Systems-driven Improvements to Radar-only Ego-motion Estimation



Roberto Aldera

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

Systems-driven Improvements to Radar-only Ego-motion Estimation
Candidate: Roberto Aldera
Supervisor: Paul Newman
Examiners: Nick Hawes and Sen Wang
Date of examination: 10 January 2022

University of Oxford
Mobile Robotics Group
Oxford Robotics Institute
Department of Engineering Science

To Granddad,
for always cheering me on,
encouraging me to more,
and that Oxford day trip fifteen years ago

Statement of authorship

This thesis is submitted to the Department of Engineering Science, University of Oxford, in fulfilment of the requirements for the degree of Doctor of Philosophy. This thesis is entirely my own work, and except where otherwise stated, describes my own research.

Roberto Aldera
Pembroke College
October 2021

Systems-driven Improvements to Radar-only Ego-motion Estimation

Roberto Aldera

Pembroke College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2021

Abstract

Ego-motion estimation in robotics is typically performed with cameras or laser sensors using established techniques that have proven successful in a variety of applications. While these advances have expanded the capabilities of autonomous systems, they are limited by sensor hardware that is adversely affected by poor weather and lighting conditions. Sensing with radar offers a robust means to perceive the environment in even the harshest conditions where these optical sensors fail.

This thesis presents strategic improvements to a radar-only ego-motion estimation pipeline to extend its utility as a core component in an autonomous system. By leveraging information pertaining to the context in which the sensor is used, we offer a means to reduce processing time by filtering out spurious returns, predict failure using an introspective component, and estimate ego-motion with much greater accuracy by considering motion constraints.

The first contribution uses a weakly-supervised approach to train a neural network that filters out radar returns not likely to correspond to static objects in the scene. By passing only those returns that are useful for ego-motion estimation, processing times are reduced by a factor of 2.36. Secondly, we propose a system that interprets the information encoded within the data association algorithm such that the overall confidence in a match between two radar landmark sets can be determined. A classifier is trained to distinguish between strong and weak correspondences and flag poor performance prior to taking action, reducing failure frequency by 24.7 %. Finally, we present a method which leverages information regarding the physical constraints of the robot platform to refine landmark correspondences. We discard those that suggest extreme ego-motion when under the assumption that the vehicle traverses paths of constant curvature. This includes a way to estimate relative pose from a single landmark observed twice, and reduces translational error in odometry by 45.94 %.

In each instance, validation is performed on real data captured from a vehicle-mounted scanning radar to measure the performance improvement of the proposed subsystem against a baseline Radar Odometry (RO) system. Through these data-driven advances, we seek to further the use of radar as a navigation sensor, so as to enable our robots to roam with greater efficiency, safety, and accuracy than before.

Acknowledgements

There are many people I have the pleasure of thanking for the part they've played during the course of my DPhil journey. Firstly to my supervisor Paul, for his enthusiasm, optimism, guidance, and all-round exuberance over the past four years. Thank you for the funding, time, and other resources you've invested in me, and for giving me an excuse to visit Iceland (for science, of course).

My two favourite postdocs, Matt and Daniele — thank you for your time and patience with me in every meeting, debugging session, and hallway conversation (when those were still a thing we got to do) during my time in Oxford. Matt, thank you for your thorough proof-reading of this thesis and for your pointers where there were gaps (or cliff edges). Daniele, thank you for always being available to answer my many questions, and for helping to further my understanding of both Python and regional Italian dishes, not necessarily in that order.

To the wider ORI, both past and present, thank you for making this an exciting place to work, and for accompanying me on what would otherwise be a very lonely road. The combined effort from the engineers, admin team, and other academics really enables us students to focus on the research knowing that we are well-supported by those in the George building. I would like to say a big thank you to Sarah for taking the time early on to teach me how the magic of RO works, and for offering guidance while I found my feet here. And of course those who I was fortunate enough to work alongside in Room 2: Oliver, Fabian, Sasha, Rob, Charlie, Marc, Mark, Georgi, and Chia-Man, thank you for the company and the laughs.

Thank you to the Pembroke community, who have made my Oxford experience so unique. I'll miss Regina/Reginald's obscure MCR notices, the annual Master's ramble, boat burnings (I don't row but we seem to be quite good at winning), and obsession for everything that is pink.

To those I get to call friends in Oxford from outside the ORI, who just think we play with robots all day: thank you for reminding me how awesome this field is, and for nodding patiently when I've tried to explain what it is I'm working on to you. St Aldates postgrads, there are too many of you to thank by name without me accidentally leaving someone off, but I think that's a testament to how blessed I've been to have had such a vibrant community around me for these four years. Thank you for the way you've all made Oxford feel like home, and just for the authenticity and love you all bring to the fellowship we have together in Christ.

Mom and Dad, thank you for supporting and encouraging me through all my years of growth and in my education. Thanks too for the years of cooking lessons, good home-cooked food certainly makes the 8-month British winter a little easier. And Mams and Bruce, thank you for cheering me on regularly from a distance, having already seen Erin through her own PhD. Caterina, Chris, and little Alex, having you so nearby during this time has been a joy. Thank you for being so instrumental at the start of this journey, having us around for a braai on a regular basis, and introducing me to Tesco's £3 meal deal.

Erin, my love, thank you for your constant support and companionship. It's been such a gift to have you by my side for this adventure. Thank you for being a soundboard, cheerleader, and signpost to our God throughout this time of many ups and downs. Thank you for giving so much of your time to proof-read this thesis, and tolerating my use of the Oxford comma.

And finally to the one who formed me, knows me, and gives me peace, thank you Jesus for making a way in the wilderness and rivers in the desert, to you alone be the glory.

The LORD is my light and my salvation;
whom shall I fear?
The LORD is the stronghold of my life;
of whom shall I be afraid?

Psalm 27:1 (ESV)

Contents

1 Introduction **1**

1.1 Motivation 1
1.2 Contributions 3
1.3 Publications 4
1.4 Thesis outline 5

2 Preliminaries **6**

2.1 Introduction 7
2.2 The need for a robust sensor 7
2.2.1 Classifying sensors 8
2.2.2 Frequency of operation on the electromagnetic spectrum . . 9
2.2.3 Operating in challenging environmental conditions 10
2.3 Ego-motion estimation in robotics 14
2.3.1 Vision-based methods 16
2.3.2 Laser-based methods 18
2.4 Sensing with radar 20
2.4.1 Frequency-Modulated Continuous-Wave radar 20
2.4.2 Compiling a radar scan 21
2.4.3 Radar configuration settings 22
2.4.4 Complexities in radar measurements 23
2.5 Navigation with radar 29
2.5.1 Applications of radar in autonomous systems 29
2.5.2 Radar odometry 32
2.6 Components of MRG Radar Odometry 34
2.6.1 Landmark extraction 34
2.6.2 Point descriptors 37
2.6.3 Unary match assignment 37
2.6.4 Selecting correspondences 38
2.6.5 Computing rigid-body motion 40
2.7 Identifying areas in need of improvement 40
2.7.1 Using MRG RO as a baseline 41
2.7.2 Building on the MRG RO foundation 41

3 Pre-filtering radar scans for faster ego-motion estimation 44
3.1 Overview of contribution . . . 44
3.2 Integrated manuscript . . . 45
3.3 Further insights . . . 54
3.3.1 Choosing what to label . . . 54
3.3.2 Odometry source for automatic annotation . . . 55
3.3.3 Further discussion on original results . . . 56
3.3.4 Alternative baselines alongside the learnt approach . . . 57
3.3.5 Updated related work . . . 58
3.4 Contribution review . . . 59

4 Introspective radar odometry 62
4.1 Overview of contribution . . . 62
4.2 Integrated manuscript . . . 63
4.3 Further insights . . . 74
4.3.1 Why use a spectral correspondence technique for radar? . . . 74
4.3.2 Effects of scene symmetry on correspondence success . . . 75
4.3.3 Revisiting the Kalman filter component . . . 76
4.4 Contribution review . . . 82

5 Leveraging a motion model to increase ego-motion accuracy 85
5.1 Overview of contribution . . . 85
5.2 Integrated manuscript . . . 86
5.3 Further insights . . . 97
5.3.1 Selecting matches using bounds for θ-values . . . 97
5.3.2 Attempting to refine match selection . . . 99
5.3.3 Relevance to alternative RO systems . . . 99
5.4 Contribution review . . . 100

6 Conclusion 103
6.1 Summary of contributions . . . 103
6.2 Future work . . . 105
6.2.1 Velocity-aware automatic annotation . . . 105
6.2.2 Pre-filtering radar scans for other tasks and settings . . . 106
6.2.3 Reducing the sensitivity of the introspection system . . . 106
6.2.4 Integrating an external odometry fallback . . . 107
6.2.5 Introspection for topometric localisation . . . 107
6.2.6 Sparse odometry with match refinement . . . 108
6.2.7 Using individual match quality in a learnt approach . . . 108
6.3 Closing remarks . . . 108

References **110**

Abbreviations **119**

1 Introduction

Contents

1.1 Motivation 1
1.2 Contributions 3
1.3 Publications 4
1.4 Thesis outline 5

1.1 Motivation

Autonomous robots are becoming increasingly capable of navigating through complex environments due to recent advances in vision and laser-based ego-motion estimation systems. As these machines begin to see use in a wider variety of real-world applications, the limitations of our sensors and algorithms are more often exposed. In particular, reliance on a small subset of sensing technologies makes it difficult, if not dangerous, to operate autonomous systems in places where lighting is harsh or absent, or where rain, fog, snow, or dust can obscure the scene.

This thesis seeks to expand the operational domain of autonomous robots by using a radar to estimate ego-motion, enabling navigation in settings where traditional sensors fail. By relying on Frequency-Modulated Continuous-Wave (FMCW) millimetre-wave (MMW) radar alone, we can focus on what can be achieved



Figure 1.1: Autonomous navigation in challenging environments requires a sensor that is robust when operating in poor lighting, weather, and other difficult environmental conditions. Clockwise from top left: data collection in Iceland, wading through a section of an off-road 4x4 course in rural Buckinghamshire, surveying a solar farm in Oxfordshire with a Husky robot, and trials at Blenheim Palace on the Hulk platform [1].

when other sensors may be unreliable, and equip autonomous robots with knowledge of where they are and what surrounds them in unknown environments without external infrastructure, even in challenging environmental conditions. Figure 1.1 depicts four of the robots from the Oxford Robotics Institute that have been equipped with radar to sense their surroundings, operating in a wide variety of locations.

However, radar returns are complex — measurements contain rich information that can be challenging to interpret, particularly when this sensor is mounted on a moving vehicle. Previous approaches have relied on purpose-built infrastructure like radar beacons or have attempted to treat radar as if it were a Time-of-Flight (ToF) laser, discarding useful information in the process of attempting to get past these difficulties. Recent work on Radar Odometry (RO) [2] has opened up the possibility of using radar for robust ego-motion estimation, but without sufficiently considering the context in which this sensor operates.

The work presented in this thesis leverages that under-utilised context and asks: "How can we use information at a systems-level to improve radar-only ego-motion estimation?" Answering this question leads us to take a systems-driven approach in developing informed algorithms that are faster, introspectively aware, and produce estimates with greater accuracy than was possible before.

1.2 Contributions

This thesis makes several contributions to radar-only ego-motion estimation that use a systems-based approach to improve speed, robustness, and accuracy. By leveraging the context in which robots find themselves to inform specific components of the system, the following contributions are made:

- An automatic annotation pipeline that allows for rapid labelled data generation to identify stable portions of the scene
- A method for pre-filtering radar returns to pass only static structure in a scene from a moving vehicle
- An introspection framework to determine the reliability of a final scan match between radar landmark sets for either ego-motion or localisation
- A fast, portable technique to classify the feasibility of individual landmark matches between scans by leveraging knowledge on the dynamic constraints

These contributions extend the capability of the original Mobile Robotics Group (MRG) RO pipeline [2], interfacing with the components of the system at different points as illustrated in Figure 1.2. A more detailed overview of the existing subsystems and their functionality is provided in Section 2.6, with only a high-level perspective shown here to contextualise the components presented in this thesis.

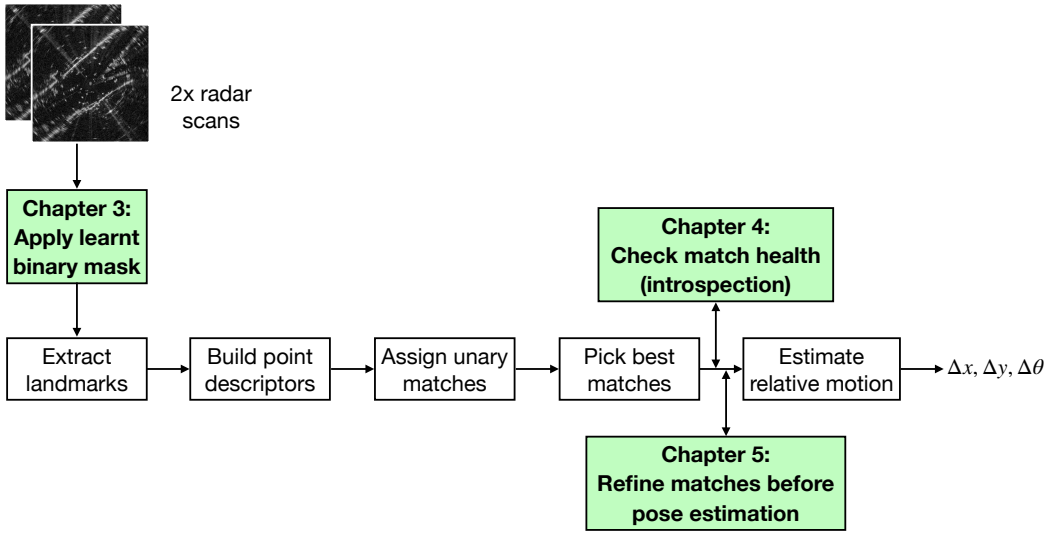


Figure 1.2: An overview of the baseline MRG RO [2] system, indicating where each of the core contributions interfaces with the existing components.

1.3 Publications

The contributions in this thesis are centred on three publications that each target a specific component of the MRG RO pipeline [2]. They are briefly outlined below:

- A method to pre-filter radar scans to remove spurious returns is presented in "Fast Radar Motion Estimation with a Learnt Focus of Attention using Weak Supervision" [3]. By learning how to mask out radar complexities, this technique more than halves the processing time with negligible impact on odometry accuracy.

- Adding introspection to the data association task and using this to inform a simple motion model is presented in "What Could Go Wrong? Introspective Radar Odometry in Challenging Environments" [4]. Having a means of "knowing when we don't know" allows a confidence in the data association result to be determined and reduces odometry estimate failure occurrence by almost 25%.

- Assessing the plausibility of individual matches by using knowledge of the platform constraints is detailed in "What Goes Around: Leveraging a Constant-curvature Motion Constraint in Radar Odometry" [5]. By considering the

underlying motion model of this system, the final pose estimate error can be reduced by over 45 %, compared to when these constraints are not utilised.

1.4 Thesis outline

This thesis details how the contributions in each of the aforementioned publications drive fundamental improvements to a radar-only ego-motion estimation pipeline. Each component is aimed at a specific area in need of improvement: speeding up processing time, adding introspective capability in challenging conditions, and finally, refining ego-motion estimates themselves to reduce odometry errors.

Chapter 2 outlines the necessary preliminaries, detailing the context in which radar finds itself as a sensor in a sector traditionally dominated by vision and Light Detection And Ranging (LiDAR) systems. Understanding the the capabilities and vulnerabilities of this sensor and the diversity of previous applications gives necessary background in which to situate more recent related work, in particular that of our baseline MRG RO system. An overview of each component in this particular RO implementation provides background information that aids in understanding where significant and novel improvements could be made, and how the core contributions presented in this thesis interface with these related needs.

Chapters 3 to 5 present the three contributions that detail the approaches to enhance the speed, reliability, and accuracy of the MRG RO pipeline, and include further insights and discussion not included in the original manuscripts. Finally, Chapter 6 discusses these advances and avenues for improvement, followed by a conclusion.

2
Preliminaries

Contents

2.1 Introduction 7
2.2 The need for a robust sensor 7
2.2.1 Classifying sensors 8
2.2.2 Frequency of operation on the electromagnetic spectrum 9
2.2.3 Operating in challenging environmental conditions . . . 10
2.3 Ego-motion estimation in robotics 14
2.3.1 Vision-based methods 16
2.3.2 Laser-based methods 18
2.4 Sensing with radar 20
2.4.1 Frequency-Modulated Continuous-Wave radar 20
2.4.2 Compiling a radar scan 21
2.4.3 Radar configuration settings 22
2.4.4 Complexities in radar measurements 23
2.5 Navigation with radar 29
2.5.1 Applications of radar in autonomous systems 29
2.5.2 Radar odometry 32
2.6 Components of MRG Radar Odometry 34
2.6.1 Landmark extraction 34
2.6.2 Point descriptors 37
2.6.3 Unary match assignment 37
2.6.4 Selecting correspondences 38
2.6.5 Computing rigid-body motion 40
2.7 Identifying areas in need of improvement 40
2.7.1 Using MRG RO as a baseline 41
2.7.2 Building on the MRG RO foundation 41

2.1 Introduction

This chapter establishes the necessary context for radar's use in mobile robotics as a navigation sensor. Related work is presented alongside core concepts where relevant, and provides background on the different approaches to problems in this domain.

Radar remains a relatively underutilised sensor in mobile robotics, but has begun to see more widespread use in tasks previously dominated by vision and LiDAR systems. Section 2.2 describes how sensors are classified, the factors that determine their capabilities, and the limitations they face in adverse environmental conditions. After covering these important characteristics, Section 2.3 looks at how the task of ego-motion estimation has been approached with vision and LiDAR systems in the past, and provides the context necessary to understand how the radar-specific themes in this thesis relate to the concepts in more established techniques.

