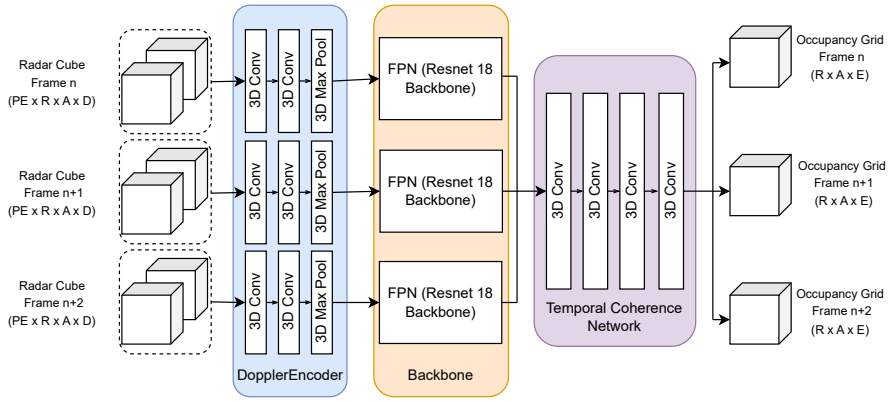


Figure 6.5: Proposed network architecture for the data-driven detector composed of three sub-networks. First, the *DopplerEncoder* network aims to encode the Doppler information so that its information is retained even if not directly comparable with ground truth lidar data. Then, a standard FPN with a Resnet *backbone* is used. Note that the three branches process three frames of data separately but share the same weights. Finally, the three outputs are concatenated to produce an input tensor to the *temporal coherence network*, which generates the final occupancy grid for each frame.

with

$$p_t = \begin{cases} p & \text{if} \quad y = 1 \\ 1-p & \text{otherwise} \end{cases} \tag{6.5}$$

where $y \in \{\pm 1\}$ is the ground-truth class (i.e., detection or not), α_t the weighting factor to take into account data imbalance defined as $\alpha \in [0,1]$ for class 1 and $1-\alpha$ for class -1, and $\gamma > 1$ is the focusing factor. This loss is especially interesting in radar data, since high RCS targets can be easily detected, but low RCS targets or targets located at a far distance are more challenging to detect, and this can be taken into account by the γ parameter.

In terms of training-testing split, 90% of the data from five of the seven recorded scenarios have been used to train the network using Adam optimizer, leaving 10% for validation. The network was trained using the DelftBlue Supercomputer [131] from TU Delft. The remaining two recorded scenarios are used as a test set, i.e., with data completely new and unseen for the network.

6.4. RESULTS

The trained neural network can estimate the 3D occupancy grid for each radar cube and thus act as a detector. It should be noted that all the results presented in this section are evaluated using only the test set, composed of the 2 scenes left out from the training process. This ensures data independence and, while still collected in the same geographical area, the capability of the proposed method to generalize to unseen data with different characteristics.

Two main performance metrics are used to evaluate the results of the proposed neural network: the usual probability of detection and probability of false alarm metrics and the Chamfer distance (CD). In both cases, the lidar data is used as a reference, either in

the occupancy grid format for the P_d and P_{fa} computation or in the point cloud format for the Chamfer distance. While different definitions are given for the Chamfer distance in the literature, in this work, the following is used:

$$CD(S_1, S_2) = \frac{1}{|S_1|} \sum_{x \in S_1} \min_{y \in S_2} ||x - y||_2 + \frac{1}{|S_2|} \sum_{y \in S_2} \min_{x \in S_1} ||x - y||_2, \tag{6.6}$$

where S_1 and S_2 are the two sets of points being compared (e.g., the lidar points assumed as ground truth vs the points from the 3D occupancy grid provided by the proposed data-driven detector), and $|S|$ is the cardinality of the set. The closer the two sets of points are the better, and so the lower the Chamfer distance.

However, it is important to note that a caveat is needed when analyzing the P_d and the P_{fa} metrics. A small misalignment in the calibration of only a few centimeters in range or of a small angle will cause the probability of detection to fall drastically, while the probability of false alarms will rise, as can be seen in the examples presented in Figure 6.6. For instance, considering the example in Figure 6.6b, even though the P_{fa} of this case is numerically the same as in the case represented in Figure 6.6a, the impact in terms of quality of the perceived environment can be very different, especially taking into account the small cell dimensions. While this is not a problem in the proposed method (as the network used as data-driven detector is able to learn offsets such as those in this example), it may affect the other methods used in this section for benchmarking, such as the different variants of CFAR detectors. Moreover, since the radar resolutions are worse than the lidar's, many targets will be overestimated in size, raising the P_{fa}. These false alarms are, in general, assumed to be less relevant for assessing the quality of automotive radar since a small overestimation of objects in the order of centimeters (i.e., few lidar resolution cells) may not be as bad as detecting isolated ghost targets. Nevertheless, all the false alarms are treated equally in the assessment in this work since an extra clustering or tracking stage may be needed to distinguish between these unfavorable cases in terms of P_{fa}. An example of this phenomenon can be seen in Figure 6.6c. On the other hand, it can be seen how the Chamfer distance is able to capture these spatial relationships, yielding different values for the three different cases. Considering all this, a point cloud level metric like the Chamfer distance is considered a better evaluation metric for this work.

Table 6.3 shows the performance of the proposed method with the three aforementioned metrics averaged over the whole test set and compared with different alternative approaches for detection. Specifically, the different rows on the table are:

- *Proposed Method*: the results of the proposed method are explained in the previous section and with the overall architecture shown in Figure 6.5.
- *No Doppler & Quantile*: an approach similar to the one presented in [27], where only those power cells with values higher than the 0.9 quantile are kept and the rest are set to zero. Furthermore, the Doppler information is collapsed by taking the mean over the Doppler dimension. This is used to 'sparsify' the data and speed up processing, with the risk of cutting out weakly reflecting targets. Since there is no Doppler data anymore, the *DopplerEncoder* subnetwork is removed from the general architecture of the proposed data-driven detector.

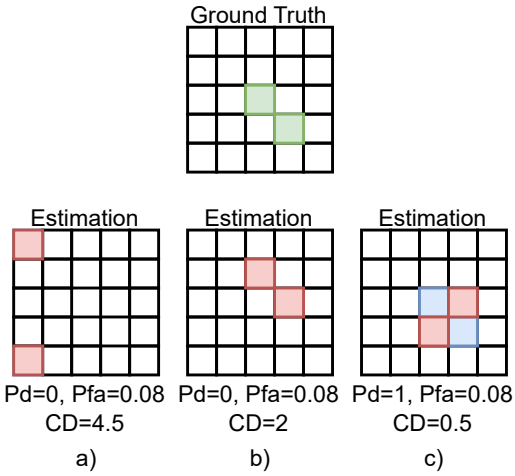


Figure 6.6: Illustration of the problem in computing the P_d, P_{fa}, and Chamfer Distance (CD) as performance metrics. In a), a case where two ghost targets are created. In b), a calibration misalignment shifts the detection cells, raising the P_{fa} as if two ghost targets were created. In c), the problem of the overestimation of target size. These three cases have nominally the same P_{fa}, but the implications for overall scene perception are completely different. It can be seen how the CD captures the spatial relationships and yields a better value in cases b) and c), where the false alarms have less impact from an application point of view.