Radar offers a unique modality of sensing, and there are a number of important considerations to understand when using it in a mobile robotics task, which are discussed in Section 2.4. Following this, the applications of radar for ego-motion estimation and the operating principles of our baseline MRG RO system are outlined in Section 2.5 and Section 2.6 respectively, to identify its prior limitations and motivate the need for the contributions presented in this thesis.

2.2 The need for a robust sensor

Mobile robots perceive everything about their surroundings and internal state through their sensors, which makes the reliability of these devices critical to success. In particular, tasks in challenging conditions or environments can become difficult, or simply infeasible, when sensors fail to function in these circumstances. This thesis makes use of radar as a sensor to offer solutions to issues frequently encountered in autonomous navigation. Radar is able to avoid the difficulties that vision and laser sensors face when operating in adverse weather and lighting conditions primarily because it has a different mode of operation. To contextualise this, we must first

consider the different classes of sensors and their sensing frequencies, before looking at how this impacts their ability to perceive the environment.

2.2.1 Classifying sensors

Sensors are formally classified by whether they are designed to measure internal or external quantities, and whether they do this independently or by relying on external sources. *Exteroceptive* sensors are defined by their ability to observe the surrounding environment in which a robot may find itself through light, sound, or range measurements. This is in contrast to *proprioceptive* sensors that capture information about the robot's internal state like motor shaft angles, temperature, or acceleration. Sensors can also be classified as either *active* or *passive*, determined by whether they rely on signals already present in the environment or emit their own energy and measure a response. The use-case is important to consider when selecting a specific class of sensor, as their different sensing mechanisms provide certain strengths and weaknesses that are exemplified by scenarios discussed in the remainder of this subsection [6].

Exteroceptive sensors like cameras, LiDAR, and radar are well-suited for use in ego-motion estimation tasks, as they leverage the strong prior that the surrounding environment is static — allowing us to assume that the dominant perceived motion of selected features has arisen from ego-motion alone. On the contrary, proprioceptive sensors like encoders measuring rotations of an axle for wheel odometry are prone to errors introduced by wheel slip or other difficult road conditions, and are not feasible in the air or at sea [7]. While passive sensors such as cameras are usually available at a lower relative cost, active sensors that do not rely on external signals to obtain reliable measurements, like LiDAR and radar, are often more suitable in difficult lighting and weather conditions. Before discussing these attributes in greater detail, it is important to consider the frequency of electromagnetic waves at which sensors operate and how this determines their capabilities.

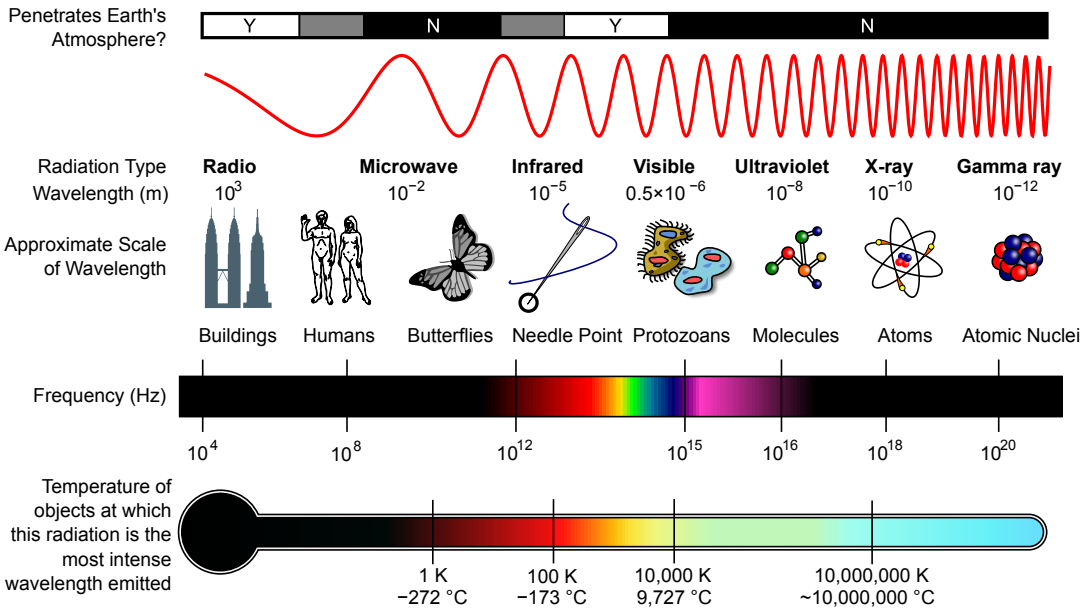


Figure 2.1: A signal's frequency along the electromagnetic spectrum determines how it interacts with the environment and gives each sensing modality its own unique characteristics[1].

2.2.2 Frequency of operation on the electromagnetic spectrum

For systems that make use of vision, LiDAR, sonar, or radar, the frequency of electromagnetic radiation used to sense the environment determines their capabilities, and can be understood by considering where they operate on the electromagnetic spectrum (shown in Figure 2.1). Optical sensors like cameras and lasers operate at wavelengths around the visible and infrared range respectively, typically between 380 nm to 750 nm for vision, and 1.5 µm to 10 µm for LiDAR systems. This allows them to measure relatively fine detail in the scene, but also makes them prone to attenuation by airborne particles.

Radar and sonar on the other hand operate at lower frequencies, usually with wavelengths in the range of 3 mm to 30 cm for radar and 1.5 cm to 1.5 m for sonar devices, although these figures vary considerably depending on the application [8], [9]. These longer wavelengths allow them to sense beyond the limitations of optical

[1]Wikimedia Commons page on EM Spectrum. Released under licence CC BY-SA 3.0, `https://commons.wikimedia.org/wiki/File:EM_Spectrum_Properties_edit.svg`

sensors, but at the cost of not being able to describe the environment in as high a resolution as would be possible with a shorter wavelength. The characteristics of sensors operating at different portions of the electromagnetic spectrum are particularly relevant when considering how they perform in realistic conditions outside of a controlled environment.

2.2.3 Operating in challenging environmental conditions

Deploying mobile robots in the real world forces them to face environmental factors that are often out of their control, and that can present difficulties when they are attempting to navigate, perceive, or plan their way through the world. This section looks at how lighting, weather, and other factors affect sensors.

Sensing in poor lighting

As mentioned previously, passive sensors like cameras have difficulty functioning in sub-optimal lighting conditions. Direct sunlight, shadows, darkness at night, or environments that are generally poorly lit present challenges for vision that limit the utility of systems that rely on these devices. Images from a vehicle-mounted camera that were captured at night and in direct sunlight are shown in Figures 2.2a and 2.2b, illustrating these difficulties. In contrast, active sensors like LiDAR, sonar, and radar are unperturbed by these external effects, as they generate their own transmission signals instead of relying on emissions in favourable environmental conditions. However, there is a caveat for LiDAR sensors, in that they are affected by harsh sunlight, although in practice this is not as significant a problem as it is for vision [10].

Sensing in harsh weather

Operating reliably in poor weather is difficult. Particles like fog, dust, rain drops, snowflakes, or spray from other vehicles are usually between 1 µm and 5 mm in diameter and will obstruct signals with a wavelength that is shorter than this [9]. This is problematic for optical sensors as they tend to have wavelengths on this scale, and so are prone to suffering significant attenuation and occlusion. Figures 2.2c



(a) Night, urban **(b)** Low sun, urban **(c)** Rain, off-road **(d)** Snow, urban

Figure 2.2: Relatively common adverse lighting and weather conditions make ego-motion estimation difficult for vision. This motivates the case for alternative sensors which avoid these failure modes.

and 2.2d show how cameras are not only vulnerable to being blinded by poor weather conditions like rain, fog, and snow at a distance [11], [12], but that it is often the droplets and snowflakes accumulating on lenses that are problematic. Similarly, LiDAR is also unreliable in poor weather [13]–[16], and can be observed qualitatively alongside MMW radar data in Figure 2.3.

The RADIATE dataset [17] includes data collected in these challenging conditions, and also provides quantitative evidence of performance degradation in object detection during poor conditions when using optical sensors as opposed to radar. For robot navigation specifically, Hong et al. [18] show their radar Simultaneous Localisation and Mapping (SLAM) system's strong performance in extreme weather conditions, which gives an indication of this sensor's reliability, even while a large portion of the scene is obscured due to the front of the radar being partially encased in snow (see [18] for relevant figures).

Additional benefits of sensing at a longer wavelength

Besides reliable operation in challenging lighting and weather conditions, sensing at longer wavelengths has some additional benefits too. For example, deploying mobile robots in dusty environments can be challenging, not only due to the presence of dust clouds or trails created by vehicles that obscure an optical sensor's view (as shown in Figure 2.4), but also because of the need to keep sensor lenses clean. Unlike vision and LiDAR which are both prone to this problem, a radar's sensitive components are usually protected in a secure, fixed housing that keeps particles out while still allowing radar signals to pass through the material. This avoids the

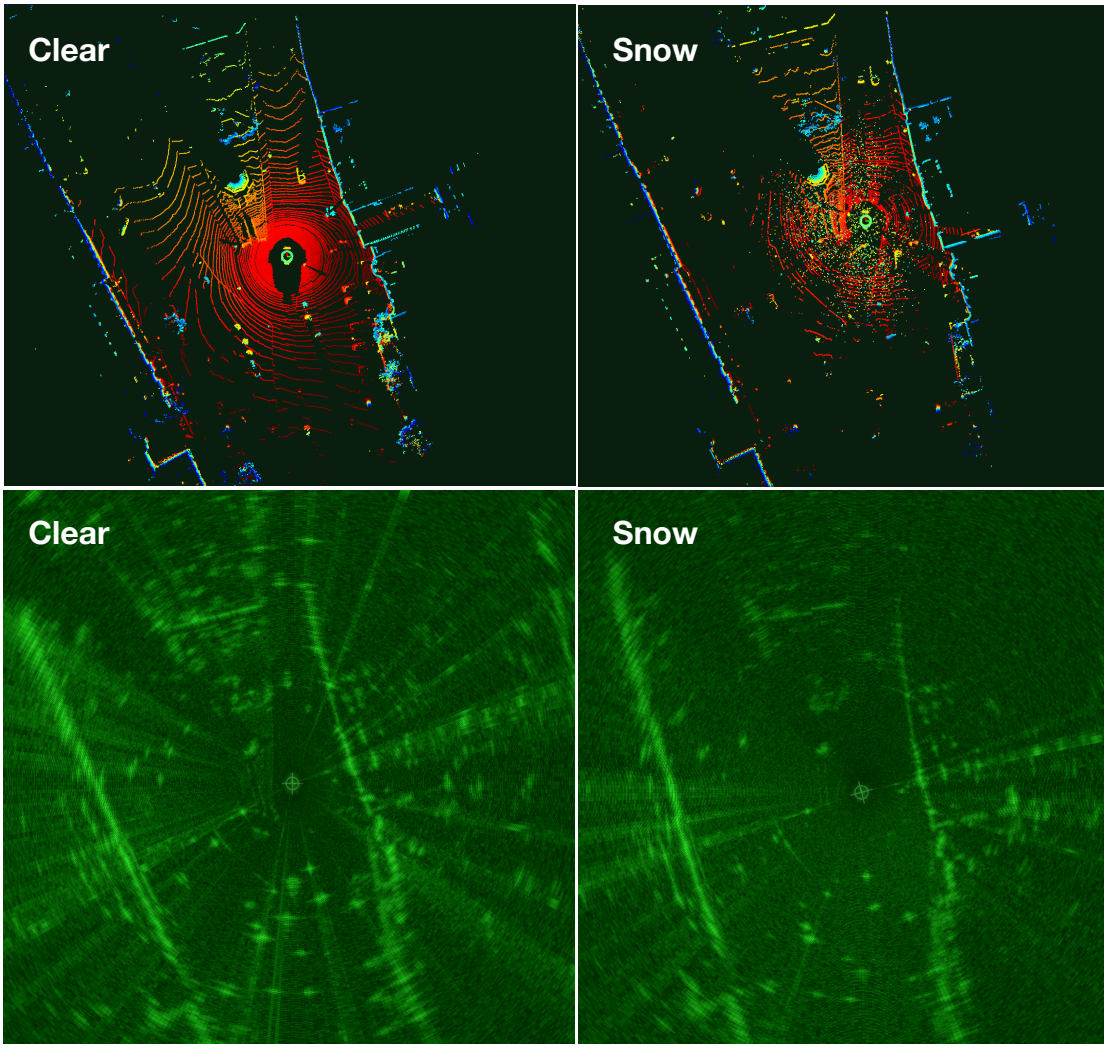


Figure 2.3: LiDAR is susceptible to snow due to the wavelength that it uses to sense the environment. Here, we see a qualitative comparison between LiDAR and radar on a clear day as compared to a snowy day when passing the same location in Oxford. The radar scans show little change in appearance, bar some sensor artefacts discussed in Section 2.4.4, and with power returns only marginally attenuated — visible as a slight fading of the walls in the snow scan. On the contrary, the LiDAR returns are clearly detecting the nearby falling snowflakes, which are also partially occluding sections of the scene at greater ranges evident along the roads. This can be problematic for navigation and perception algorithms as these undesirable returns add further noise to the LiDAR data stream that would need to be mitigated, and obscure other objects in the vicinity. For another example of LiDAR in the snow compared with radar, see the related figures from the RADIATE dataset [17].

need to wipe sensors when dust or water droplets have accumulated, although the latter can also be mitigated in software for cameras in the rain [19].

Ultimately, the fact that radar operates at a longer wavelength means that these difficulties posed by dust, poor weather, and unfavourable lighting are avoided. This provides a significant advantage over the vision or LiDAR systems so often relied upon in robotics to perform ego-motion estimation, localisation, and perception tasks. Radar sensors are still in relatively active development and are continuing to improve in their functional sensing range, operating frequency, range resolution, and angular resolution. If previous trends in vision and LiDAR are repeated, the cost of future radars should make them more accessible too. But in the meantime, some early adaptors of low-cost radars have used them to successfully augment LiDAR systems, allowing tasks to be completed in adverse conditions that would otherwise not have been possible with traditional sensors alone [20], [21].

Another advantage of using these longer wavelengths is the increase in functional sensing range. Aside from requiring less transmitted power to operate, a radar bathes the scene in electromagnetic energy as opposed to a LiDAR's highly focused laser beam transmissions, and so is more likely to detect subtle distant features that might otherwise occur between two neighbouring azimuths. This is particularly important when navigating feature-sparse environments like rural Iceland as in Figure 2.5, and notably, was an issue which was mentioned as one of the four areas requiring improvement following Visual Odometry (VO)'s implementation on Mars [22]. Closer to home, this limitation can also raise safety concerns when there is a need to detect obstacles sooner when moving at high speeds than would be possible with vision or LiDAR [23]. And unlike sonar, radar's operating wavelength is still short enough to provide reasonable resolution, making it feasible to use in addressing some of these fundamental tasks in robotics.

Before focusing on radar as a sensor that can be useful for navigation, it is important to give an overview of the established ego-motion estimation methods that have been refined over decades in vision, and then look at how laser-based

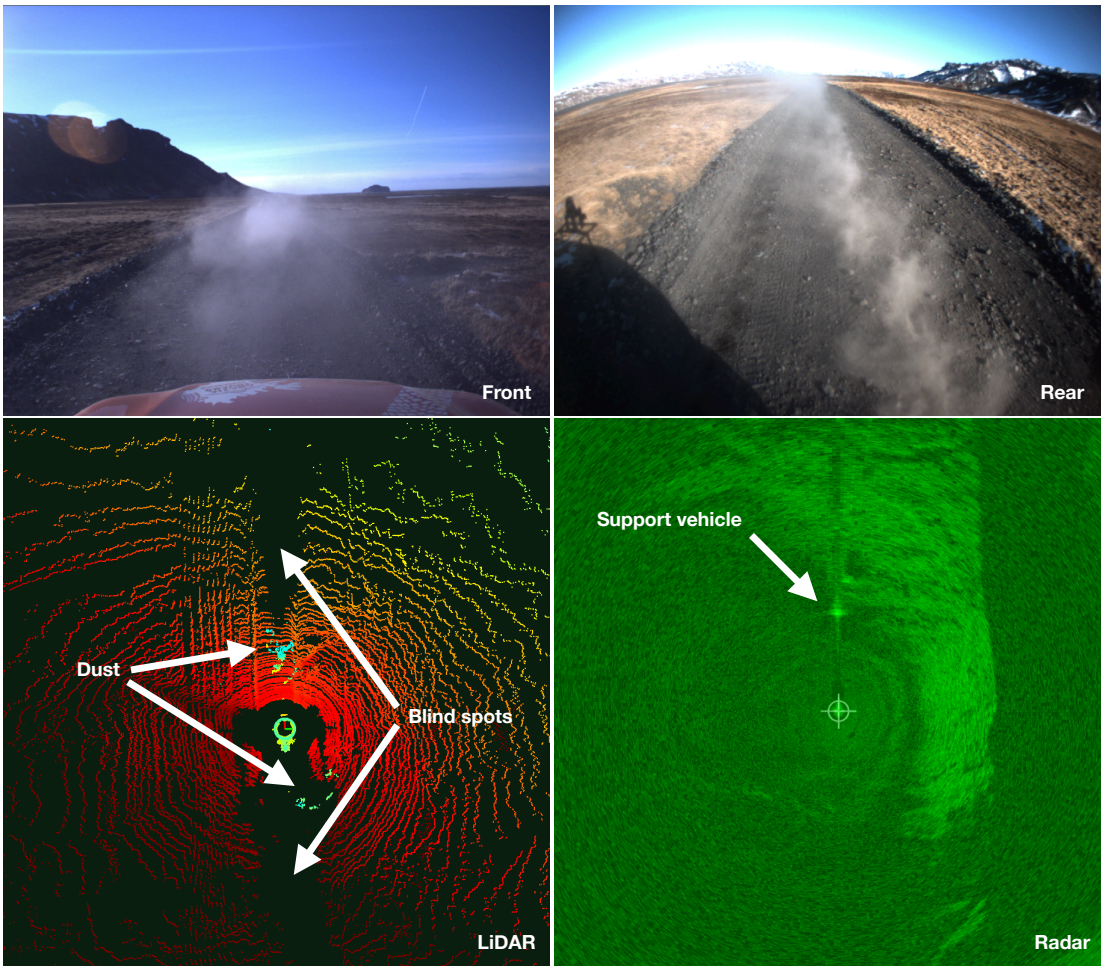


Figure 2.4: Dust is a constant reality when driving along off-road routes, evident here in an ORI dataset we captured in Iceland. The top two images show the front and rear camera views, with the corresponding LiDAR and radar scans below. The support vehicle is driving ahead, obscured visually in the camera view by the dust it has thrown up. The LiDAR beam is reflecting off this dust cloud and registering the dust itself as obstacles ahead and, to a lesser degree, behind the vehicle, while also being blinded in the rear view by the dust trail created by the vehicle on which it is mounted. Lastly, the radar scan shows a clear return from the support vehicle ahead, and no disturbances from either the dust ahead or behind the vehicle. This radar scan has been cropped to make the comparison to LiDAR clearer, but is only approximately to the same scale.

approaches solving the same core problem have had to approach it differently because of the key differences between these two popular robotics sensors.

2.3 Ego-motion estimation in robotics

Ego-motion estimation in mobile robotics typically needs to be within a few centimetres of ground truth to meet the levels of accuracy required to navigate safely

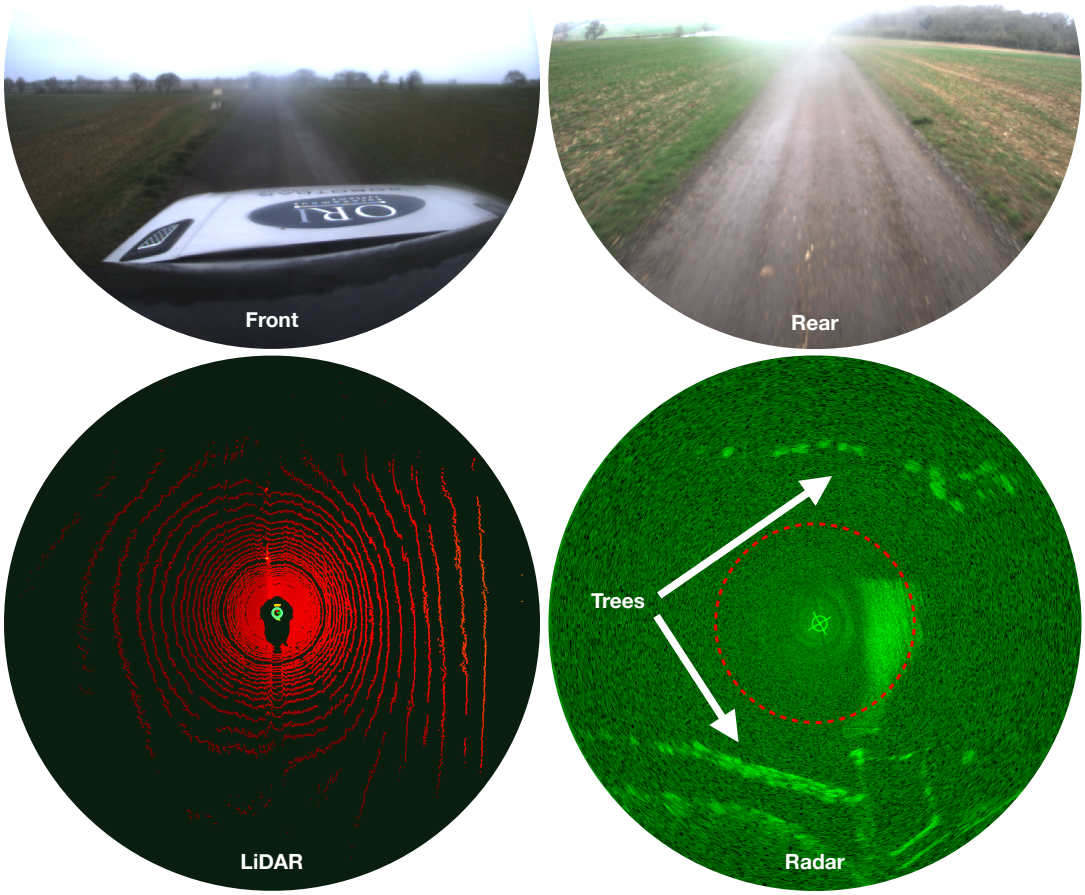


Figure 2.5: Observations from front and rear cameras, as well as the LiDAR sensor, fail to detect distinctive features in wide open areas. Conversely, the superior range of radar allows it to observe the wooded area behind the vehicle as well as the row of trees up ahead, providing a reliable set of features to use when navigating. The cameras used here have a wider field-of-view compared with those in Figure 2.4, and both the LiDAR and radar scans have been slightly cropped down from their full range for clarity. The dotted red circle on the radar scan indicates the approximate sensing range of the LiDAR, which is unable to observe the distant scene structure that is useful for navigation.

and effectively, and is usually performed with cameras or LiDAR sensors. This section aims to outline how this task has been approached in the past using these technologies.

2.3.1 Vision-based methods

The use of cameras as a robotics sensor is an obvious and popular choice, owing largely to their low cost, information richness, and the ease with which data can be interpreted, debugged, and processed. Unlike using an encoder to measure wheel rotation for an odometry estimate, a visual system provides a better sense of relative motion that is robust to wheel slippage, and functions just as well in aerial or aquatic domains inaccessible to regular wheel odometry.

Visual Odometry (VO) is arguably the most prominent pose estimation system in the vision context, having been in some form of development since the 1980s in the computer vision community as part of work on structure from motion (SFM), prior to the seminal publication by Nistér et al. where the term VO was coined [7], [24]. The components of a VO system are usually tasked with taking a pair of images captured at two different poses, detecting features, matching these features between observations, and then estimating the motion that occurred from the first pose to the second. Usually at each pose, a pair of images is captured from two cameras that are a fixed distance apart to allow for scale to be inferred in the scene. This is known as stereo VO, but the same result can be achieved with a single camera (monocular VO) if needed, requiring a few additional steps to recover the scene scale [24].

For feature-based VO, features need to be extracted from a given image such that they are distinctive and easy to identify from a nearby pose. These methods have become more popular compared with earlier correlation-based frameworks where a feature is tracked into the next image, which are of limited use when the distance between poses increases as is the case in more realistic outdoor applications [25], or when performing localisation with a prior map. Feature extraction requires some form of corner or blob detector, which is used to identify features that are likely to persist between frames due to their distinctive texture, intensity, or colour [6]. Some

of the more popular algorithms for feature detection include SIFT [26], FAST [27], SURF [28], and Harris [29]. These are specifically designed for use in the vision context, and depending on the operating environment (indoor, outdoor, urban, rural, aerial) a selection is made to perform well in a particular setting. These features then need to be expressed as unique descriptors, prior to a matching step. They typically rely on the richness present in images, where pixels surrounding a feature encode information that persists between observations. A detailed description of these algorithms is outside the scope of this thesis, but a summary of their specific functions can be found in [6], [25].

With feature descriptors in hand, the next step involves creating links between features in different images through a data association step. This can be done by finding the Euclidean distance between feature vectors, or more elegantly, to avoid the computational load of checking every possible feature comparison, a region around the feature can be searched for the closest correspondent (in feature space). Once associations have been computed, a refinement process is required to reject outliers that are likely present due to image blur, variable illumination, occlusions, or other complexities. Random Sample Consensus (RANSAC) [30] is the established method for removing these correspondences in VO, and has been shown to improve accuracy significantly [25].

To estimate the relative pose between observations, one final step is required that takes the inlier set of feature correspondences and computes the underlying motion. This relies on the construction of an essential matrix which captures the epipolar geometry constraints, and relates the translation and rotation between two camera positions to each other. Using singular value decomposition (SVD), it is then possible to solve for the final pose estimate, as detailed further in [7]. In addition to estimating poses, these systems also often include some form of pose graph optimisation which seeks to refine estimates [31], and can use loop closures to add valuable constraints when revisiting the same location if navigating within a map [32]. It is also worth mentioning windowed bundle adjustment, which attempts

to do this task locally, tracking features through more than just two frames and iteratively refining ego-motion estimates [25].

With over 40 years of research progress behind them, vision-based approaches to estimating ego-motion are impressively capable. Modern VO is fast, accurate, and robust to many of the difficulties encountered in real-world robotics that previously limited our capabilities to indoor or offline use. Perhaps the best evidence for this statement is its implementation on Mars, where VO has been used on several missions [22]. More recently, deep learning methods have demonstrated new ways to improve on this mature area of research, where using end-to-end approaches allows us to avoid hand-tuning parameters or even sensor calibration entirely [33], [34]. The age-old trade-off between processing speed and accuracy has been disrupted by hybrid methods like DPC-Net, which rapidly processes sparse VO with a fast classical pipeline to prioritise computational efficiency at the cost of producing coarser estimates. These small errors that are introduced are then corrected for by a network, such that the end result achieves accuracy on par with dense methods that have a higher computational cost [35].

With an overview of vision-based ego-motion estimation methods in robotics having been covered, it is now time to discuss how this task is approached in laser-based systems, where many of the specific techniques from vision no longer apply.

2.3.2 Laser-based methods

The use of LiDAR in robotics has brought new capabilities to the field, providing range measurements that are robust to scene texture and lighting. There are different considerations to be taken into account when using this rotating range sensor, where instead of pixels in an image, measurements consist of a number of point returns that make up a scan. Compensating for the rotation rate of the sensor when mounted on a moving platform can be important, especially at higher speeds if the sensor's frequency of rotation is low [36], [37].

Similar to the approach of VO, an established LiDAR odometry method like LOAM seeks to extract features from scans, find correspondences between them,

and then solve for the relative motion that aligns these correspondences [38]. Some key differences exist, particularly when extracting features from these unevenly distributed raw points.

Usually once points have been processed by a motion distortion correction step, features are extracted from points that represent surface planar patches and sharp edges as these are likely to persist between scans. Other constraints can also be included, such as limiting the number of points in specified subregions, or excluding those that are on patches that risk running parallel to the laser's beam [38].

Some form of the Iterative Closest Point (ICP) method is commonly used to find the alignment between LiDAR scans, which performs well if the scans are initially well-aligned [39]. As alluded to earlier, addressing the need for velocity correction is often worthwhile in a mobile robotics context for a rotating range sensor, and improves ICP results when motion distortion is present [40].

Similarly, as in the vision context, learnt approaches have emerged to contribute novel solutions that avoid the discrete steps and hand-tuning of parameters required for classical methods [41]. For supervised systems like LO-Net [41], an extensive labelled dataset with an accurate ground truth signal is required to achieve satisfactory generalisation performance. Unsupervised methods, which avoid this requirement, are starting to emerge and although still in their infancy, present a viable alternative to previous classical approaches [42].

In an extension of the contribution made by LOAM in [38], V-LOAM uses a VO method for high frequency estimates with lower fidelity, which improves the original LiDAR-only method's ability to handle more aggressive manoeuvres by combining the strengths of both sensing modalities [43]. In practice these multi-sensor approaches tend to outperform systems that rely on only one sensing medium, particularly when complementing each other as in [43], as evidenced in odometry metrics for urban driving like the KITTI dataset [44].

However, vision, laser, and methods that combine both of these sensor types are still prone to failures that are not feasible to address by simply making our algorithms more adaptive. As covered in Section 2.2, radar offers answers to some of these

limitations, but also comes with its own set of challenges. The next section looks at our radar sensor in more detail, and discusses some of the particulars of its operation.

2.4 Sensing with radar

Radars have come a long way since their early applications in detecting distant aircraft and shipping vessels, where their primary task of target detection was achieved by locating large metal craft in an open setting. Modern radars are smaller, cheaper, and operate at sensing resolutions that allow them to detect ordinary obstacles and features in an operating environment and build up an image of a scene [45]. Having discussed the need for a sensor that is reliable in harsh weather and lighting conditions in Section 2.2, this section outlines how radars operate, and expands on some of the challenges that arise due to the complexity of radar data.

2.4.1 Frequency-Modulated Continuous-Wave radar

The radar used in our navigation context is a FMCW scanning radar. It operates by emitting a frequency-modulated radio wave that reflects off objects in the environment, with some of this reflection returning to the receiver [45]. The phase shift between the transmitted and received frequencies induces a beat frequency which is measured to infer target range. The power return observed by the receiver will depend on an object's Radar Cross Section (RCS), which is determined by its orientation, surface area, and material properties. A single reading therefore encodes considerable information regarding an object's geometry, relative orientation, and physical makeup which should be taken into account by downstream algorithms that need to extract relevant information from this rich stream of data [2].

Our application uses a Navtech CTS350-X FMCW scanning radar with a baseband operating frequency of 77 GHz, so as to benefit from a portion of the frequency spectrum that sees lower atmospheric attenuation [46]. This millimetre-wave (MMW) radar provides higher resolution measurements compared with other microwave radars operating at lower frequencies. As it rotates its antenna about a

vertical axis, the sensor's measurement stream can be processed as an image made up of rows and columns, which correspond to readings at specific angles and ranges.

The reader is directed to Adams et al. [45] for a thorough presentation of the operating principles of FMCW radar which have been summarised here. The next section discusses aspects of the imaging process that are relevant to our application when compiling azimuth readings into scans for use in the navigation task.

2.4.2 Compiling a radar scan

For each sweep, the FMCW radar processes m azimuths, which are one-dimensional power-range spectra made up of n range bins. These azimuth returns, each corresponding to a specific angular slice, are combined sequentially to create a full scan. The left panel of Figure 2.6 shows an example of a polar image produced from a single scan, where each row of pixels corresponds to an azimuth's power-range spectrum. Scans can also be interpreted in Cartesian coordinates as shown in the right panel of Figure 2.6, with each power-range spectrum making up a ray originating from a single central point. The coordinates corresponding to a Cartesian point (p_x, p_y) can be expressed in terms of the range r and bearing angle θ as:

$$p_x = r\cos(\theta)$$
$$p_y = r\sin(\theta)$$

which can also be interpolated bilinearly to create a Cartesian image [45]. However, as with any rotating range sensor, interpolating the sparse information at greater ranges creates a kind of "smearing" effect not present in the native polar representation. In practice, this is avoided by rendering returns as points rather than building a Cartesian grid of pixels, but it is worth noting so as to explain the effect seen at the edges of the left panel in Figure 2.6.

Once a scan has been compiled and expressed in a Cartesian form, it starts being useful for grid-based navigation as is used in this thesis. Much information is encoded in readings that span several azimuths and bins, as well as in the positions and orientations of objects in the scene with respect to one another. These

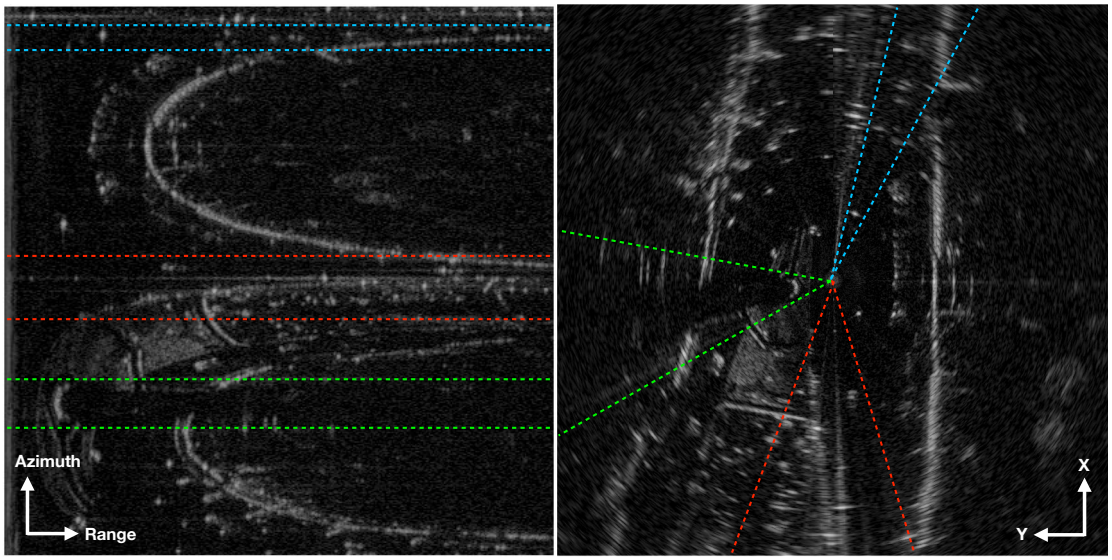


Figure 2.6: A portion of a complete radar scan is shown, cropped here to approximately the first 110 m of each azimuth so as to display some of the finer detail more clearly. In a polar radar image shown on the left, each azimuth is rendered as a row, with range increasing from left to right. The corresponding Cartesian radar image on the right gives a more intuitive birds-eye view of the scene, where the rectangular sections indicated in blue, red, and green in the left image correspond to wedges in the Cartesian projection on the right. Notice too that when interpolating these polar readings in Cartesian space, features tend to blur increasingly at greater ranges.

relationships between objects in the environment are the foundation on which many navigation solutions are built, as will be expanded upon in Section 2.6.