- *No Doppler*: the proposed method without any predefined threshold on the data, but with the Doppler information collapsed by taking the mean of the data and thus without the *DopplerEncoder* subnetwork in the architecture. This is done to perform an ablation study on the importance of the Doppler domain features and analyze the effect of each block independently.
- *Quantile*: the proposed method, but with the pre-detection fixed threshold based on the 0.9 quantile inspired by [27].
- *No Time*: in order to assess the impact of inputting several frames into the network and use the temporal evolution of the scene, this tests the proposed method without the *Temporal Coherence* subnetwork in the architecture, essentially an ablation study without inter-frame temporal information.
- *OSCFAR*: a 2D OS CFAR in range-angle, followed by a 1D OS CFAR in Doppler. While multiple different CFAR alternatives have been tested (i.e., different combinations of CA and OS CFAR detectors), only the best implementation is reported here for conciseness. An analysis with different variations has been presented in [127] for completeness. Following [132], the rank has been set to 0.75 times the number of training cells, and no guard cells have been used.

As it can be seen in Table 6.3, the highest P_d and the lowest Chamfer distance is achieved by the proposed method while maintaining a similar P_{fa}. On the other hand,

Table 6.3: Performance Results of the proposed method for data-driven detection, different variations of the method, and the best-performing CFAR detector implemented.

Method	P_d (%)	P_{fa} (%)	Chamfer distance (m)
Proposed Method	62.13	2.77	1.54
No Doppler & Quantile	52.97	2.63	2.16
No Doppler	50.44	2.50	2.13
Quantile	57.9	2.85	1.92
No Time (single frame)	58.08	2.63	2.16
OSCFAR	0.41	0.015	6.73
No Elevation cases			
Proposed Method	74.83	1.12	2.92
Quantile	74.09	1.11	2.78
OSCFAR	11.56	3.1	4.11

applying the quantile cut and removing the Doppler information similarly to [27] reduces the P_d from 62.13% to 52.97%, and worsens the Chamfer distance from 1.54m to 2.16m. Looking at the results for the other versions, it can be seen that this drop in performance is mostly due to the removal of the Doppler information. Using only the quantile-based threshold may be a good trade-off since the performance degradation is not substantial, but the computational cost is reduced. Looking at the version without the *Temporal Coherence* subnetwork, which is trained on single frames, it can be seen how all the metrics are worse than in the baseline. Thus, including temporal information in the network is a good strategy to boost performance, with the only downside of increasing slightly the training time due to the extra layers. Finally, it can be seen that the conventional OS CFAR is the method that performs the worst, with a much higher Chamfer distance of 6.73m.

In order to have a fairer comparison against the conventional CFAR detector, a 2D version of the proposed method has also been evaluated by disregarding the elevation information, as this can only be estimated rather poorly due to the unfavorable design of the radar array. To this end, the proposed NN has been trained without elevation information, discarding the virtual channels in the Z-direction and, thus, treating it as a ULA in the azimuth direction. For completeness, the implementation with a quantile-based threshold has also been assessed in this new analysis. The results are shown in Table 6.3 under the 'No Elevation cases' rows. For these tests, the P_d of the OS CFAR approach is increased to 11.5%, but the P_{fa} is also raised. This is mainly due to detections triggered in the adjacent angle bins of a target generating "ring-like" patterns due to side lobes, a phenomenon also mentioned in [27]. Both the proposed method and the proposed method with the quantile-based threshold are shown to outperform the conventional OSCFAR in the three metrics.

In addition to the quantitative results, some qualitative results are also presented to show the performance of the proposed method visually. In Figure 6.7, a challenging frame from the radar point of view is shown, where the vehicle is going under a large but relatively not tall bridge. The 3D point cloud generated with the proposed method is

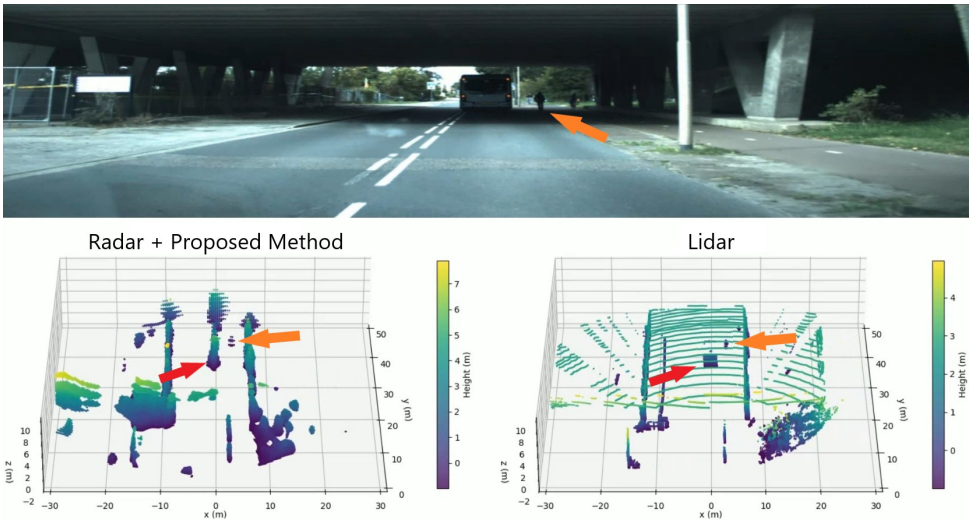


Figure 6.7: Example frame in a challenging situation for the radar system, where the vehicle is going under a large bridge. In the top figure, the camera image is shown for reference. On the left, the point cloud generated with the proposed data-driven method is shown, and on the right, the original point cloud provided by the lidar. The red arrows point to the bus under the bridge and the orange arrow to the pedestrian next to it. Note that the color in the point clouds refers to the height of the objects.

shown in the left plot, with the original lidar on the right plot. As it can be seen, the road is clear of false alarms, and the bus (in red arrow) and pedestrian (in orange arrow) are clearly detected. The bus and the ceiling merge due to the poor elevation resolution of the radar data, but they could be split and identified in Doppler.

Figure 6.8 shows another scene where the resulting point clouds have been projected onto the camera image to provide a sense of the 3D scene (top), but the bird's eye view projection is also shown (bottom). For simplicity, the point clouds have been cropped to a maximum range of 30 meters. Moreover, as a visual aid in the bird's eye view, cyclists are highlighted with an orange hexagon, cars with a red hexagon, and a large van with a light blue hexagon. In Figure 6.8a, the original lidar point cloud is presented, where many details of the scene can be appreciated. Figure 6.8b shows the detections generated using the proposed data-driven detector, and as can be seen, most of the details of the relevant targets are preserved. Objects are slightly overestimated in size, but the overall scene is clear. Also, the shape of the objects is preserved, especially in the case of cars and the shown large van. Finally, Figure 6.8c shows the output of the previously-mentioned best-performing CFAR detector, where it can be seen how the output is much sparser in terms of detected points, and also missing one of the cyclists in the scene.