2.4.3 Radar configuration settings

The Navtech CTS350-X device is configurable to a certain extent in both hardware and software. The beam width can be adjusted by fitting different radar antennae, allowing for wider or narrower coverage of the scene for each azimuth. Two of these antennae were trialled, which provided either a 1.8° or 3.6° azimuthal width. The 1.8° antenna is used for all experimental data in this thesis, as it provided finer angular resolution where there was no clear benefit to using the additional azimuth overlap.

Additionally, although the number of bins in a single power-range return is fixed (currently a maximum of 3768 for this sensor), it is possible to set the range resolution for each bin to either 0.0438 m or 0.1752 m to adjust the level of detail that is captured. However, increasing bin resolution will decrease the sensor's overall

range because of this upper limit on the number of bins.

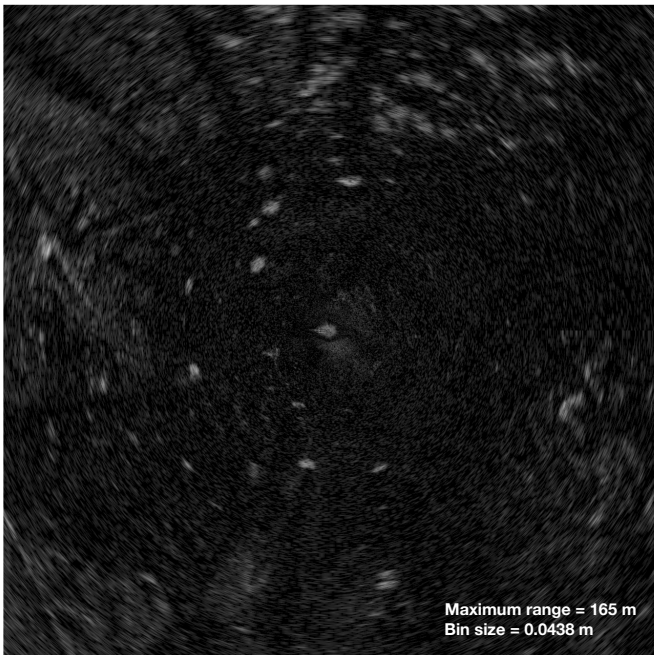
Figure 2.7 shows two radar scans captured in the same location along a route we traversed through a boulder field in Thórsmörk, Iceland. This provides an example of an environment in which it is advantageous to use the lower resolution configuration, so as to maximise the radar's range and capture more of the scene layout in this wide open space. On the contrary, Figure 2.8 presents a setting in which the higher precision offered by the high-resolution configuration may be more appropriate, as is usually the case in urban areas where features are usually much closer to the vehicle, while those that are further away are likely to be obscured.

The rotation rate of the sensor is nominally set to 4 Hz, capturing a complete scan of the scene every 250 ms. This could be increased to 8 Hz, but practically 4 Hz is adequate as it is fast enough to capture the static environment when the sensor is in motion at city driving speeds of up to 30 mph. Most of our data collections have taken place at these speeds, where there may be up to 3.3 m of translation between scans. Higher rates of rotation could prove advantageous in contexts where greater speeds are expected (up to 70 mph on highways), but the adverse effects of high-speed motion can also be mitigated in software through some form of velocity correction, as detailed in related work by Burnett et al. [47], or in the LiDAR-based solution by Hong et al. [40]. Additionally, it may also be beneficial to favour a higher rate of rotation when the task involves detecting or tracking dynamic objects in the scene.

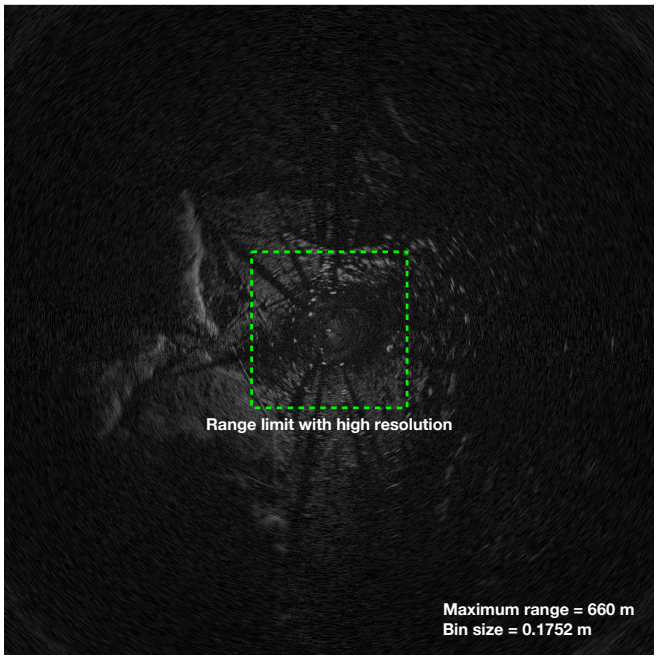
Before moving towards using the information encoded in radar scans, it is important to understand the difficulties that must be considered when working with this sensor, as discussed next.

2.4.4 Complexities in radar measurements

Interpreting data from radar is inherently complex due to the way in which observations are made, creating difficulties when trying to distinguish surrounding objects from the raw measurement stream. Effects like speckle noise, saturation, and atmospheric attenuation [45] need to be considered and mitigated — the

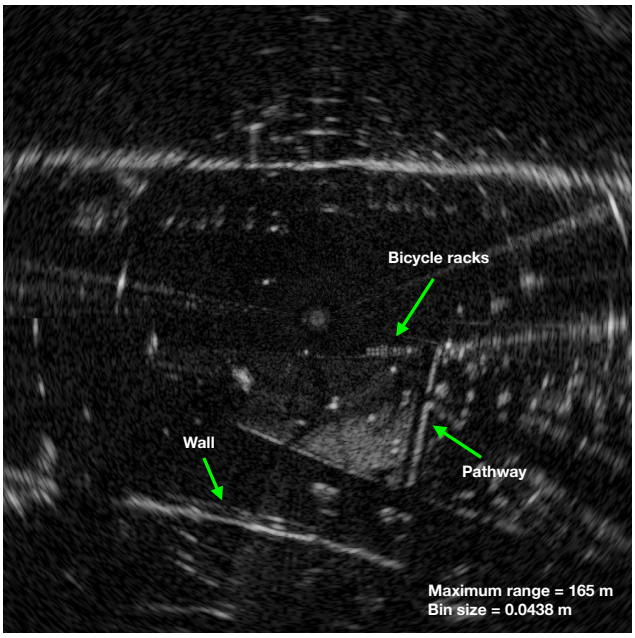


(a) High-resolution, low-range configuration

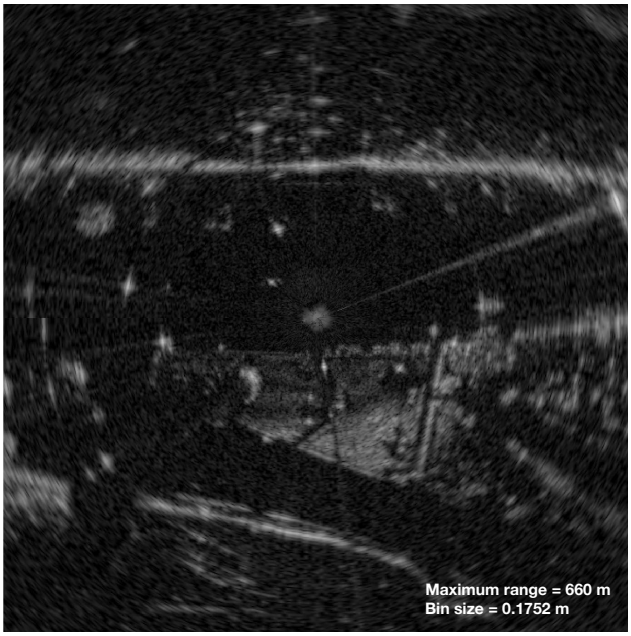


(b) Low-resolution, high-range configuration

Figure 2.7: The resolution per bin is configurable for the Navtech CTS350-X radar, shown here for the two settings of 0.0438 m and 0.1752 m in the same location while driving through a boulder field in Thórsmörk, Iceland. A dotted green box shows the extent of the environment observed in the high-resolution (lower range) scan in the context of the low-resolution (higher range) scan. The boulders scattered around the central vehicle are observable in both scans, but the surrounding hillside and other stable features can only be captured when making use of the extended range configuration.



(a) High-resolution, low-range configuration



(b) Low-resolution, high-range configuration, cropped to scale of Figure 2.8a for comparison

Figure 2.8: As in Figure 2.7, two radar scans in the same location but with different range resolutions are compared with one another from a route in central Oxford, with the low-resolution scan shown at the same scale as its counterpart. In this urban scene, greater range is not as advantageous because buildings, vegetation, and other vehicles occlude distant features. In addition, this environment has many features useful for navigation that are close to the vehicle, making distant observations less important. The scans appear relatively similar, but note the indicated instances where objects in the scene are clearer — bicycle racks, walls, and the nearby pathway appear in sharper focus as their range has been resolved with greater precision.

frequency of radar's electromagnetic waves interacting with the environment make it inappropriate to treat it as a ToF laser, which uses a much shorter wavelength than MMW radar as already discussed in Section 2.2.2.

Signals are also prone to taking multiple return paths from a surface to the receiver, manifesting as "ghost" objects, as shown in Figure 2.9. These multipath reflections are problematic as they do not correspond to the real parts of the static scene that are useful for ego-motion estimation, path planning, segmentation, or other tasks performed by a mobile autonomous robot. They are often difficult to detect and identify as spurious effects, resulting in them being naïvely processed as landmarks.

Additionally, the main beam is flanked by side lobes that transmit power in unintended directions. Although this power is significantly lower than that of the main beam, these returns can project objects into the current azimuth that are actually either side of it, creating further confusion as to which returns in the scene correspond to the real environment. Figure 2.10 shows the effects of saturated azimuths that usually occur when nearby vehicles cause occlusions, such as buses that pass by regularly in urban settings. Also illustrated in Figure 2.10 is a phenomenon that occurs due to internal noise within the radar hardware itself, where certain harmonics can manifest as false returns at a constant range in every azimuth, creating a ring or occasionally multiple rings in ways that are not feasible to predict.

There is a mature understanding of these effects in the radar literature [45], but they remain somewhat unaddressed in mobile robotics, making the task of using radar as a navigation sensor in this context appear more daunting than it ought to be. There are a number of mitigation strategies that have been developed for radar in the past outside of the robotics context which are of some use here, but these methods often do not adequately reduce the complexities discussed. They are often best suited to detecting targets from "noise", rather than capturing a full scene in detail in the presence of these artefacts. There are new hurdles to clear when mounting a traditionally stationary sensor on a mobile platform in an uncontrolled environment, but this also presents opportunities to leverage what

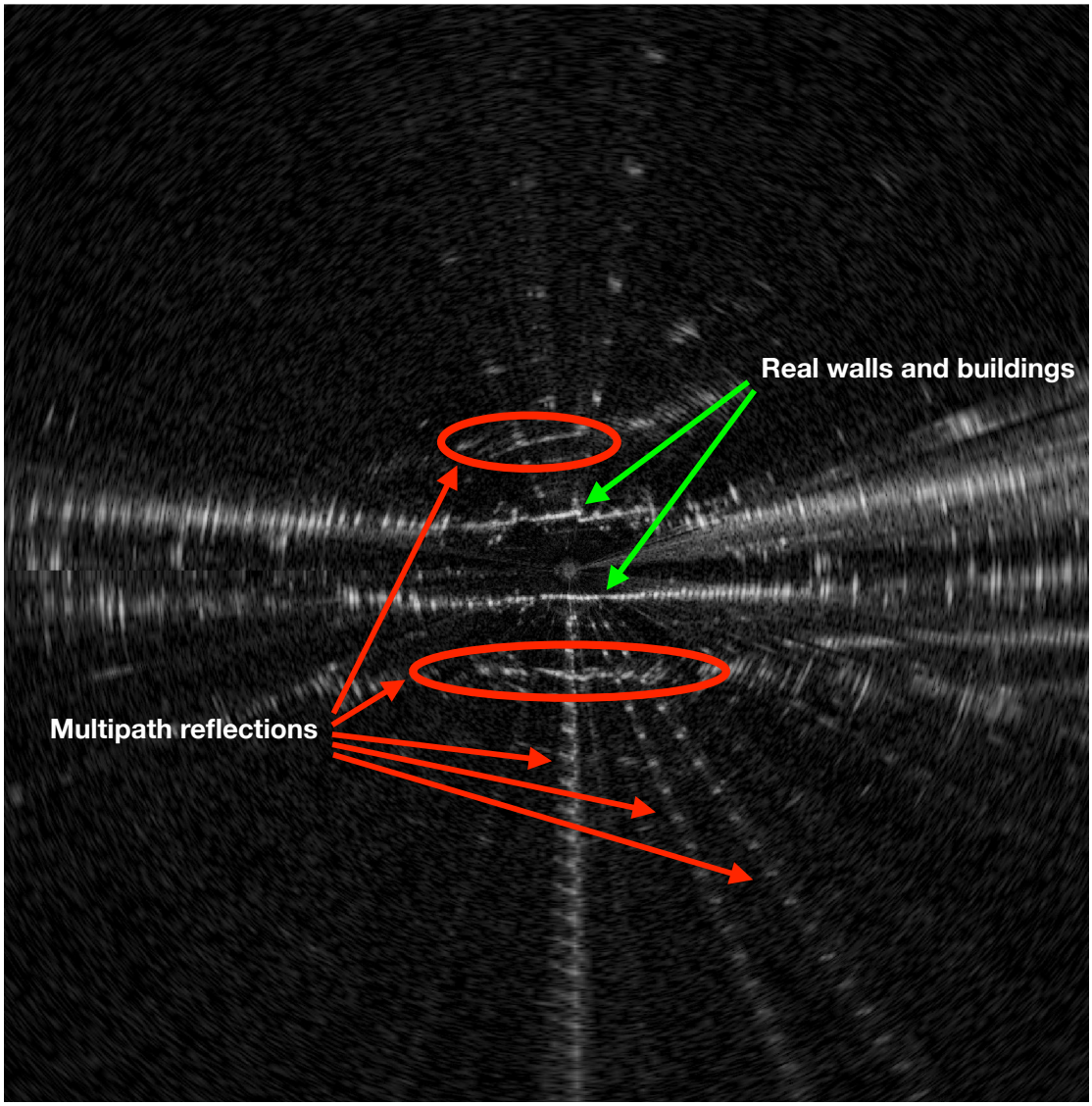


Figure 2.9: Multipath reflections occur when radar transmissions are able to reflect off objects in the scene and return via another route. In this scene, the robot is travelling along a street that is flanked by walls and buildings, as indicated. Some of the transmissions have rebounded off the walls to such an extent that a significant portion of a "ghost" wall has appeared, indicated in the two ellipses. Additionally, some of the azimuths that are particularly corrupted by these reflections are indicated, where a repeating power return appears at fixed intervals and adds further confusion to the scene.

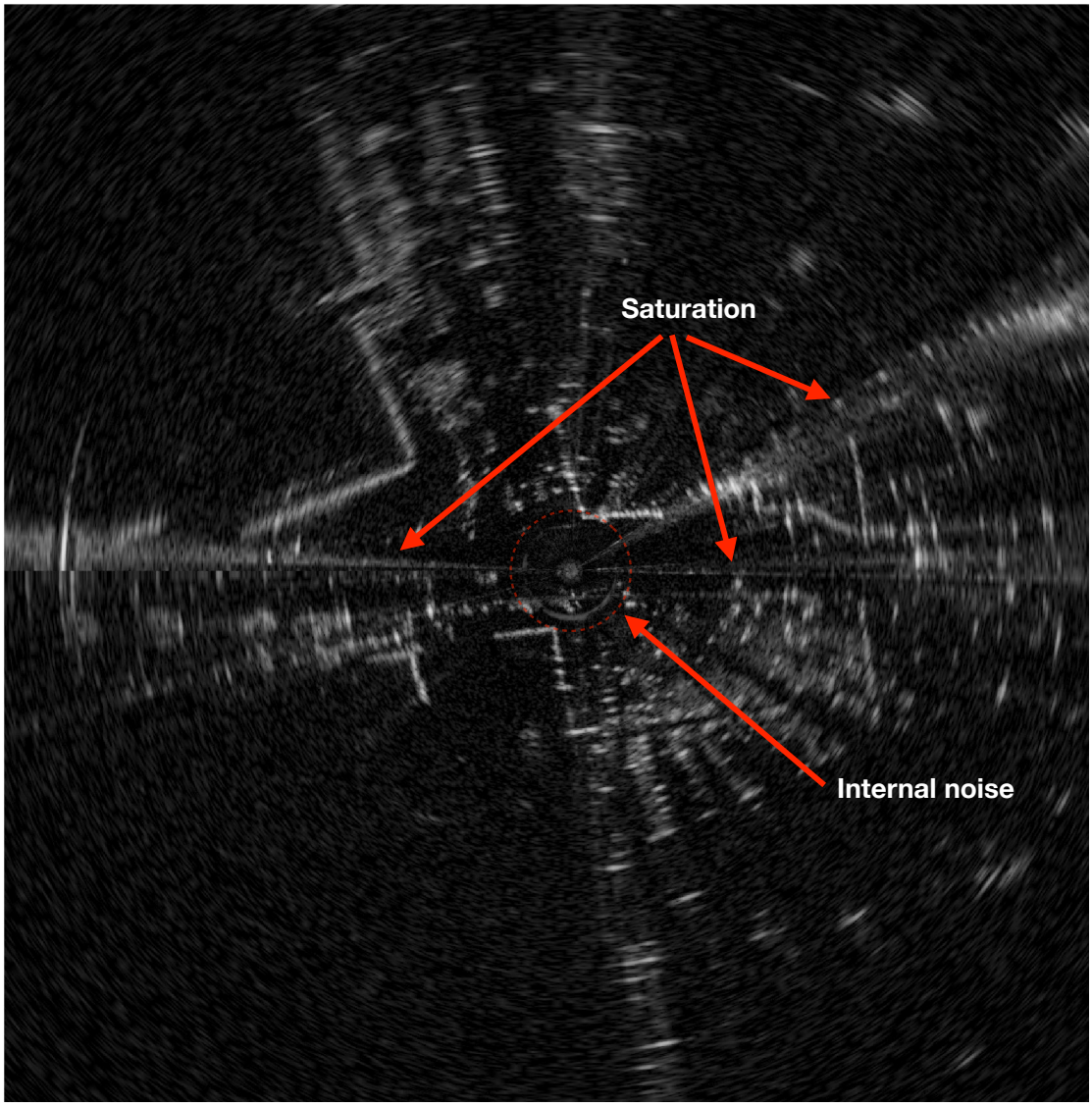


Figure 2.10: This radar scan shows two different kinds of artefact that can arise in radar data. Firstly, there are several azimuths that appear to be saturated, obscuring what might have occupied the space in the scene they were tasked with sensing. This usually occurs due to occlusions or reflections, particularly in this urban setting where other vehicles and buses regularly pass nearby. Secondly, a faded ring can be seen around the radar at the centre of the scan just inside the dotted ring annotation. This is a result of noise within the radar device itself that manifests as a spurious power return at a fixed range, creating the ring effect. Occasionally two or three rings occur at fixed ranges, although in practice, saturation and the multipath effects illustrated in Figure 2.9 are greater cause for concern in the navigation use-case.

we know about the surroundings and exploit prior knowledge on the consistency of the scene in which the robot finds itself.

2.5 Navigation with radar

In order to navigate through an environment with any sensor, observations between poses need to be associated with one another, as has been discussed in the vision and laser contexts in Section 2.3. The way in which this problem is subdivided and approached will differ from one method to the next, with varying degrees of directness, from methods with discrete steps to end-to-end systems that attempt to avoid this modularisation. This is often not a clean split between classical and machine learning approaches, and there has been much work done that seeks to leverage the strengths of both of these categories, some of which is contained in this thesis. In this section, applications of these techniques are presented to provide relevant background to radar's use in navigation, before discussing the role of radar in the ego-motion estimation context in greater detail.

2.5.1 Applications of radar in autonomous systems

Early uses of radar primarily consisted of target detection and tracking either in the air or at sea [48], where radar's ability to detect objects reliably over vast distances in all conditions has been unrivalled. Relying on radar for navigation was first adopted for the guidance of homing missiles, where both passive and more advanced active approaches were developed successfully. The target could be passively sought out and tracked via its own electromagnetic emissions, or actively honed in on by transmitting a radar signal either from the missile itself or from another base radar that could illuminate the target [49].

Systems that first began to use radar to navigate through an environment, rather than just to a designated target, looked at observing features on the ground from an aircraft, which could then be matched when observed from another location to determine relative motion [50], [51]. The work in [50] used an omnidirectional radar to measure echo pulses off the ground from an altitude of 600 m to 750 m to perform

localisation. This concept depended on there being detectable correlation between radar pulses that observed the same area, which when aligned would yield the relative motion. A short window of consecutive averaged echo pulses was compared with every other captured image in the flight sequence using the normalised inner product as a similarity metric. This technique proves successful in localising the aircraft and determining its ego-motion when previously gathered "memory" data is available. Early work in Synthetic Aperture Radar (SAR) also attempted a correlation approach using template matching [51]. Although it showed promise as a theoretical advance on simulated radar images, in practice they encountered difficulty when processing real data due to their proposed speckle model's inability to handle the complexities and variability of radar. Along a similar line, edge-matching [52], [53] has been successful when trying to correlate radar scans with a map or an overhead image — the latter task having seen new development that utilises a learnt approach to localise between radar scans and widely available satellite imagery that avoids first needing to build radar maps [54], [55].

Early ground-based systems navigate with a radar by detecting purpose-built radar beacons which are specifically designed to have a large RCS for reliable detection. These act as landmarks in the environment that can then be used to infer the relative pose of the radar between two observations [56]–[58]. After work on the SLAM problem had also seen success making use of radar as a sensor [58], there have been several developments that extend this work by making use of radar's multiple range returns, enabling operation without the need for specialised beacons in the environment [18], [59]–[63].

Estimating pose by matching points in 2D range scans is an effective navigation solution [64]. By registering the current observation against either a scan from a stored database (map) or a previous scan, it is possible to estimate the relative pose for an odometry task as well as to localise a robot in a previously mapped environment [65]. Inspired by some of the previously mentioned successes of the LiDAR-based solutions in Section 2.3.2, the ICP algorithm [39] has also been

used in the radar localisation task, although this method is poorly adapted to handle large pose offsets [2], [66].

Ego-motion can also be inferred from the discrepancies within maps that arise when robot motion has been sensed poorly, in a way that recovers the motion while also improving map quality [67]. Along a similar line of thought, the distortions in radar scans due to ego-motion can be used to solve for relative poses when travelling at high speeds, instead of treating these distortions as noise [68].

Methods that do not seek to extract points prior to alignment have turned to the Fourier-Mellin Transform (FMT) [69] to register radar scans as images and find the translation and rotation between them by making use of all the available information [70], [71], as had been done in the past with visual ego-motion estimation [72]. However, these methods can be prone to errors that arise from the complexities in radar data not being taken into account, as not all information contained in a scan is relevant when trying to align a static scene [3], [73].

Doppler radar systems, commonly used in automotive applications, provide measurements of the relative speed and direction of the surrounding landmarks that have been detected and processed. This information can also be used to infer ego-motion, which assumes that the largest subset of landmarks will correspond to the static scene and therefore encode the robot's ego-motion [74], [75]. Another interesting localisation approach uses Ground-penetrating Radar (GPR) to detect stable features below the driving surface. This approach achieves impressive cross-track accuracy below 5 cm at speeds of up to 60 mph, provided the trajectory on which localisation is required overlays the earlier mapping run very closely [76].

The recent increase in capability of modern radar devices has seen other novel applications of this sensor in the tasks of localisation [77], [78] place recognition [79], [80], perception [81], [82], semantic segmentation [83], occupancy grid mapping [84], and terrain classification [85], [86]. Although impressive progress has been made on many of these fundamental robotics problems using vision and LiDAR as discussed in Section 2.3, these capabilities are significantly impaired when environmental conditions become challenging. By directing our efforts towards solving some of

these tasks using only radar — a sensor that is reliable in many of the conditions that cause optical sensors to fail — the capabilities of mobile robots are increased beyond what was previously possible.

2.5.2 Radar odometry

Many of the contributions mentioned in the previous section serve to advance the theory and understanding of how we might use radar-only systems to navigate. Extending these developments to function at scale on a robot in the real world when faced with a variety of environments and practical concerns is a significant challenge. VO systems have been widely used in robotics tasks and successfully deployed in outdoor settings and diverse weather and lighting conditions [87], but are reliant on cameras which are prone to failure in some of these circumstances. LiDAR sensors have also seen extensive use in odometry systems evaluated on larger datasets [38], and have been coupled with vision systems to leverage the benefits they both have to offer [43], although are still prone to their shared weaknesses outlined in Section 2.2.3.

Radar Odometry (RO) has recently begun to see use in this domain that was previously dominated by vision and laser sensing modalities, and has also taken on a number of different forms. Cen and Newman [2], [88] developed a system that extracts landmarks from a pair of scans and solves for the rigid-body motion between them after a robust data association process. This had proved challenging in the past, particularly due either to poor landmark extraction, which failed to capture salient features in the scene without the assistance of radar beacons, or unreliable data association, when rotations or translations were too large for traditional point registration methods to handle [2].

Self-supervised learning has been shown to be useful in extracting robust keypoints from radar scans, which encode ego-motion information that can be inferred from their motion between frames [89]. Correlation-based methods that do not extract landmarks look to solve the odometry problem directly, and have made use of the FMT [70], [71], [90] as well as learnt methods, where the robot's

motion is exploited to mask out portions of the radar scan that would otherwise degrade pose estimation accuracy [73].

While some methods seek to approach the odometry task without appealing to machine learning techniques [2], [18], [90]–[92] and others make extensive use of them [73], [89], a recent hybrid approach has shown that there are advantages to employing deep learning methods when handling radar's complexity, which are complemented by the use of classical techniques for pose inference [16].

These radar ego-motion estimation methods can also be divided between point-based and dense approaches, although often the improved accuracy of dense methods is not as critical as the speed offered by point-based techniques. This is usually the case when online operation is required on a robot navigating through its environment, where a higher update rate allows for improved planning, safer obstacle avoidance, and has a lower computational cost.

Presently, the RO implementations that could be considered most similar to our baseline MRG RO system and the further contributions in this thesis would be the recent work on CFEAR by Adolfsson et al. [91] and HERO from Burnett et al. [16]. CFEAR presents a purely classical (or a learning-free) point-based RO design, and has recently achieved an impressive odometry performance of 1.76% in translational error on the Oxford Radar Robotcar Dataset [91], [93]. The hybrid approach of HERO that was mentioned earlier could be considered similar to some of the contributions in this thesis, in that it consists of a point-based method that uses deep-learning tools for a section of an RO pipeline. HERO achieves 1.98% translational error on the same metric detailed above [16], which improves on an earlier result by another learned point-based system [89]. Both of these advances have been developed since the majority of the work presented in this thesis, and no doubt further significant contributions to RO will continue to be made in the near future in this active area of robotics research.

2.6 Components of MRG Radar Odometry

Having covered the core contributions to field of RO that have explored a variety of approaches to the task of ego-motion estimation, we now turn our focus to a more detailed description of MRG RO. Specifically, we must understand the relevant components of Cen and Newman's work [2] in order to contextualise the contributions made in this thesis.

Before expanding on each subsystem within the MRG RO pipeline, it is useful to have a sense of the set of tasks assigned to each component and how these interface with each other. Figure 2.11 provides a high-level overview of the processes used to estimate the relative pose between two radar scans. To start, landmarks are extracted from each of the two scans that capture key features in the scene. These landmarks are then used to build point descriptors, which provide a way to identify a landmark based on the distances and bearings to the surrounding landmarks. The point descriptors are used to assign each landmark in one scan to a counterpart in the other scan to obtain an initial alignment, which serves to reduce the dimensionality by avoiding having to compare every point to every other point. Having built these initial correspondences, the compatibility score of each proposed unary match is then computed in a second data association step, from which the top 50 % are selected. Finally, this subset of landmark matches is used to compute the rigid-body motion of the robot using a least-squares method that estimates the translation and rotation between two point sets.

With these steps covered, we can now discuss these components in greater detail in the remaining subsections, starting with landmark extraction.

2.6.1 Landmark extraction

The task of landmark extraction from radar data can present many difficulties, as already outlined in Section 2.4.4. The use of radar beacons placed in the environment provides a way of simplifying this problem, but is not feasible when operating at scale [56], [57], [67]. This approach also limits radar's utility as it fails to make use of the rich information captured by radar [45], [60].

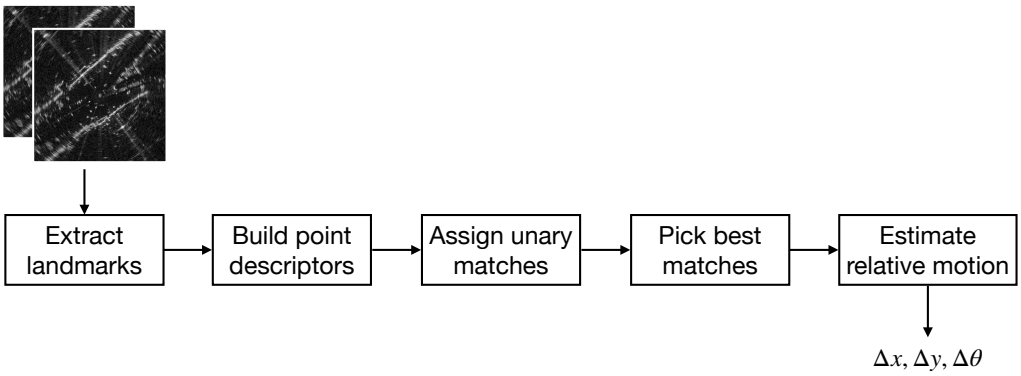


Figure 2.11: A high-level overview of our MRG RO pipeline that computes relative motion between two radar scans

To address this, landmark extraction can be performed with probabilistic methods to identify features naturally present in the scene. Adaptive detection techniques like those used in Constant False Alarm Rate (CFAR) algorithms rely on statistical principles associated with the sensor itself, and aim to detect landmarks with predefined probabilities of false alarm and missed detection [45], [94]–[96]. Although these techniques have been successful in mitigating some of radar's noise characteristics, the detections are still complex to interpret and still contain artefacts not useful for navigation. CFAR algorithms can also be difficult to tune for ego-motion estimation tasks, where the distribution of landmarks in the environment can change drastically from one section of a route to another.

Adolfsson et al. [91] seek to address CFAR's limitations with their CFEAR system mentioned earlier, devising a landmark extraction method that limits the number of detections per azimuth to the k-highest power returns. This comes at the cost of discarding true features that are less reflective, but whose omission is shown to have negligible impact on odometry performance. An extension of [91] attempts to improve CFAR with a Bounded False Alarm Rate (BFAR) method that applies an affine transformation to the CFAR noise floor estimate used in landmark detection, using a simple grid search to select parameters that minimise odometry error [97]. This sees a reduction in translational errors from 1.76 % to 1.55 %, improving on the accuracy achieved with CFEAR's sparser features [91].

Image feature detection techniques have also been applied to this landmark extraction problem [61], but tend to perform poorly on radar data compared with methods designed specifically for the radar domain [2], [98]. Feature extraction using machine learning approaches has also been applied to radar in numerous applications [99]–[102], with recent work in radar ego-motion estimation showing the success that can be achieved with self-supervised methods that optimise for odometry accuracy [89].

To extract landmarks from the surrounding scene one azimuth at a time, Cen and Newman [2] find peaks in the power-range spectrum that likely correspond to real objects. This is achieved by estimating the noise statistics of each signal, and then scaling the received power in each bin by the probability that it corresponds to a true landmark. Further work sought to reduce the number of tunable parameters that this initial method requires so as to maintain reliable performance in different domains [88]. However, this approach is still prone to detecting false landmarks, seeing up to twice as many detections arising than are necessary [3]. The first contribution in Chapter 3 addresses this issue and shows the clear decrease in the computational load when reducing these false positives.

One final detail worth noting is the need to account for inaccurate landmark positions that arise due to ego-motion distortion. During the radar's quarter-second rotation sweep duration, the robot starts and ends the sweep in different physical locations when on the move. As is commonplace in the literature, we collect azimuthal measurements taken at different time points and bundle them into scans with a single timestamp, when in fact the measurement stream is continuous, so instead of tracing a circle when in motion, it takes on a helical form. This effect can be accounted for by proportionately adjusting the position of observed landmarks using interpolated poses between two relative poses. A more thorough investigation into this distortion and the need to account for it has shown quantitatively how it affects ego-motion accuracy, particularly at higher speeds, as well as its significance over more minor effects like Doppler shift in radar measurements [47].

2.6.2 Point descriptors

With landmarks successfully extracted, the next task involves associating these points between the two scans. In order to make use of the spectral technique presented in [103] that is used in solving this task, the set of possible matches must first be limited to avoid having to compute the compatibility for every possible landmark association. A proposed *Radial Statistics Descriptor* [2] is used to describe landmarks by their spatial relationships to the surrounding points. This feature descriptor encodes the range and bearing distribution of neighbouring points, a technique which is robust under large translations and rotations between observations, provided there is sufficient overlap.

Independent work by Burnett et al. [47] verified the benefit of using these specialised descriptors as compared with ORB features, which are more commonly used in vision [104]. When attempting to localise against a scan stored in a map but seen from a different viewpoint, the difference in RCS was significant enough to cause inconsistencies in the detected ORB features [47].

2.6.3 Unary match assignment

Having generated point descriptors for each landmark, the next step involves matching landmarks corresponding to the first set of descriptors with their counterparts in the second. A greedy algorithm runs through each possible descriptor pairing, assigning landmarks from the first set with their nearest neighbour in feature space in the second set using the Euclidean distance between feature vectors. Once a landmark has been paired, it is removed from the set of candidates left to consider, ensuring that there will be only one correspondence per landmark.

This is by no means an optimal solution, as it still involves comparing every landmark with every possible remaining partner, and neglects instances where multiple landmarks sharing the same correspondence may have been more suitable. However, it serves as a sufficient pre-processing step to make the size of the correspondence problem in the next step more manageable.

An intuitive way to limit the number of comparisons here is to introduce a motion prior, and only consider points that are within a fixed distance threshold of each other as Hong et al. describe [18]. This is in addition to their pairwise consistency constraint that rejects matches where the magnitude of the difference in the distance between two points in one scan is greater than some fixed threshold [18].