Finally, an example of results where the elevation information is disregarded in the detection process is presented in Figure 6.9. Here, the figure shows the camera image for visual reference (top), and the comparison of the resulting point cloud from the radar data with the proposed data-driven detector (left), the original lidar data (center), and the point cloud from the radar data with the best-performing implemented CFAR. Note that cars are highlighted in red, and there are "ring-like" detections (highlighted in green)

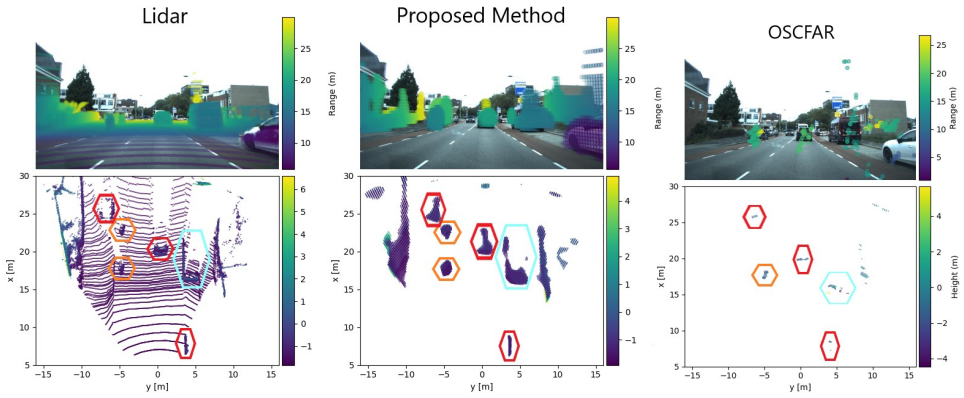


Figure 6.8: Example of data frame in urban scenario with related detections. In (a), the original lidar point cloud projected onto the camera as well as a bird's eye view. In (b), the radar point cloud generated with the proposed data-driven detector. In (c), the radar point cloud generated with the best-performing CFAR implemented (i.e., 2D OS-CFAR in range-azimuth, followed by an OS-CFAR in Doppler).

due to the high side lobes of the van, which can be seen in the figure generated using the CFAR detector. This phenomenon raises the P_{fa} and is an expected behavior that has been reported in other automotive radar datasets [27] when using CFAR detectors. As also reported in the previous qualitative examples, the point cloud generated by the proposed data-driven detector is denser than the CFAR-generated one and conserves the correct location and shape of most objects.

6.5. CONCLUSIONS

This work introduces an innovative data-driven detector for automotive radar and the *RaDelft* dataset, a newly collected multi-sensor real-world dataset. The proposed detector is trained exclusively from unlabeled synchronized radar and lidar data, thus eliminating the need for costly manual object annotations. Two types of performance metrics were employed to validate the method, i.e., conventional probability of detection & probability of false alarm, alongside the Chamfer distance, a point cloud-level metric designed to capture spatial relationships and similarities between point clouds. The proposed method reduces by 4.2 meters (77% reduction) the Chamfer distance when compared with conventional OSCFAR detectors and by 0.62 meters (28% reduction) when compared with the state-of-the-art. Also, it significantly increases the probability of detection. Moreover, an ablation study showed that including temporal information in the process is important, and Doppler information is especially crucial for our model's good performance. Results show that the probability of detection is increased from 50.44% to 62.13%, and the Chamfer distance is reduced by 27% when using Doppler information. However, it is important to point out that the proposed detector cannot control the false alarm rate, which may be a key feature for some applications, and lacks explainability (i.e., it is a black-box model with inputs and outputs).

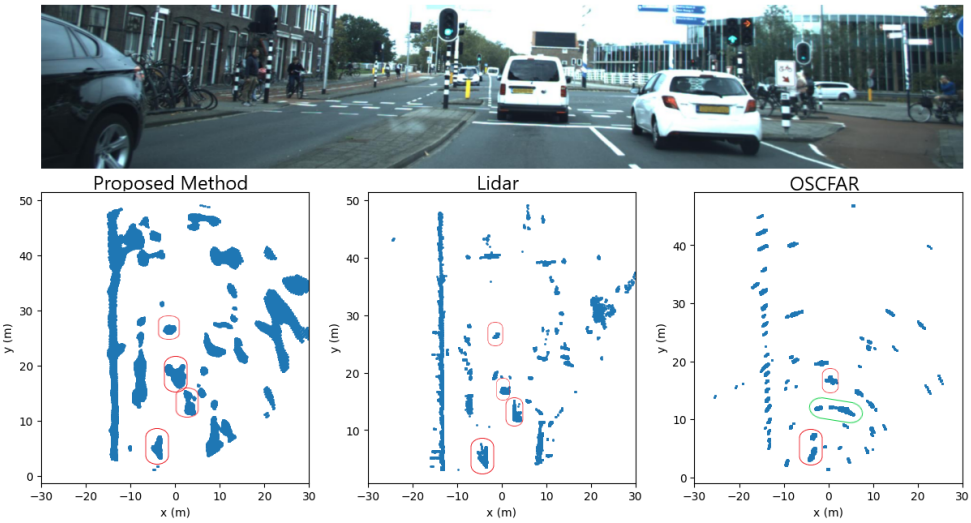


Figure 6.9: Example of data frame where the elevation information is disregarded from the detection process. In the top figure, the camera image is shown for reference. In the bottom part of the figure, the original lidar point cloud is shown (center), with the point cloud generated by the proposed data-driven detector (left) and by the best-performing implemented CFAR (right).

For the experimental evaluation of the proposed approach, a comprehensive dataset encompassing over 30 minutes of actual driving scenarios was collected using a vehicle equipped with both lidar and radar sensors, resulting in 16975 radar frames paired with corresponding lidar ground truth. Compared with other existing datasets, *RaDelft* provides raw data from a commercial 4D imaging radar needed for radar practitioners for many research lines. Moreover, it contains data processed at other levels (e.g., radar cubes and point clouds) suitable for researchers with different backgrounds and interests. The dataset is publicly available, with code to parse, visualize, and process the data, as well as the code to reproduce the results reported in this work.

7

Conclusions

7.1. MAJOR RESULTS AND NOVELTIES

This thesis introduces new techniques and algorithms to enhance the performance of automotive radars at different levels of their processing chain and for different perception tasks. They can be grouped into machine learning and compressive sensing techniques.

7.1.1. RESULTS BASED ON MACHINE LEARNING TECHNIQUES

- *Angular resolution enhancement using Machine Learning (Chapter 3)*

  A novel framework based on neural networks has been proposed to enhance the angular resolution of automotive radars. The proposed approach artificially expands the antenna aperture using a self-supervised scheme, where data from a high angular resolution radar, i.e., a radar with a large aperture antenna, is used to train a deep neural network to extrapolate the antenna element's response of a compact, low angular resolution radar. This framework has been validated with simulated and experimental data collected with a commercial automotive radar. A case study with two pedestrians walking towards the radar at the same speed and range shows that using an 86 virtual element array, the two pedestrians can be resolved 94.5% of the frames, while with an array of 44 virtual elements, only 30.1% of the frames. If the proposed method is used before the angle estimation, the probability of resolving the two pedestrians is increased to 55.6%. This result shows that the maximum distance at which both pedestrians are perceived as two different targets is increased when using the proposed method. Therefore, the proposed method can be used for more accurate and faster responses from the vehicle's planning and decision-making system.

- *Deep Automotive Radar Detector (Chapter 6)*

  The problem of detecting multiple extended targets in automotive scenes with high-resolution imaging radars has been studied. A data-driven approach is proposed where unlabeled synchronized lidar data is used as ground truth to train a deep neural network with only radar data as input. Two types of performance metrics were employed to validate the method, i.e., conventional probability of detection and probability of false alarm, alongside the Chamfer distance, a point cloud-level metric designed to capture spatial relationships and similarities between point clouds. The proposed method reduces the Chamfer distance by 4.2 meters (78% reduction) compared with conventional OSCFAR detectors and by 0.62 meters (28% reduction) when compared with the state-of-the-art. Also, it significantly increases the probability of detection. Moreover, an ablation study showed that including temporal information in the process is important, and Doppler information is especially crucial for our model's good performance. Results show that the probability of detection is increased from 50.44% to 62.13%, and the Chamfer distance is reduced by 27% when using Doppler information.

- *RaDelft Dataset (Chapter 6)*

  To support the development of machine learning algorithms applied to radar data, a novel, large-scale, real-life, and multi-sensor has been recorded using a demon-

strator vehicle in different locations in the city of Delft. *RaDelft* provides raw data from a commercial 4D imaging radar needed for radar practitioners for many research lines. Moreover, it contains data processed at other levels (e.g., radar cubes and point clouds) suitable for researchers with different backgrounds and interests. The dataset is publicly available, with code to parse, visualize, and process the data.

7.1.2. RESULTS BASED ON COMPRESSIVE SENSING TECHNIQUES

In the field of angle estimation using automotive radars, Compressive Sensing techniques stand out for their unique properties, such as providing higher resolution than FFT-based implementations with a single snapshot or being able to deal with sparse arrays. This thesis addresses three main problems while using CS: how to select a proper sparse array valid for using CS, how to reduce the main drawback of CS, i.e., the computational load, and how to estimate the angular space when the scene is not sparse.

- *Bayesian Compressive Sensing Applied to Automotive Radar (Chapter 4)*

  Bayesian compressive sensing is a family of algorithms that tackles the CS problem from a probabilistic point of view. When used, it not only provides the estimations but also their associated uncertainties. Using this uncertainty, a novel algorithm has been proposed to generate the optimal array topology for angle estimation using BCS in MIMO radars. Based on the differential entropy of the measurements, the algorithm has been executed offline with simulated data to generate the array topology and tested with real measurements using a commercial automotive MIMO radar. The results show that the generated array outperforms the average random array, which is usually considered the best strategy to generate CS sensing matrices. Moreover, is it possible to correctly resolve close space targets using 50% of the elements compared with a fully dense ULA.

  The second novelty of this work considers the computational complexity of BCS. Using a split-and-merge approach algorithm followed by a correction stage, it is proved that the proposed algorithm can reduce by a factor of 10 the computational time of the standard BCS. Moreover, the method does not reduce the accuracy and probability of resolving close space targets. Results, validated with experimental measurements, show that the proposed method overperforms FFT-based and single-snapshot MUSIC in accuracy and resolution probability, while running 10 times faster than the MUSIC algorithm.

- *Total Variation Compressive Sensing Applied to Automotive Radar (Chapter 5)*

  The angular estimation for extended targets with a single snapshot in short-range imaging radar has been analyzed. Conventional compressive sensing methods can not be directly applied due to the non-sparsity characteristics of the scene, and super-resolution algorithms based on subspace methods are not applicable to the constraints of this problem (single snapshot and coherent sources). Therefore, two novel methods have been proposed based on total variation compressive sensing, where sparsity is enforced in the signal's gradient. An independent azimuth-elevation estimation for each range cell is used in the first method, while a joint

3D range-azimuth-elevation optimization problem is proposed in the second one. The proposed algorithms have been verified on simulated data generated in a realistic manner using 3D CAD models of different cars, pedestrians, and bicycles. The results of the two methods have been compared with a conventional Fourier beamformer and the standard compressive sensing algorithm in terms of shape reconstruction. The intersection over union between the ground truth shape and the estimated shape using the four methods has been computed to evaluate the shape reconstruction capabilities. The two proposed methods consistently outperform the FFT-based and conventional CS methods, with a maximum IoU increase of 3.5x in good SNR conditions for the 2D version.

7.2. RECOMMENDATIONS FOR FUTURE WORK

Due to time limitations during this PhD project, some research lines or further development of the ideas shown in this document have not been fully performed. A list of the possible research lines as a continuation of the work in this PhD thesis is presented below.

- *Target Classification and/or Semantic Segmentation using the RaDelft Dataset*

  This research line is currently being explored by the TU Delft master's student Botao Sun, who will graduate at the end of October 2024. The first step of this is to generate class labels using lidar and camera data to serve as ground truth to the radar data. These labels, generated with the automatic labeling algorithm developed by Botao Sun, have already been added to the publicly shared RaDelft dataset [122]. Then, a semantic segmentation neural network can be designed to classify the radar targets directly in the radar cube. The initial results of this research line have been sent to the 2025 International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [133].

- *2D array design for BCS DoA*

  In the work presented in Chapter 4, the proposed method is only tested with 1D arrays, but the extension to 2D arrays can be easily performed. The main limitation here is the need for 2D commercially available MIMO arrays for the validation of this research line, so this line of work would involve not only the design of the array, but also the manufacturing and actual testing of it, where the manufacturing aspect can for example consider non-idealities such as the effect of hardware imperfections and mutual coupling between antennas.

- *Extension of the TVCS*

  The work presented in Chapter 5 proved in simulations that, by using TVCS, the object's shape can be better estimated. However, this is the only chapter in which the proposed method has not been tested with experimentally measured data. Again, the main limitation is the lack of 2D sparse arrays to test the algorithm in the real world. Moreover, the better shape estimation enabled by this method would, in principle, lead to improved classification performance. However, this hypothesis must be confirmed by implementing a classification stage after conventional

angular estimation methods, and then comparing the results to the ones obtained with TVCS.

- *Comprehensive radar signal processing pipeline*

  This thesis proposed different methods for different, separate steps in the radar signal processing pipeline, from enhancing the radar signal before the angle estimation to the detection stage. However, most of the proposed methods have been developed and studied as an independent block. A future research line can combine all these different methods into a single, end-to-end processing pipeline and assess the compound effect in the system's final performance.

BIBLIOGRAPHY

[1] F. Sezgin, D. Vriesman, D. Steinhauser, R. Lugner, and T. Brandmeier, "Safe autonomous driving in adverse weather: Sensor evaluation and performance monitoring," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–6.

[2] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian, "The rise of radar for autonomous vehicles: Signal processing solutions and future research directions," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 20–31, 2019.

[3] "Nxp radar - investor relations." [Online]. Available: https://investors.nxp.com/static-files/2cae75de-3618-47ec-97ca-37571cb9b983

[4] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98–117, jul 2020.

[5] N. Scheiner, O. Schumann, F. Kraus, N. Appenrodt, J. Dickmann, and B. Sick, "Off-the-shelf sensor vs. experimental radar - how much resolution is necessary in automotive radar classification?" in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 2020, pp. 1–8.

[6] H. Buddendick and T. F. Eibert, "Incoherent scattering-center representations and parameterizations for automobiles," *IEEE Antennas and Propagation Magazine*, vol. 54, no. 1, pp. 140–148, 2012.

[7] M. Bühren and B. Yang, "Automotive Radar Target List Simulation based on Reflection Center Representation of Objects," *Workshop on Intelligent Transportation (WIT)*, no. September, pp. 161–166, 2006.

[8] M. Andres, P. Feil, and W. Menzel, "3D-scattering center detection of automotive targets using 77 GHz UWB radar sensors," *Proceedings of 6th European Conference on Antennas and Propagation, EuCAP 2012*, pp. 3690–3693, 2012.