2.6.4 Selecting correspondences

Once the set of unary matches has been constructed, the next step involves identifying which of these associations are reliable for use in ego-motion estimation. As alluded to in Section 2.6.2, a correspondence technique based on the compatibility of each match is used to select the final set of associated landmarks [103]. Each correspondence is compared with every other possible correspondence, so that for n unary matches there are $n \times n$ comparisons. This forms an $n \times n$ compatibility matrix M, where each element contains the computed score associated with the link between those two correspondences, based on their similarity in the distance to another point. This score $s_{(i,j),(i',j')}$ is computed as:

$$s_{(i,j),(i',j')} = \frac{1}{1 + |d_{i,j} - d_{i',j'}|}$$

where $d_{i,j}$ is the distance between the points (i, j) in the first point set, and $d_{i',j'}$ is the distance between points (i', j') in the second set. Stated plainly, when a correspondence between two points in the first set is similar to a correspondence between two points in the second set, the distances between them should also be similar. This will drive the score to be near 1, while incorrect links between matches where there is a significant discrepancy will not share this geometric similarity, and so will push the score nearer to 0.

After the compatibility matrix has been populated with these scores, the principal eigenvector of this matrix is obtained in order to determine the most suitable matches, as outlined in earlier work by Shapiro and Brady [105] and the contribution of Scott and Longuet-Higgins [106]. Specifically, as per Leordeanu and Hebert's

method [103], we seek to find the set of correspondences belonging to the cluster C that maximises the cluster score S_C:

$$S_C = \sum_{a,b \in C} M(a,b)$$

where $a = (i, i')$ and $b = (j, j')$ are two correspondences between the two point sets, and recalling that M is described earlier as the compatibility or affinity matrix containing the scores computed as $s_{(i,j),(i',j')}$. A cluster can be represented by a binary vector x, with elements set to either 0 or 1 at $x(a)$ depending on whether a is contained in the cluster. The score for a particular cluster can then be written as:

$$S_C = \sum_{a,b \in C} M(a,b) = x^T M x$$

Now, the optimal solution x^* that represents the cluster with the maximum score is found by solving:

$$x^* = \arg\max_{x \in \{0,1\}} (x^T M x)$$

It is beneficial to relax these discrete values in order to improve computational efficiency, allowing the elements in x to take on real values between 0 and 1. These values can now be interpreted as an indication of the association of each element to the cluster, rather than as a simple indicator of their membership status. Finally, in this form, finding the x^* that maximises the cluster score is equivalent to finding the principal eigenvector of M, as per the Rayleigh-Ritz ratio [107]. And with this principal eigenvector in hand, its elements can be interpreted as a relative confidence value in a particular association [103]. This concept of match confidence is leveraged in our RO introspection work ahead in Chapter 4 that investigates failure modes of the correspondence algorithm, such as excessive symmetry or a lack of stable features in the scene [4].

The last step in the correspondence selection process is to add the matches corresponding to the highest principal eigenvector elements first, working through the overall set. This process terminates when the top 50 % of matches have been added, or the remaining elements are below a minimum confidence value.

2.6.5 Computing rigid-body motion

Now that the final associations between the two point sets are computed, the last stage in the process of ego-motion estimation requires finding the underlying relative pose that caused their displacement. Given two sets of associated points, the procedure developed by Challis is capable of computing the relative translation and rotation that will align them [108]. This method requires at least two correspondences to constrain the problem and find a solution for the robot's pose in $\Delta x, \Delta y, \Delta\theta$. The rigid-body transformation parameters map points in one reference frame to another, specifically as:

$$p'_i = Rp_i + v$$

where the position of a point p_i is moved to p'_i through a rotation matrix R and translation vector v, which we seek to recover. This derivation is presented in detail in [108], where ultimately the SVD is used to find R between the re-centred point sets using the decomposition of cross-dispersion matrix C, before computing v as:

$$v = p'_i - Rp_i$$

Two interesting additions that are not implemented in our original MRG RO system but provide useful extensions to this SVD method are noted here for the sake of completeness. Firstly, a weighting factor for each match can be set in the rigid-body method as mentioned by Challis [108], which reduces the influence of poorer matches over their more stable counterparts. If desired, the eigenvector confidence score could be included in this step as a real value, instead of being limited in its use as a binary mechanism to filter inliers from outliers. Secondly, a form of uncertainty can be computed by iteratively running the SVD method on a small subset of matches and calculating the standard deviation of the resulting pose estimates.

2.7 Identifying areas in need of improvement

This section briefly motivates the reasoning behind using this particular RO implementation, and explicitly identifies areas in the original system that present

opportunities for innovation. These topics are picked up in Chapters 3 to 5, where the core contributions made in this thesis are presented.

2.7.1 Using MRG RO as a baseline

As already stated, the contributions contained in this thesis make use of the MRG RO system first proposed by Cen and Newman [2] that has been covered in detail in Section 2.6. This point-based odometry pipeline provides a way to solve for the relative motion between two radar scans despite all their complexities, in varied environments, and at scale. As shown in Figure 2.11, this system is made up of separable components that each have a specific role to play, and if required, can be specialised or replaced entirely by other subsystems to improve performance in a specific area. In addition, our baseline system requires no training data prior to deployment, and runs without the need for a GPU. This data-requirement attribute was particularly relevant for the first and second contributions in this thesis, which pre-dated the *Oxford Radar Robotcar Dataset* [93], as prior to this dataset's release, the quantity of publicly available radar data with a reliable ground truth odometry signal was insufficient to train a network.

The MRG RO pipeline offers a suitable starting point for the contributions in Chapters 3 to 5, and provides a firm foundation on which to build. Specifically, it is reliable despite radar's unique artefacts, does not require scans to be well-aligned prior to estimation, is relatively fast compared with dense methods, estimates with centimetre accuracy compared to a ground truth signal, and does not depend on any additional sensors. Although this system showed ego-motion estimation performance on par with VO, there are still several shortcomings that were unaddressed in the initial work.

2.7.2 Building on the MRG RO foundation

This subsection provides an overview of key areas in need of improvement in the MRG RO system, and specifies our contributions in these areas.

Filtering out radar complexities for faster processing speeds

As stated by Cen in [109], the processing speed scales quasi-quadratically with the number of landmarks extracted from a scene, which can impact odometry update rates. When landmarks were in excess of around 800, estimation time was near to the sensor's update period of 250 ms, as shown in [3]. This makes real-time operation difficult to ensure, which impedes the practical utility of the system without some form of landmark limiting, as proposed in [88]. The landmarks that are detected by the extraction algorithm contain many false positives as mentioned in Section 2.6.1, which directly impact the processing time required by the other downstream subsystems. The contribution in Chapter 3 discusses this disadvantage and proposes a filtering mechanism to identify and remove spurious radar returns.

Addressing scan matching odometry failures through introspection

Another consideration not addressed in the MRG RO system that would benefit performance involves developing a way to assess the quality of proposed solutions introspectively. Without this, the MRG RO pipeline produces ego-motion estimates that can fail critically when there are highly symmetric features or occlusions in the scene, and has no way in which to flag this vulnerability. Chapter 4 picks up on this need, and describes the development of the missing component which classifies poor performance prior to mitigating it.

Leveraging a motion model to refine estimation accuracy

As a radar scan-matching system, MRG RO is in fact not specialised for use in odometry. This is beneficial in part, as it allows for the flexibility to perform both odometry and localisation using the same core pipeline with no specific motion model constraints, with the only major requirement being that there must be sufficient scene overlap. However, this generality comes at a cost which becomes apparent when considering MRG RO's ego-motion estimates closely, and with knowledge of the motion constraints of the platform on which the radar is mounted. Specifically in our case, infeasible motion is observable when assessing estimates of lateral

translation, where centimetre-scale errors that corrupt the odometry solutions are present. In response to this, the contribution detailed in Chapter 5 shows that by acknowledging a vehicle's non-holonomic constraints, the associations that are the sources of these errors are easier to identify and discard.

With an understanding of the motivation behind using radar and the capabilities and limitations of established odometry methods, an appreciation for the details of our MRG RO system, and insight into areas that can be built upon, we can continue to our first contribution in Chapter 3.

3

Pre-filtering radar scans for faster ego-motion estimation

Contents

3.1 Overview of contribution 44
3.2 Integrated manuscript 45
3.3 Further insights 54
3.3.1 Choosing what to label 54
3.3.2 Odometry source for automatic annotation 55
3.3.3 Further discussion on original results 56
3.3.4 Alternative baselines alongside the learnt approach . . . 57
3.3.5 Updated related work 58
3.4 Contribution review 59

3.1 Overview of contribution

Extracting landmarks from radar scans is challenging, particularly when faced with radar-specific complexities that are not identified as spurious returns. As mentioned in Section 2.7.2, processing times in this point-based system are sensitive to excessive landmark detection, causing a bottleneck that prevents reliable real-time operation. The data association methods implemented in our MRG RO pipeline are robust to most of the errors in landmark extraction, but are still burdened by needing to

process those instances that are not part of the static scene and, ultimately, not useful for navigation. This creates an additional load on the system, which has no way to distinguish between parts of the scene which contain landmarks that correspond to real objects, and those that are a result of the difficulties listed in Section 2.4.4.

In this publication, a weakly-supervised approach to pre-filtering radar scans allows us to learn which returns in a scan are likely to represent static landmarks in the environment, which are useful in the navigation task. By masking landmarks that are not relevant for ego-motion estimation, this work reduced the processing time without compromising on ego-motion estimation accuracy. Following the completion of this work, this approach of using the robot's motion through the scene to differentiate between desirable and undesirable radar returns has been leveraged as part of a self-supervised correlation-based RO implementation [73]. Many of the difficulties that arise due to the complexity present in radar data can be mitigated by using this masking approach and incorporating the strong prior that radar returns corresponding to static objects are consistent between frames.

3.2 Integrated manuscript

The primary numerical result of this contribution was to **increase processing speed by a factor of 2.36** as compared with the baseline MRG RO implementation. This manuscript was presented at the 2019 IEEE International Conference on Robotics and Automation (ICRA) in Montréal, Canada.

Fast Radar Motion Estimation with a Learnt Focus of Attention using Weak Supervision

Roberto Aldera, Daniele De Martini*, Matthew Gadd*, Paul Newman
Oxford Robotics Institute, Dept. Engineering Science, University of Oxford, UK.
{roberto,daniele,mattgadd,pnewman}@robots.ox.ac.uk

***Abstract*— This paper is about fast motion estimation with scanning radar. We use weak supervision to train a focus of attention policy which actively down-samples the measurement stream before data association steps are undertaken. At training, we avoid laborious manual labelling by exploiting short-term sensor coherence from multiple poses in the presence of an external ego-motion estimator (for example, wheel odometry). In this way, we generate copious annotated measurements which can be used for training a learning algorithm in a weakly-supervised fashion. We demonstrate the validity of the approach in the context of a Radar Odometry (RO) task, pre-filtering raw data with a popular image segmentation network trained as presented. We evaluate our system against 26 km of data collected in Central Oxford and show consistent motion estimation with greatly reduced radar processing times (by a factor of 2.36).**

***Index Terms*— radar, sensing, ego-motion estimation, field robotics**

I. Introduction

Radar is a sensor that continues to receive relatively little attention as a source of pose information for field robots in unstructured environments – typically being relegated to collision avoidance systems in domains such as driving assist systems. Frequency-Modulated Continuous-Wave (FMCW) scanning radars, in contrast to more widely used cameras and lidar sensors, operate well under variable weather and lighting conditions. Due to the long wavelength of radio waves and processing techniques, radar sensors receive multiple returns from a single azimuthal transmission and can operate out to many hundreds of metres. In many ways this sounds like the ideal sensor. However, radar measurements are complex. The beam is not narrow and tightly focused, returns are affected by various noise sources, and the interaction of the electromagnetic wave in the environment is far more complex than that of time-of-flight (TOF) lasers [1]. Consequently, comprehension of radar scans for precise ego-motion estimation requires dealing with complex measurement patterns which are not intuitive. In the end, as it often does, this boils down to a data association problem: "what detail in the last frame is relevant to the detail seen in this frame?" This paper is about making that task simpler.

Figure 1 shows an example of the Cartesian projection of a complete raw scan as collected by a FMCW scanning radar, where the peaks of the indicated reflections cannot be completely resolved into objects at those locations – even by a human expert. Indeed, the energy reflected back to the source sensor is a complex function of both geometry and material properties. A small nail in a wall at 400 m range is an ideal natural feature, but may appear as "bright" and as large as a car or be considered a nuisance glint – only briefly visible.

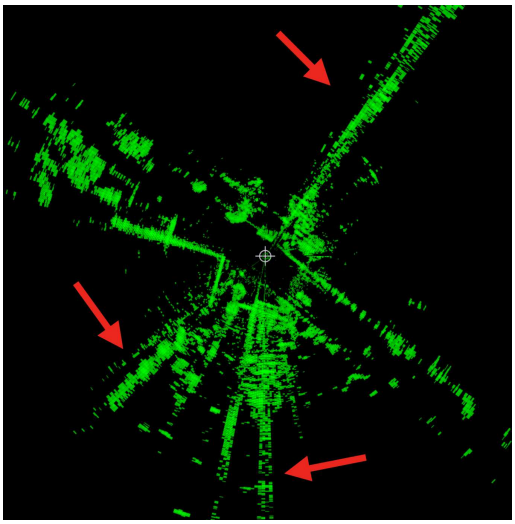


Fig. 1: An unfiltered radar scan: returns which are not spatially coherent from multiple viewpoints have been highlighted by red arrows.

This paper presents in Section IV a simple but effective method for down-sampling the measurements in radar scans to ease the computational complexity of data association between two scans. As our goal is ego-motion estimation or localisation, we leverage the straightforward point that ideally we would only operate with returns from artefacts in the scene that are visible from multiple views. Accordingly, we set ourselves the task of building a filter that only passes through measurements which, in the context of the entire scene, appear to be visible across wide baselines. We approach this task by building a classifier which marks individual reflections as either suitable for the data association stage or not.

The training of such a classifier may be challenging. Certainly manual labelling is arduous, but even if that were ignored, it is not a trivial task to decide what detail in one scan is visible in another given the remarkable way radar data is liable to change under small pose increments (totally unlike first return lidars or cameras). We create annotations

* The authors contributed equally to this work.

in a weakly-supervised fashion – simply accumulating radar returns in consecutive scans and looking for spatial coincidence. To do this, of course, one needs an external ego-motion source at training time, for which we use Visual Odometry (VO) [2].

Section V demonstrates the feasibility of the procedure in training a segmentation network based on U-Net [3]. We present and analyse in Section VI the impact of this approach upon the performance of a state-of-the-art Radar Odometry (RO) algorithm [4], where we achieve consistent motion estimation with radar processing times reduced from $168.12\,\mathrm{ms}$ to $71.10\,\mathrm{ms}$ as averaged over three representative trials in an urban environment.

II. Related works

Many techniques have been developed to extract landmarks from radar data, mostly based on probabilistic distributions and integrated directly into landmark extraction algorithms, e.g. in Constant False Alarm Rate (CFAR) processor and derivates [5, 6, 7] or in [4]. Although Deep Learning (DL) techniques have been proven to be effective in many applications for noise filtering [8, 9, 10], there is little or no work that exploits them in directly filtering range measurements [11], although many can be found on end-to-end classification, mostly concerning Automatic Target Recognition (ATR) problems [12, 13, 14].

Indeed, the drawback of such techniques is the need for an exhaustive set of annotated training examples in learning how to perform a specific task. The data annotation procedure can be a costly process in terms of both time and expertise since, in many situations, ground truth annotation can not only take much time, but it may require a specific knowledge of the system that only an expert can provide. Thus, it is desirable to build Machine Learning (ML) frameworks that can work with weak supervision. In [15] three categories of Weakly-Supervised Learning (WSL) are described – the proposed method falling into the *inaccurate supervision*, i.e. where information in the ground truth data presented to the model may suffer from errors.

Similar approaches can be found in other domains, where accurate human annotation has been replaced by less accurate, but automatic, labelling procedures. Most closely related to our method is the work of Barnes et al. [16] which uses a learned model to mask sensor observations for the purposes of improving a vision-based odometry system. As in this work, the authors use automatic annotation from a localisation system to generate ground truth pixel-wise mask labels for training a CNN by analysing geometric consistency. Additionally, the authors look at geometric consistency over multiple traversals of the same route, whereas we look only at consistency over a single traversal. However, we are the first to apply this method to radar domain data.

In a similar fashion [17] use lidar records for training a DL model to perform a road marking segmentation on images. They exploit material reflectance to detect road markings with a push-broom laser scanner and build the ground truth images by projecting the results onto the video stream from the camera as the vehicle moves using odometry information.

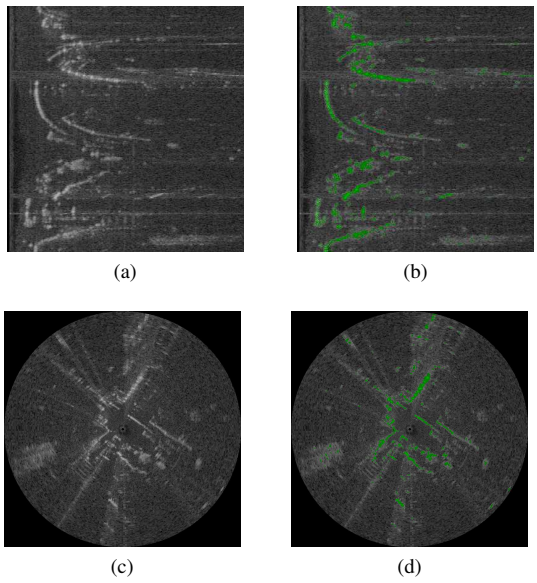


Fig. 2: Examples of scans assembled from successive azimuth readings from a FMCW scanning radar before and after the proposed annotation procedure. (a) and (c) show the raw data in polar (range vs azimuth) and Cartesian space respectively, whereas the corresponding annotated scans (green labels indicative of class **KEEP**) are shown in (b) and (d) respectively. For clarity, these examples have been cropped to display 384 of the full 2000 bins available that are used to train the network discussed in Section V.