[9] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*. Scitech Publishing Inc, 2010, vol. 1.

[10] Y. Cheng, J. Su, M. Jiang, and Y. Liu, "A novel radar point cloud generation method for robot environment perception," *IEEE Transactions on Robotics*, vol. 38, no. 6, pp. 3754–3773, 2022.

[11] J. Yoon, S. Lee, S. Lim, and S.-C. Kim, "High-density clutter recognition and suppression for automotive radar systems," *IEEE Access*, vol. 7, pp. 58 368–58 380, 2019.

[12] D. Gusland, J. M. Christiansen, B. Torvik, F. Fioranelli, S. Z. Gurbuz, and M. Ritchie, "Open radar initiative: Large scale dataset for benchmarking of micro-doppler recognition algorithms," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.

[13] A. Palffy, J. Dong, J. F. Kooij, and D. M. Gavrila, "CNN based road user detection using the 3d radar cube," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1263–1270, 2020.

[14] Y. Jin, M. Hoffmann, A. Deligiannis, J.-C. Fuentes-Michel, and M. Vossiek, "Semantic segmentation-based occupancy grid map learning with automotive radar raw data," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 216–230, 2024.

[15] H. Sun, F. Brigui, and M. Lesturgie, "Analysis and comparison of mimo radar waveforms," in *2014 International Radar Conference*, 2014, pp. 1–6.

[16] L. Xu, J. Lien, and J. Li, "Doppler–range processing for enhanced high-speed moving target detection using lfmcw automotive radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 1, pp. 568–580, 2022.

[17] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for mimo radar," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 419–430, 2014.

[18] I. Roldan, F. Fioranelli, and A. Yarovoy, "Total Variation Compressive Sensing for Extended Targets in MIMO Radar," *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, vol. 2022-June, pp. 61–65, 2022.

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

[20] S. Yuan, F. Fioranelli, and A. G. Yarovoy, "3drudat: 3d robust unambiguous doppler beam sharpening using adaptive threshold for forward-looking region," *IEEE Transactions on Radar Systems*, vol. 2, pp. 138–153, 2024.

[21] J. Fuchs, M. Gardill, M. Lübke, A. Dubey, and F. Lurz, "A machine learning perspective on automotive radar direction of arrival estimation," *IEEE Access*, vol. 10, pp. 6775–6797, 2022.

[22] I. Bilik, S. Villeval, D. Brodeski, H. Ringel, O. Longman, P. Goswami, C. Y. B. Kumar, S. Rao, P. Swami, A. Jain, A. Kumar, S. Ram, K. Chitnis, Y. Dutt, A. Dubey, and S. Liu, "Automotive multi-mode cascaded radar data processing embedded system," in *2018 IEEE Radar Conference (RadarConf18)*, 2018, pp. 0372–0376.

[23] D. Brodeski, I. Bilik, and R. Giryes, "Deep radar detector," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.

[24] D.-H. Paek, S.-H. Kong, and K. T. Wijaya, "K-radar: 4d radar object detection for autonomous driving in various weather conditions," in *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.

[25] Y. Lin, X. Wei, Z. Zou, and W. Yi, "Deep learning based target detection method for the range-azimuth-doppler cube of automotive radar," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 2023, pp. 2868–2873.

[26] F. E. Nowruzi, D. Kolhatkar, P. Kapoor, F. Al Hassanat, E. J. Heravi, R. Laganiere, J. Rebut, and W. Malik, "Deep open space segmentation using automotive radar," in *2020 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, 2020, pp. 1–4.

[27] D.-H. Paek, S.-H. Kong, and K. T. Wijaya, "K-radar: 4d radar object detection for autonomous driving in various weather conditions," in *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. [Online]. Available: https://openreview.net/forum?id=W_bsDmzwaZ7

[28] D. Gusland, S. Rolfsjord, and B. Torvik, "Deep temporal detection - a machine learning approach to multiple-dwell target detection," in *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 203–207.

[29] O. Schumann, J. Lombacher, M. Hahn, C. Wohler, and J. Dickmann, "Scene Understanding with Automotive Radar," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 188–203, 2020.

[30] Y. Wang, Z. Jiang, Y. Li, J.-N. Hwang, G. Xing, and H. Liu, "Rodnet: A real-time radar object detection network cross-supervised by camera-radar fused object 3d localization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 954–967, 2021.

[31] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Tracking of extended objects with high-resolution doppler radar," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1341–1353, 2016.

[32] M. Hassan, F. Fioranelli, A. Yarovoy, and S. Ravindran, "Radar multi object tracking using dnn features," in *2023 IEEE International Radar Conference (RADAR)*, 2023, pp. 1–6.

[33] R. Perez, F. Schubert, R. Rasshofer, and E. Biebl, "Single-frame vulnerable road users classification with a 77GHz FMCW radar sensor and a convolutional neural network," *Proceedings International Radar Symposium*, vol. 2018-June, pp. 1–10, 2018.

[34] W. Kim, H. Cho, J. Kim, B. Kim, and S. Lee, "Yolo-based simultaneous target detection and classification in automotive FMCW radar systems," *Sensors (Switzerland)*, vol. 20, no. 10, pp. 1–15, 2020.

[35] X. Cai, M. Giallorenzo, and K. Sarabandi, "Machine Learning-Based Target Classification for MMW Radar in Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8858, no. c, 2021.

[36] C. Lundquist, L. Hammarstrand, and F. Gustafsson, "Road Intensity Based Mapping Using Radar Measurements With a Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1397–1408, apr 2011.

[37] I. Orr, M. Cohen, and Z. Zalevsky, "High-resolution radar road segmentation using weakly supervised learning," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 239–246, 2021.

[38] P. Stoica and J. Li, *MIMO Radar Signal Processing*. Wiley, 2009.

[39] J. Bergin and J. Guerci, *MIMO Radar: Theory and Application*. Artech House, 2018.

[40] J. Li and P. Stoica, "MIMO Radar with Colocated Antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, sep 2007.

[41] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[42] R. Feng, F. Uysal, P. Aubry, and A. Yarovoy, "MIMO–monopulse target localisation for automotive radar," *IET Radar, Sonar & Navigation*, vol. 12, no. 10, pp. 1131–1136, oct 2018.

[43] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, mar 1986.

[44] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *Adaptive Antennas for Wireless Communications*, vol. 37, no. 7, pp. 224–235, 2009.

[45] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[46] S. Fortunati, R. Grasso, F. Gini, and M. S. Greco, "Single snapshot DOA estimation using compressed sensing," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 2297–2301, 2014.

[47] R. Heckel, "Super-resolution MIMO radar," in *2016 IEEE International Symposium on Information Theory (ISIT)*, vol. 2016-Augus. IEEE, jul 2016, pp. 1416–1420.

[48] S. Brisken, F. Ruf, and F. Höhne, "Recent evolution of automotive imaging radar and its information content," *IET Radar, Sonar and Navigation*, vol. 12, no. 10, pp. 1078–1081, 2018.

[49] M. Stolz, M. Wolf, F. Meinl, M. Kunert, and W. Menzel, "A New Antenna Array and Signal Processing Concept for an Automotive 4D Radar," *2018 15th European Radar Conference, EuRAD 2018*, pp. 63–66, 2018.

[50] T. I. Inc. (2019) Design guide: Tidep-01012—imaging radar using cascaded mmwave sensor reference design (rev. a). [Online]. Available: https://www.ti.com/lit/ug/tiduen5a/tiduen5a.pdf

[51] C. Alistarh, L. Anitori, S. K. Podilchak, J. Thompson, P. D. Hilario Re, M. Sellathurai, G. Goussetis, and J. Lee, "Millimeter-wave Automotive Radar using Extrapolation for Improved Angular Resolution," *EuRAD 2020 - 2020 17th European Radar Conference*, no. January, pp. 394–397, 2021.

[52] B. Cho and S. Sun, "Cross-range resolution improvement in forward-looking imaging radar using autoregressive model-based data extrapolation," *IET Radar, Sonar & Navigation*, vol. 9, no. 8, pp. 933–941, oct 2015.

[53] H.-W. Cho, W. Kim, S. Choi, M. Eo, S. Khang, and J. Kim, "Guided Generative Adversarial Network for Super Resolution of Imaging Radar," in *2020 17th European Radar Conference (EuRAD)*. IEEE, jan 2021, pp. 144–147.

[54] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, ser. Springer Texts in Statistics. Springer International Publishing, 2016.

[55] J. Dickmann, J. Klappstein, M. Hahn, N. Appenrodt, H.-L. Bloecher, K. Werber, and A. Sailer, "Automotive radar the key technology for autonomous driving: From detection and ranging to environmental understanding," in *2016 IEEE Radar Conference (RadarConf)*. IEEE, may 2016, pp. 1–6.

[56] P.-J. Chung, M. Viberg, and J. Yu, "Chapter 14 - doa estimation methods and algorithms," in *Academic Press Library in Signal Processing: Volume 3*, ser. Academic Press Library in Signal Processing, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds. Elsevier, 2014, vol. 3, pp. 599–650.

[57] W. Liao and A. Fannjiang, "Music for single-snapshot spectral estimation: Stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, 2016.

[58] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, oct 2017.

[59] H. L. Van Trees, "Chapter 8 - parameter estimation i: Maximum likelihood," in *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory*. Wiley, 2002, vol. 3.

[60] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8, 1983, pp. 336–339.

[61] M. A. Alrmah, S. Weiss, and S. Lambotharan, "An extension of the music algorithm to broadband scenarios using a polynomial eigenvalue decomposition," in *2011 19th European Signal Processing Conference*, 2011, pp. 629–633.

[62] Y. Yu, A. P. Petropulu, and H. V. Poor, "Compressive sensing for mimo radar," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3017–3020.

[63] T. Strohmer and B. Friedlander, "Compressed sensing for mimo radar - algorithms and performance," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, 2009, pp. 464–468.

[64] Z. Yang, J. Li, P. Stoica, and L. Xie, "Chapter 11 - sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, pp. 509–581.

[65] M. Weiß, "Chapter 6 - sparsity-based radar technique," in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, pp. 277–316.

[66] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[67] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[68] L. Stankovic, D. P. Mandic, M. Dakovic, and I. Kisil, "Demystifying the coherence index in compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 152–162, 2020.

[69] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[70] A. Faul and M. Tipping, "Analysis of sparse bayesian learning," in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14. MIT Press, 2001.

[71] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse bayesian models," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, C. M. Bishop and B. J. Frey, Eds., vol. R4. PMLR, 03–06 Jan 2003, pp. 276–283, reissued by PMLR on 01 April 2021.

[72] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.

[73] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Directions-of-arrival estimation through bayesian compressive sensing strategies," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3828–3838, 2013.

[74] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.

[75] S. Ji and L. Carin, "Bayesian compressive sensing and projection optimization," in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 377–384.

[76] T. M. Cover and J. A. Thomas, "Differential entropy," in *Elements of Information Theory*. John Wiley & Sons, Ltd, 2005, pp. 243–259.

[77] L. Lamberti, "Bayesian compressive sensing for radar based doa estimation using sparse sensor arrays," 08 2022. [Online]. Available: https://repository.tudelft.nl/islandora/object/uuid%3A903e6afc-0feb-4f6c-a758-afca1cc2d3ee?collection=education

[78] M. Courcoux-Caro, C. Vanwynsberghe, C. Herzet, and A. Baussard, "Sequential sensor placement using bayesian compressed sensing for source localization." IEEE, 1 2021, pp. 241–245.

[79] J. H. Ender, "On compressive sensing applied to radar," *Signal Processing*, vol. 90, pp. 1402–1414, 5 2010.

[80] M. Carlin, P. Rocca, F. Viani, G. Oliveri, and A. Massa, "Multi-resolution BCS-based approach for DoA estimation," *IEEE Antennas and Propagation Society, AP-S International Symposium (Digest)*, vol. 2015-October, pp. 1074–1075, 2015.

[81] Q. Zhang, "Probability of resolution of the music algorithm," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 978–987, 1995.

[82] J. E. Evans, D. Sun, and J. Johnson, "Application of advanced signal processing techniques to angle of arrival estimation in atc navigation and surveillance systems," Massachusetts Inst of Tech Lexington Lincoln Lab, Tech. Rep., 1982.

[83] T.-J. Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 3, pp. 527–536, 1985.

[84] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 806–811, 1985.

[85] Q. Chen and R. Liu, "On the explanation of spatial smoothing in music algorithm for coherent sources," in *International Conference on Information Science and Technology*, 2011, pp. 699–702.

[86] H. L. Van Trees, *Optimum array processing*. John Wiley & Sons, 2002.

[87] Z. Yang and Y. R. Zheng, "A comparative study of compressed sensing approaches for 3-d synthetic aperture radar image reconstruction," *Digital Signal Processing*, vol. 32, pp. 24–33, 2014.

[88] Z. Tan, Y. C. Eldar, and A. Nehorai, "Direction of arrival estimation using co-prime arrays: A super resolution viewpoint," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5565–5576, 2014.

[89] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for mimo radar," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 419–430, 2014.

[90] Y. Yu, A. P. Petropulu, and H. V. Poor, "Mimo radar using compressive sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 146–163, 2010.

[91] J.-Y. Lee and L. Greengard, "The type 3 nonuniform fft and its applications," *Journal of Computational Physics*, vol. 206, no. 1, pp. 1–5, 2005.

[92] C. Li, W. Yin, and Y. Zhang, "Tval3: Tv minimization by augmented lagrangian and alternating direction algorithm," 2009. [Online]. Available: https://www.caam.rice.edu/~optimization/L1/TVAL3/

[93] C. Li, W. Yin, H. Jiang, and Y. Zhang, "An efficient augmented lagrangian method with applications to total variation minimization," *Computational Optimization and Applications*, vol. 56, no. 3, pp. 507–530, Dec 2013.

[94] Q. Kong, R. Gong, J. Liu, and X. Shao, "Investigation on reconstruction for frequency domain photoacoustic imaging via tval3 regularization algorithm," *IEEE Photonics Journal*, vol. 10, no. 5, pp. 1–15, 2018.

[95] X. Liu, L. Zhang, Y. Zhang, and L. Qiao, "A comparative study of four total variational regularization reconstruction algorithms for sparse-view photoacoustic imaging," *Computational and Mathematical Methods in Medicine*, vol. 8, pp. 1046–1055, Oct 2021.