Finally, in [18], in the same spirit as us, the authors apply a WSL approach to train a network for object detection on dynamic grid maps. They exploit temporal and spatial relationships to extract moving objects and their shapes using a lidar sensor. Indeed, once an object is observed, it continues being observed until it exits the field of view, updating shape and trajectory information. Then, this information is propagated backwards in time to refine the annotation for more consistency in the labels.

III. Preliminaries

The sensor we employ is a FMCW scanning radar which rotates about its vertical axis while continuously sensing the environment through the transmission and reception of frequency-modulated radio waves. While rotating, the sensor inspects one angular portion (*azimuth*, α) of space at a time and receives a power signal that is a function of reflectivity, size, and orientation of objects at that specific azimuth and at a particular distance, ρ. The radar takes measurements along an azimuth at one of M discrete intervals and returns N power readings which we refer to as *bins*. One full rotation across all M azimuths is called a *scan*, $\mathcal{S}$. Furthermore, let $\mathbf{s}(k) \in \mathbb{R}^{N\times 1}$ be the power-range readings at time step k, where $t(k) = t_k$ is the time value at k and $\alpha(k) \in A$ is

the azimuth associated with the measurement. The element $s_i(k) \in \mathbf{s}(k)$ is the power return at the i-th range bin, with $i \in \{1, \ldots, N\}$. The measurement range is given by $\rho_i(k) = \beta(i - 0.5)$, where β is the range resolution of the radar.

IV. Methodology

In RO tasks such as [4], data association involves correlating the distance between pairs of landmarks identified in subsequent frames. The aim of this work is to develop a front end that simplifies this data association burden by focusing its attention on a stream of raw azimuth messages that are a filtered subset of reflections likely to be visible across wide baselines.

To this end, we automatically annotate each portion of the sensor's polar space representation as belonging to the class **KEEP** or not. Here **KEEP** is defined as a power bin that contains an object which has been observed to be wide baseline visible. This means that noise effects, dynamic objects, and reflections from objects that are highly specular or not visible from multiple views constitute the **REJECT** class.

The details of the proposed pipeline will be described in more detail from Section IV-A to Section IV-C, and are summarised here as:

(A) collection of a whole-scan measurement and conversion into Cartesian space;
(B) comparison of the spatially accumulated measurements and annotation of the power bins by thresholding;
(C) updating the internal representation of the environment and conversion of the current scan back to polar space.

A. Data Accumulation in Cartesian Space

The labelling process is carried out upon scans $\mathcal{S}(\bar{k}, \bar{k} + M)$. To this end, M azimuth measurements $\mathbf{s}(k), k \in (\bar{k}, \bar{k} + M]$ are collected and processed to represent the Cartesian environment prior to annotation. From now on, we will abbreviate $\mathcal{S}(\bar{k}, \bar{k} + M)$ as $\mathcal{S}_{\bar{k}}$, where $\bar{k}$ represents the time step at the last azimuth reading of the previously assembled scan, $\mathcal{S}_{\bar{k}-M}$.

Each measurement bin is processed separately to project it into Cartesian space. Assuming that the vehicle is moving on a flat ground plane, the Cartesian representation of the i-th bin within an azimuth reading is given by

$$\mathbf{x}_i(k) = \left[\rho_i(k)\cos\alpha(k), \rho_i(k)\sin\alpha(k), 0, 1\right]^T \quad (1)$$

specified using a projective space, $\mathbb{P}^3$, and the associated homogeneous co-ordinate system.

The rigid-body transformation $T(\bar{k}, k) \in \mathbb{SE}(3)$ between the pose of the robot at time steps coinciding with the start of the window and the current azimuth reading[1] ($\bar{k}$ and k respectively) can be used to project this sensor-centric Cartesian point $\mathbf{x}_i(k)$ into a common frame of reference as $\bar{\mathbf{x}}_i(k) = T(\bar{k}, k) \cdot \mathbf{x}_i(k)$ where the transformation

$$T(\bar{k}, k) = \begin{bmatrix} \mathbf{R}(\bar{k}, k) & \mathbf{t}(\bar{k}, k) \\ \mathbf{0} & 1 \end{bmatrix}$$

is obtained by exploiting any reliable external odometry measurement, in our case Visual Odometry (VO) [2]. In general, however, let $\bar{h}$ and h be the closest time steps of the external odometry source with respect to the radar measurement samples such that $t_{\bar{h}} \leq t_{\bar{k}}$ and $t_h \geq t_k$. Then, the transformation $T(\bar{k}, k)$ can be calculated as

$$T(\bar{k}, k) = T(\bar{k}, \bar{h}) \cdot T(\bar{h}, h) \cdot T(h, k) \quad (2)$$

and

$$T(\bar{h}, h) = \prod_{l=\bar{h}}^{h-1} T(l, l+1) \quad (3)$$

The terms $T(\bar{k}, \bar{h})$ and $T(h, k)$ are obtained by interpolation in $[\bar{h}, \bar{h} + 1]$ and $[h - 1, h]$, respectively: the rotational component is obtained by Spherical Linear Interpolation (SLERP) on a spherical surface traced by a unit quaternion, as described in more detail in [19], and the translational component is obtained by a constant velocity interpolation. Since the time step $\bar{k}$ defines the last time step of the previous scan, the term $T(\bar{h}, \bar{k})$ is available from previous computations.

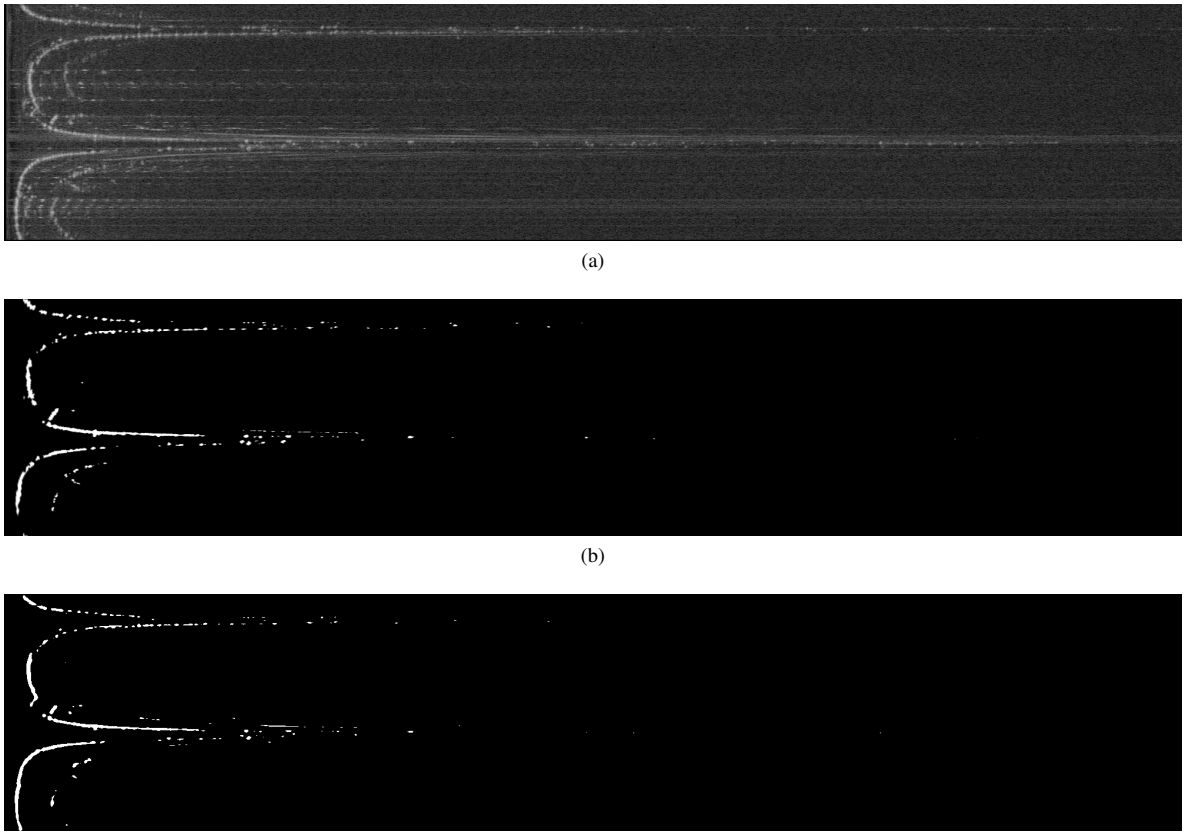
Once the position of each bin is projected into Cartesian space with a common reference frame, it is possible to build the Cartesian representation of the current scan, $\mathcal{S}_{\bar{k}}$. The chosen representation for the Cartesian space is a fixed size 2-dimensional grid, $\mathcal{G}_{\bar{k}}$, of size $2N \times 2N$, the values of which are derived from the power readings from incoming and projected bins. The two representations, Cartesian and polar, do not match perfectly, since the sparsity of the information in polar space depends on the distance from the radar, requiring the interpolation of each cell power value $\mathcal{G}_{\bar{k}}(u, v)$, with $u, v \in [-N, N]$, among the four closest bin measurements. For this task we employed a weighted polar to Cartesian data conversion as described in [1]. Figures 2a and 2c show the polar and Cartesian representations of a whole scan respectively.

B. Labelling Bins in Cartesian Space

Let $\mathcal{W}(k) = \{\mathcal{G}_{k-l*M} : l \in \{1, \ldots, w\}\}$ be a vector of Cartesian grids each of size $2N \times 2N$. $\mathcal{W}$ contains, at the time step k, the previous w scans represented as described above. Moreover, the relative transformations between them are recorded in $\mathcal{T}(k)$ by chaining transformations computed by eq. (2).

At this point, we can construct a histogram $\mathcal{H}_k$ of size $2N \times 2N$ by projecting each grid in $\mathcal{W}(k)$ into a common reference attached to the current grid, $\mathcal{G}_k$, and by counting the number of overlapping cells which have a power value greater than some label threshold, τ. This count is assigned to $\mathcal{H}_k(u, v)$, with $u, v \in [-N, N]$.

[1] In practice, we find a single pose for an entire scan (and not for every azimuth) to yield comparable performance in our weakly-supervised approach at a much improved processing speed for label generation.



(a)

(b)

(c)

Fig. 3: Visual comparison of a complete radar scan in polar coordinates (range vs azimuth): (a) shows the original data, while (b) and (c) show classification masks from the labelling procedure and the U-Net network respectively.

If the number of occurrences in a single histogram cell exceeds a given threshold, σ, the corresponding cell in $\mathcal{G}_k$ is annotated as **KEEP**, otherwise it is assigned to the **REJECT** class. Figure 2d shows the results of the labelling task performed on the measurements. In addition to reflections and noise, dynamic objects have not been labelled as **KEEP**s.

C. Updating the Environment's Polar Space Representation

Once the histogram calculation is completed, it is possible to update the internal representation of the environment; discarding the oldest scan grid from $\mathcal{W}$ and transformation from $\mathcal{T}$ and replacing them as more azimuth readings become available. Since annotated bins in polar space are required, the last step is to transform the annotated grid, $\mathcal{G}_k$, back into polar space. To do so, we exploit the weights derived by the projection into Cartesian space and annotate as **KEEP** each bin that most contributed to the power value in each corresponding **KEEP** grid cell. Figure 2b shows the results of the classification task on the polar measurements.

V. Experimental Setup

We discuss in this section an integration of our pre-filtering technique within a RO system as described in [4], leveraging U-Net [3]. The raw scans (indicated in the results as **ro**), scans with generated masks applied directly (labelled as **gt**), scans filtered by the network output (**unet**), and outputs from VO (**vo**) [2] are all compared with each other.

A. Hardware

The tests are performed using data collected from the *Oxford RobotCar* platform [20]. We employ a CTS350-X Navtech FMCW scanning radar without Doppler information, mounted on top of the platform with an axis of rotation perpendicular to the driving surface. It is characterised by an operating frequency of 76 GHz to 77 GHz, yielding up to 2000 range readings with a resolution of $0.25\,\mathrm{m}$, each constituting one of the 400 azimuth readings with a resolution of 2° and a scan rotation rate of $4\,\mathrm{Hz}$.

Images (for VO) were gathered by a Point Grey Bumblebee XB3 camera, mounted on the front of the platform facing towards the direction of motion. The camera is characterised by $1280\times960\times3$ resolution, $16\,\mathrm{Hz}$ FPS, $1/3''$ Sony ICX445 CCD, global shutter, $3.8\,\mathrm{mm}$ lens, 66° HFoV, $12/24\,\mathrm{cm}$ baseline.

The training and the evaluation have been run on a Dell PowerEdge machine with 3.33 GHz Intel Xeon processors, 192 GB RAM, 2666 MT/s DDR4.

B. Odometry

Our implementation of VO uses FAST [21] corners combined with BRIEF [22] descriptors, RANSAC [23] for outlier

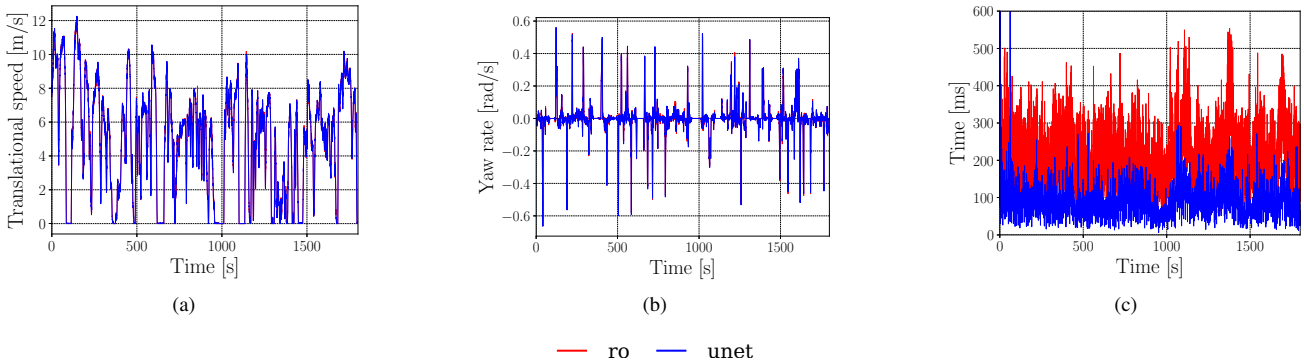


Fig. 4: Profiles of translational (a) and rotational (b) velocities with computational time for odometry estimation (c), for a $9\,\mathrm{km}$ portion of a $10\,\mathrm{km}$ loop, corresponding to the Cloudy condition summarised in Table I.

rejection, and nonlinear least-squares refinement. In order to verify the compatibility of the labelling method with existing landmark extraction algorithms, we exploit the RO method as described in [4], without any changes to the algorithm.

C. U-Net Training

The network is trained by providing it with whole-scan images of raw data in polar space and the corresponding annotated image obtained from the proposed method, where each pixel row corresponds to a single azimuth measurement and each pixel column corresponds to a bin, the value of which is set to `255` if the bin belongs to the **KEEP** class or `0` otherwise. As shown in Figure 3, each single-channel grayscale image is of resolution $M \times N$, where $M = 400$ and $N = 2000$ are the cardinalities of the azimuth set in a single scan and of the bin set in a single azimuth reading respectively, as described in Section IV.

The loss function we exploited is a sum of Dice coefficient and cross-entropy loss and we apply early stopping at 5 epochs, obtaining a validation Dice coefficient of $64.32\,\%$ and a validation loss of 0.37.

The dataset employed consists of three trials, collected in different weather conditions, one of which is used for training the network. This trial consists of a single loop under sunny weather conditions (favourable to VO). It has been divided by a proportion of $90\,\%$-$10\,\%$ for training and validation sets respectively, shuffled in order to avoid learning environment-specific geometric features (such as street or building layout) present within training data and absent from validation, or vice versa. The remaining trials (Cloudy, Rainy, and Night) are used for the evaluation of the proposed method, as described below.

Augmentation on the dataset has been performed as random noise on the input image to mimic speckle noise in radar data and random flips on the horizontal axes. While horizontal flips can be interpreted as a reverse in perspective, vertical flips are not physically meaningful due to the polar nature of the measurements.

	Translational speed $[\mathrm{mm\,s^{-1}}]$		
	ro	gt	unet
Cloudy	103.86	110.75	110.87
Rainy	97.06	100.67	99.89
Night	198.23	208.51	206.79
	Rotational speed $[\mathrm{mrad\,s^{-1}}]$		
	ro	gt	unet
Cloudy	5.54	5.64	5.72
Rainy	4.94	5.25	5.36
Night	6.88	7.26	7.14
	Estimation time [ms]		
	ro	gt	unet
Cloudy	232.17	107.92	99.05
Rainy	174.11	84.17	74.97
Night	98.08	36.02	39.27

TABLE I: Summary statistics for several sorties exhibiting distinct environmental conditions. The vehicle drove $8834.44\,\mathrm{m}$, $8379.14\,\mathrm{m}$, and $8846.90\,\mathrm{m}$ during the Cloudy, Rainy, and Night outings, respectively. The median error quantities are presented in millimetres and milliradians per second to expose the significant digits. Estimation time is summarised by a RMS value.

D. U-Net Inference

At run-time we produce filtering masks at an average of $13.52\,\mathrm{Hz}$ using a single GTX 1080 Ti GPU, which is more rapid than the baseline RO performance ($5.95\,\mathrm{Hz}$) and our implementation which produces ground truth labels ($3.78\,\mathrm{Hz}$) and as such it does not need to be considered when discussing timing results. The masks generated by **unet** and **gt** are applied directly onto the landmark sets to **KEEP** bins that are relevant between frames and **REJECT** those that are not.

VI. Results

This section presents metrics for the performance of our approach when deployed as described above, summarised in Table I.