[96] E. Bel Kamel, A. Peden, and P. Pajusco, "Rcs modeling and measurements for automotive radar applications in the w band," in *2017 11th European Conference on Antennas and Propagation (EUCAP)*, 2017, pp. 2445–2449.

[97] S. Lee, S. Kang, S.-C. Kim, and J.-E. Lee, "Radar cross section measurement with 77 ghz automotive fmcw radar," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–6.

[98] Y. Deep, P. Held, S. S. Ram, D. Steinhauser, A. Gupta, F. Gruson, A. Koch, and A. Roy, "Radar cross-sections of pedestrians at automotive radar frequencies using ray tracing and point scatterer modelling," *IET Radar, Sonar & Navigation*, vol. 14, no. 6, pp. 833–844, 2020.

[99] R. López-Sastre, A. García-Fuertes, C. Redondo-Cabrera, F. Acevedo-Rodríguez, and S. Maldonado-Bascón, "Evaluating 3d spatial pyramids for classifying 3d shapes," *Computers & Graphics*, vol. 37, no. 5, pp. 473–483, 2013.

[100] B. Li, A. Godil, M. Aono, X. Bai, T. Furuya, L. Li, R. López-Sastre, H. Johan, R. Ohbuchi, C. Redondo-Cabrera, A. Tatsuma, T. Yanagimachi, and S. Zhang, "SHREC'12 Track: Generic 3D Shape Retrieval," in *Eurographics Workshop on 3D Object Retrieval*, M. Spagnuolo, M. Bronstein, A. Bronstein, and A. Ferreira, Eds. The Eurographics Association, 2012.

[101] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Shape Modeling International*, Jun. 2004.

[102] R. W. Sumner and J. Popovi, "Deformation transfer for triangle meshes," *ACM Trans. Graph.*, vol. 23, no. 3, p. 399–405, aug 2004.

[103] A. Bronstein, M. Bronstein, and R. Kimmel, *Numerical Geometry of Non-Rigid Shapes*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[104] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," in *ACM SIGGRAPH 2007 Papers*, ser. SIGGRAPH '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 24–es.

[105] E. Cades. (2023) l1-magic. [Online]. Available: https://candes.su.domains/software/l1magic/

[106] W. Tong, C. Sima, T. Wang, L. Chen, S. Wu, H. Deng, Y. Gu, L. Lu, P. Luo, D. Lin, and H. Li, "Scene as occupancy," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 8372–8381.

[107] X. Tian, T. Jiang, L. Yun, Y. Wang, Y. Wang, and H. Zhao, "Occ3d: A large-scale 3d occupancy prediction benchmark for autonomous driving," *arXiv preprint arXiv:2304.14365*, 2023.

[108] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, "Towards deep radar perception for autonomous driving: Datasets, methods, and challenges," *Sensors*, vol. 22, no. 11, 2022.

[109] J. Rebut, A. Ouaknine, W. Malik, and P. Pérez, "Raw high-definition radar for multi-task learning," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 17 000–17 009.

[110] A. Kramer, K. Harlow, C. Williams, and C. Heckman, "Coloradar: The direct 3d millimeter wave radar dataset," *The International Journal of Robotics Research*, vol. 41, no. 4, pp. 351–360, 2022.

[111] M. Mostajabi, C. M. Wang, D. Ranjan, and G. Hsyu, "High resolution radar dataset for semi-supervised learning of dynamic objects," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 450–457.

[112] M. Sheeny, E. D. Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "Radiate: A radar dataset for automotive perception in bad weather," 2021.

[113] A. Ouaknine, A. Newson, J. Rebut, F. Tupin, and P. Perez, "Carrada dataset: Camera and automotive radar with range- angle- doppler annotations," in *2020 25th International Conference on Pattern Recognition (ICPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jan 2021, pp. 5068–5075.

[114] A. Zhang, F. E. Nowruzi, and R. Laganiere, "Raddet: Range-azimuth-doppler based radar object detection for dynamic road users," in *2021 18th Conference on Robots and Vision (CRV)*, 2021, pp. 95–102.

[115] T.-Y. Lim, S. A. Markowitz, and M. N. Do, "Radical: A synchronized fmcw radar, depth, imu and rgb camera data dataset with low-level fmcw radar signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 941–953, 2021.

[116] J. Yang, J. Yi, T. Sakamoto, and X. Wan, "An extended target detector using image-processing techniques exploiting energy-spillover phenomenon in radar echoes," *IEEE Sensors Journal*, vol. 23, no. 19, pp. 22 919–22 929, 2023.

[117] J. Guan and X. Zhang, "Subspace detection for range and doppler distributed targets with rao and wald tests," *Signal Processing*, vol. 91, no. 1, pp. 51–60, 2011.

[118] J. Carretero-Moya, J. Gismero-Menoyo, A. Asensio-Lopez, and A. Blanco-Del-Campo, "Small-target detection in high-resolution heterogeneous sea-clutter: An empirical analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 3, pp. 1880–1898, 2011.

[119] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-cnn: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, p. 5119–5132, Feb. 2021. [Online]. Available: http://dx.doi.org/10.1109/JSEN.2020.3036047

[120] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+1d radar in the view-of-delft dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4961–4968, 2022.

[121] J. Domhof, J. F. P. Kooij, and D. M. Gavrila, "A joint extrinsic calibration tool for radar, camera and lidar," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 3, pp. 571–582, 2021.

[122] I. Roldan, A. Pálffy, S. Yuan, S. Zhu, F. Fioranelli, D. Gavrila, J. F. P. Kooij, and A. Yarovoy, "Radelft dataset: a large-scale, real-life, and multi-sensor automotive dataset," 2024. [Online]. Available: https://data.4tu.nl/datasets/4e277430-e562-4a7a-adfe-30b58d9a5f0a/1

[123] D. Zoeke and A. Ziroff, "Phase migration effects in moving target localization using switched mimo arrays," in *2015 European Radar Conference (EuRAD)*, 2015, pp. 85–88.

[124] C. M. Schmid, R. Feger, C. Pfeffer, and A. Stelzer, "Motion compensation and efficient array design for tdma fmcw mimo radar systems," in *2012 6th European Conference on Antennas and Propagation (EUCAP)*, 2012, pp. 1746–1750.

[125] S. Sun and Y. D. Zhang, "4d automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 879–891, 2021.

[126] S. Lee, H. Lim, and H. Myung, "Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3d point cloud," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 13 276–13 283.

[127] I. Roldan, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli, and A. Yarovoy, "See further than cfar: a data-driven radar detector trained by lidar," in *2024 IEEE Radar Conference, Denver, CO*, 2024.

[128] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944.

[129] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[130] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2018.

[131] Delft High Performance Computing Centre (DHPC), "DelftBlue Supercomputer (Phase 2)," https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2, 2024.

[132] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 4, pp. 608–621, 1983.

[133] B. Sun, I. Roldan, and F. Fioranelli, "Automatic labelling & semantic segmentation with 4d radar tensor," in *ICASSP '25. IEEE International Conference on Acoustics, Speech, and Signal Processing (Under review))*, 2025.

ACKNOWLEDGEMENTS

Here, I should list the names of those who helped me during the completion of my PhD, as is standard in all the theses I have read. I should acknowledge my supervisors for their advice, guidance, and trust in my ideas. I should mention all the members of the MS3 group for creating such an amazing work and life environment for a newcomer in Delft. I should thank all my friends, both those back home and the new ones I've made, for sharing incredible moments with me over these years. I should also thank my former colleagues; the knowledge and experience I gained there made the start of my PhD journey much smoother. I should write about how my girlfriend has helped me become a better version of myself and made sure I didn't stress too much throughout the PhD. And, of course, I should thank my family for supporting me from afar.

However, instead of thanking everyone here (though I am deeply grateful to all of you), I would like to use these lines to thank the one person without whom I wouldn't be writing these words. This thesis is dedicated to my mother, Ana, who raised me with love and taught me values far more precious than anything I learned during my PhD: kindness and integrity. Thank you, Mom.

Se supone que aquí debería enumerar los nombres de todos los que me ayudaron durante mi doctorado, como es habitual en todas las tesis que he leído. Debería agradecer a mis supervisores por su consejo, guía y confianza en mis ideas. Debería mencionar a todos los miembros del grupo MS3 por crear un ambiente de trabajo y de vida tan increíble para un recién llegado a Delft. Debería agradecer a todos mis amigos, tanto a los de mi tierra como a los nuevos que he hecho, por compartir momentos increíbles conmigo durante estos años. También debería agradecer a mis antiguos compañeros de trabajo; el comienzo de mi viaje como doctorando fue mucho más fácil gracias al conocimiento y la experiencia que adquirí allí. Debería escribir sobre cómo mi novia me ha ayudado a convertirme en una mejor versión de mí mismo y de cómo se aseguró de que no me estresara demasiado durante el doctorado. Y, por supuesto, debería agradecer a mi familia por apoyarme desde lejos.

Sin embargo, en lugar de agradecer a todos aquí (aunque estoy profundamente agradecido con todos vosotros), me gustaría utilizar estas líneas para agradecer a la persona sin la cual no estaría escribiendo estas palabras. Esta tesis está dedicada a mi madre, Ana, quien me crio con amor y me enseñó valores mucho más preciados que cualquier cosa que haya aprendido durante mi doctorado: la bondad y la integridad. Gracias, mamá.

About the Author

Ignacio Roldán Montero received his B.Sc. and M.Sc. in Telecommunication Engineering at the Universidad Politécnica de Madrid, Spain, in 2014 and 2016. In 2018 he complemented his education with an M.Sc. in Signal Processing and Machine Learning at the same university. Ignacio worked for more than 5 years for Advanced Radar Technologies, a Spanish tech company focused on designing and manufacturing radar systems. In this period, he was involved in several international projects developing state-of-the-art signal processing techniques for radars. In his last stage, he focused on applying machine learning techniques to UAV detection and classification. In September 2020, he joined the Microwave Sensing, Signals, and Systems group at Delft University of Technology as a Ph.D. candidate focusing on automotive radar. He received the best student paper award at the 2024 IEEE Radar Conference held in Denver, USA, for his work on automotive radar target detection using neural networks.

List of Publications

Journal papers

3. **I. Roldan**, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli and A. Yarovoy, "A Deep Automotive Radar Detector using the RaDelft Dataset," *in IEEE Transactions on Radar Systems*

2. **I. Roldan**, F. Fioranelli and A. Yarovoy, "Total Variation Compressive Sensing for 3D Shape Estimation in Short-Range Imaging Radars," in *"IEEE Transactions on Radar Systems*, vol. 1, pp. 583-592, 2023

1. **I. Roldan**, F. Fioranelli and A. Yarovoy, "Self-Supervised Learning for Enhancing Angular Resolution in Automotive MIMO Radars," in *"IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 11505-11514, Sept. 2023.

Conference papers

5. **I. Roldan**, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli and A. Yarovoy, "See Further Than CFAR: a Data-Driven Radar Detector Trained by Lidar," *2024 IEEE Radar Conference (RadarConf24)*, Denver, CO, USA, 2024, pp. 1-6

4. L. L. Lamberti, **I. Roldan**, F. Fioranelli and A. Yarovoy, "Sparse Array Placement for Bayesian Compressive Sensing Based Direction of Arrival Estimation," *2024 IEEE Radar Conference (RadarConf24)*, Denver, CO, USA, 2024, pp. 1-6

3. **I. Roldan**, L. Lamberti, F. Fioranelli and A. Yarovoy, "Low Complexity Single-Snapshot DoA Estimation via Bayesian Compressive Sensing," *2023 IEEE Radar Conference (RadarConf23)*, San Antonio, TX, USA, 2023, pp. 1-6

2. **I. Roldan**, F. Fioranelli and A. Yarovoy, "Total Variation Compressive Sensing for Extended Targets in MIMO Radar," *"2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Trondheim, Norway, 2022, pp. 61-65

1. **I. Roldan**, F. Fioranelli and A. Yarovoy, "Enhancing Angular Resolution Using Neural Networks in Automotive Radars," 2021 *18th European Radar Conference (EuRAD)*, London, United Kingdom, 2022, pp. 58-61

Papers published during the PhD not covered in this thesis

5. B. Sun, **I. Roldan**, F. Fioranelli, "Automatic Labelling & Semantic Segmentation with 4D Radar Tensors", under review at ICASSP '25, IEEE International Conference on Acoustics, Speech, and Signal Processing.

4. N. B. Onat, **I. Roldan**, F. Fioranelli, A. Yarovoy and Y. Aslan, "Constrained Infinitesimal Dipole Modeling Assisted Ensemble Prediction of Embedded Element Patterns via Machine Learning", in IEEE Transactions on Antennas and Propagation, vol. 72, no. 9, pp. 7353-7358, Sept. 2024.

3. N. B. Onat, **I. Roldan**, F. Fioranelli, A. Yarovoy and Y. Aslan, "Dataset Dependency of Data-Driven ML Techniques in Pattern Prediction Under Mutual Coupling, " 2024 4th URSI Atlantic Radio Science Meeting (AT-RASC), pp. 1-4, 2024

2. N. B. Onat, **I. Roldan**, F. Fioranelli, A. Yarovoy and Y. Aslan, "Efficient Embedded Element Pattern Prediction via Machine Learning: A Case Study with Planar Non-Uniform Sub-Arrays," 2023 17th European Conference on Antennas and Propagation (EuCAP), Florence, Italy, 2023, pp. 1-5

1. F. Fioranelli, S. Zhu and **I. Roldan**, "Benchmarking Classification Algorithms for Radar-Based Human Activity Recognition," in IEEE Aerospace and Electronic Systems Magazine, vol. 37, no. 12, pp. 37-40, 1 Dec. 2022

AWARDS

2. Best Student Paper Award in the 2024 IEEE RadarConf, Denver, USA, for the paper "See Further Than CFAR: a Data-Driven Radar Detector Trained by Lidar"

1. Runner up for the Best Paper Award in the 2021 EuRAD for the paper "Enhancing Angular Resolution Using Neural Networks in Automotive Radars"

DATASETS

1. **I. Roldan**, "RaDelft Dataset: a large-scale, real-life, and multi-sensor automotive dataset", 4TU.ResearchData. https://data.4tu.nl/datasets/4e277430-e562-4a7a-adfe-30b58d9a5f0a

INVITED TALKS

1. **I. Roldan**, "Machine Learning & Compressive Sensing Techniques in Automotive Radars" at the EUMW 2023 Workshop: Unlocking the Potential of Radar by Compressed-Sensing, Machine-Learning, and High-Resolution Data Processing