Figure 4 shows a comparison of odometry estimation as compared to the baseline **ro** discussed above. Here, the vehicle has traversed approximately $9\,\mathrm{km}$ of a $10\,\mathrm{km}$ urban

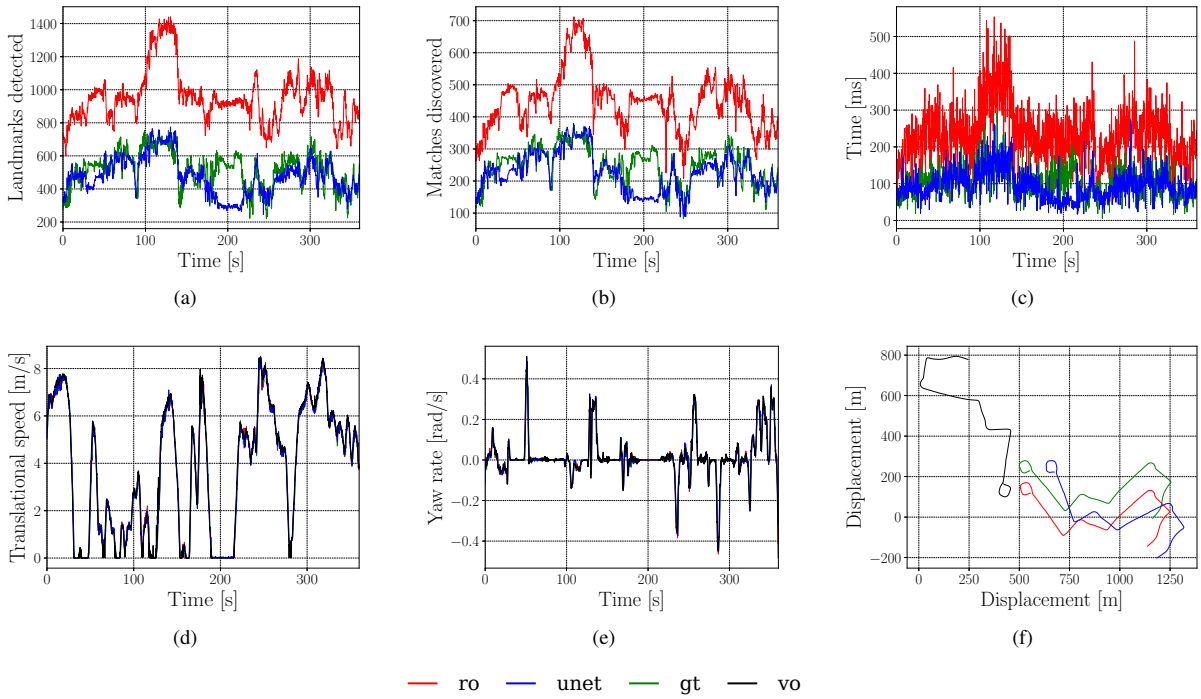


Fig. 5: Number of landmarks detected (a), number of matches between landmarks discovered (b), computational time for odometry estimation (c), translational velocity (d), rotational velocity (e), and integrated pose (f) on a short portion of a 10 km loop, corresponding to the Cloudy condition summarised in Table I. In (f) the absolute pose with respect to the start of the foray is shown, where it is clear that the drift experienced by the camera system is distinct from any of the radar systems. Nevertheless, the shape of the recovered poses, and thus the inherent egomotion, is consistent. In (a) and (b) it should be noted that by default the motion estimation algorithm attempts to match only half of the detected landmarks between frames.

route in mixed traffic and consequently exhibits a range of movement types (including regular stopping and starting at traffic control points as well as accelerating and decelerating along straight sections of road and around corners). As can be seen in Figure 4, the translational and rotational velocities recovered by using scans pre-filtered by our approach, **unet**, is consistent with the baseline **ro**. The median estimation time of 99.05 ms for this trial, as read from Table I and shown for the full trial in Figure 4, is a marked improvement over the baseline of 232.17 ms and approximates that of the ground truth annotation (107.92 ms).

In Figure 5 we consider a briefer portion of this same trial in more detail. Figure 5a includes for comparison the landmarks detected by applying the ground truth mask as discussed in Section V above (**gt**). Additionally, Figure 5d and Figure 5e include for comparison the translational and angular velocities recovered by VO.

Table I presents some summary statistics over three forays with distinct atmospheric conditions that do not correspond to the (Sunny) condition of the training data. The performance achieved by the baseline **ro**, **unet**, and **gt** is presented in terms of the median translational and angular velocity errors using **vo** as the benchmark signal. We observe in Figure 5f the integrated pose of **unet** – our fully integrated system – tracks visually that of the baseline **ro** faithfully.

VII. Conclusions

This paper demonstrates the feasibility of using a weakly-supervised learning framework to pre-filter radar measurements for use in radar-only navigation. By retaining only the information that is wide-baseline visible, our approach is able to reduce processing times (by a factor of 2.36) while maintaining the accuracy of the recovered motion. We evaluate our system over 26 km of urban driving and show that it rivals baseline RO despite having access to significantly fewer features. This contribution is more than a neat efficiency improvement – by focusing attention on relevant landmarks, the task of motion estimation can now be run in real-time on a robot using radar alone.

Acknowledgements

Roberto Aldera and Daniele De Martini are supported by the UK Engineering and Physical Sciences Research Council (EPSRC) programme grant EP/M019918/1. Matthew Gadd is supported by Innovate UK under CAV2 – Stream 1 CRD (DRIVEN). Paul Newman is supported by EPSRC Leadership Fellowship Grant EP/J012017/1. The authors would also like to thank Paul Murcutt for his software engineering support.

REFERENCES

[1] E. J. M. Adams, J. Mullane and B. Vo, *Robot Navigation and Mapping with Radar*. Artech House, 2012.

[2] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. Ieee, 2004, pp. I–I.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[4] S. Cen and P. Newman, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[5] M. A. Khalighi and M. H. Bastani, "Adaptive CFAR processor for nonhomogeneous environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 3, pp. 889–897, July 2000.

[6] P. P. Gandhi and S. A. Kassam, "Optimality of the cell averaging CFAR detector," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1226–1228, July 1994.

[7] B. Magaz, A. Belouchrani *et al.*, "Automatic threshold selection in OS-CFAR radar detection using information theoretic criteria," *Progress In Electromagnetics Research*, vol. 30, pp. 157–175, 2011.

[8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[9] J.-M. Valin, "A hybrid dsp/deep learning approach to real-time full-band speech enhancement," *arXiv preprint arXiv:1709.08243*, 2017.

[10] Y. Wang, L. Tu, J. Guo, and Z. Wang, "Residual learning based RF signal denoising," in *2018 IEEE International Conference on Applied System Invention (ICASI)*. IEEE, 2018, pp. 15–18.

[11] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar," in *2017 IEEE Radar Conference (RadarConf)*, May 2017, pp. 1703–1708.

[12] S. Chen and H. Wang, "SAR target recognition based on deep learning," in *Data Science and Advanced Analytics (DSAA), 2014 International Conference on*. IEEE, 2014, pp. 541–547.

[13] B. Xue and N. Tong, "DIOD: Fast and Efficient Weakly Semi-Supervised Deep Complex ISAR Object Detection," *IEEE Transactions on Cybernetics*, no. 99, pp. 1–13, 2018.

[14] H. Wang, S. Chen, F. Xu, and Y.-Q. Jin, "Application of deep-learning algorithms to MSTAR data," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, 2015, pp. 3743–3745.

[15] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2017.

[16] D. Barnes, W. Maddern, G. Pascoe, and I. Posner, "Driven to distraction: Self-supervised distractor learning for robust monocular visual odometry in urban environments," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1894–1900.

[17] T. Bruls, W. Maddern, A. Morye, and P. Newman, "Mark Yourself: Road Marking Segmentation via Weakly-Supervised Annotations from Multimodal Data," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[18] S. Hoermann, P. Henzler, M. Bach, and K. Dietmayer, "Object Detection on Dynamic Occupancy Grid Maps Using Deep Learning and Automatic Label Generation," *arXiv preprint arXiv:1802.02202*, 2018.

[19] K. Shoemake, "Animating rotation with quaternion curves," in *ACM SIGGRAPH computer graphics*, vol. 19, no. 3. ACM, 1985, pp. 245–254.

[20] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.

[21] E. Rosten, G. Reitmayr, and T. Drummond, "Real-time video annotations for augmented reality," in *International Symposium on Visual Computing*. Springer, 2005, pp. 294–302.

[22] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.

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

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Fast Radar Motion Estimation with a Learnt Focus of Attention using Weak Supervision
Publication Status	Published
Publication Details	**R. Aldera**, D. De Martini, M. Gadd, and P. Newman, "Fast radar motion estimation with a learnt focus of attention using weak supervision", in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 1190–1196.

Student Confirmation

Student Name:	Roberto Aldera		
Contribution to the Paper	- Developed theory and implemented automatic annotation pipeline for radar - Generated annotated data for training - Network architecture design, experiments plan (absent for final implementation used to generate results) - Contributed to final manuscript, achieving faster processing time by a factor of 2.36 - Designed poster and presented work at conference		
Signature: Roberto Aldera		Date	27/09/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Paul Newman		
Supervisor comments: The contributions are as above and Roberto drove it all.		
Signature 	Date	28/9/2021

This completed form should be included in the thesis, at the end of the relevant chapter.

3.3 Further insights

There are a number of interesting points to discuss that were not included in the original publication but are worth noting here. Since this work was presented, additional progress related to this contribution has pushed the capabilities of radar even further, and is briefly mentioned where relevant.

3.3.1 Choosing what to label

When attempting to mitigate some of the complexity present in radar data, there are two distinctive approaches that are possible with this framework. Firstly, as is the approach in our application, the static, and therefore consistent, parts of the scene can be identified and used so that only portions of the scan that are recognised by this method will be processed. The left panel of Figure 3.1 shows these sections coloured in green. This is useful in a navigation context, where passing only those measurements that are likely to represent real static objects assists downstream components that would otherwise be slowed down by needing to process all observations.

However, the alternative approach that the annotation technique provides may be more useful in other contexts. Instead of labelling elements in the scene that *are* consistent, we could also just as easily flag radar returns in the current scan that are *not* also observed in the rolling window of scans, but have a return power that is similar in magnitude to true positive landmark detections. Figure 3.1 shows these areas coloured in red in the right panel, providing a clearer comparison between this method and the first approach in the original contribution [3]. This effectively flags dynamic objects in the environment, as well as any typical radar complexities like multipath reflections or saturation as discussed in Section 2.4.4. Being able to identify these instances provides a way to screen returns that are distractions for tasks like occupancy grid mapping, path planning, or semantic segmentation. Additionally, it avoids discarding information in the scan that may provide contextual clues that can be exploited by a network, which may be useful in other applications.

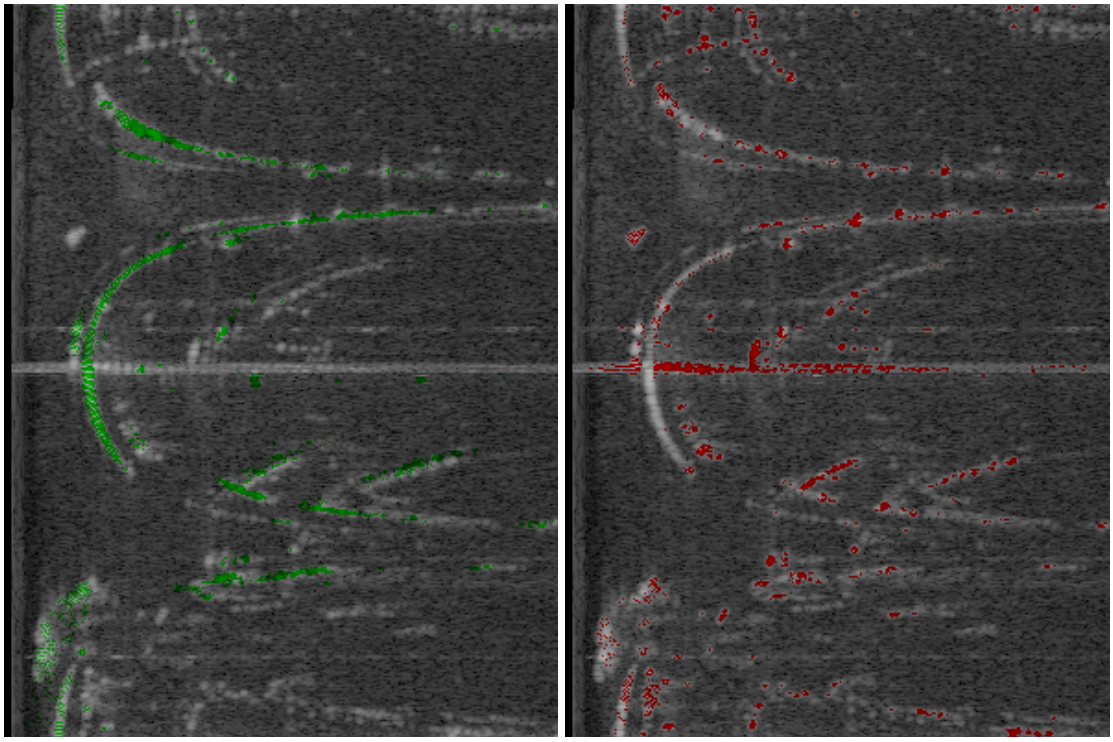


Figure 3.1: Two automatically-annotated versions of the same scan are shown from a drive through central Oxford. On the left, as in our presented method, instances that are consistent between frames and therefore likely correspond to static objects are highlighted in green. Alternatively, as shown on the right, it is also possible to instead identify sections of the current scan that do not appear in the proximal scans but are above a specific power threshold, which can serve to flag undesirable complexities in radar as well as dynamic obstacles.

3.3.2 Odometry source for automatic annotation

The accumulation of radar returns used in the automatic annotation pipeline requires an odometry source to compare data captured at different points in time in separate local reference frames along a trajectory. In the original publication, this is done using our MRG VO implementation as described by Churchill [110]. Additionally, extrinsic calibration is required to relate the radar measurements in the radar's reference frame to the VO that is expressed in the stereo camera frame. This process is detailed in the "*Sensor Calibration*" sections in both the Oxford Robotcar Dataset [111] and Oxford Radar Robotcar Dataset manuscripts [93].

In hindsight, this automatic annotation technique would likely work just as well if we had used our MRG RO as the odometry source, avoiding the additional set

of dependencies that come with VO. However, our MRG RO implementation was still relatively new at the time, and so we chose to rely on our tried-and-tested VO system as an odometry source instead.

3.3.3 Further discussion on original results

The odometry error metrics presented in the manuscript aim to show that this pre-filtering method does not impact ego-motion estimation performance, and produces similar translational and rotational errors to the unfiltered implementation of MRG RO. Pre-filtering the complexities out of the radar measurements speeds up estimation time, which forms the primary contribution in the original work. However, when taking a closer look at the quantitative results, the variance in both the translational and rotational error metrics as well as the estimation time for each system across the three sets of environmental conditions appears to be surprisingly significant, especially given that radar has been repeatedly shown to be impervious to these factors. While there is not too much of a difference between the cloudy and rainy datasets, the night outing exhibits almost twice the translational error across all three systems (unfiltered RO, pre-filtering using ground truth labels, and pre-filtering using U-Net predictions). Additionally, the estimation time more than halves across all systems. As radar measurements are not affected by ambient lighting, it's likely that these differences are due to the reduced traffic after dark in central Oxford.

This potentially has two effects: firstly, it allows for driving at higher speeds, which will magnify any pose estimation errors. And secondly, with fewer passing cars, trucks, and buses, there are not likely to be as many instances of the complexities caused by neighbouring vehicles. This tends to produce radar scans that contain fewer high-power returns, and therefore fewer landmarks. Validating these possible explanations would require further experimental work which is outside the scope of this contribution, but may be of interest in future, particularly for applications in an urban environment where these differences can occur depending on the time of day, such as travelling during rush hour or on a Sunday afternoon.

One final observation to make here relates to some of the small differences in performance between pre-filtering using ground truth labels versus the U-Net predictions. As already mentioned, the automatic annotation pipeline produces imperfect labels as part of this weak-supervision method. The aim is to train a network that learns to generalise on these examples, despite minor inconsistencies in the training data. This makes it possible that at inference time, the network may occasionally outperform the system that uses these annotated examples (our "ground truth") as it is has generalised to recognise certain features in the data that may have been undetected by the imperfect labelling pipeline. One example of this may be the reliance of the annotation pipeline on VO as its odometry source when labelling, which would be expected to be less reliable in rain or night-time conditions. As before, confirming these possible explanations would require more experimental evidence, which we leave for future investigation.

3.3.4 Alternative baselines alongside the learnt approach

The approach in this contribution showed how pre-filtering radar scans speeds up the MRG RO pipeline without negatively impacting the system's performance as measured in the error metrics. There are however a few worthwhile points of interest to discuss further on other avenues of investigation that were not mentioned before.

In order to identify regions of the radar scan that correspond to spurious returns, we use a pre-filtering technique that relies on machine learning tools. This was motivated by the level of complexity of the task of identifying these radar artefacts, where a classical algorithm like the Hough transform [112] would not be expected to succeed. Although the Hough transform has been used effectively in radar applications before to detect lines and other structures [113], it would not necessarily be able to distinguish between returns from real objects in the scene and the undesired spurious returns that appear similar in shape and size. This task requires a system that could differentiate between these two classes of high-power radar returns, which naturally aligns with the strengths of a learning-based approach that is able to utilise the whole scan context when making a decision. So although

it is important to consider a variety of techniques when designing a system such as this before reaching for the popular machine learning tools of today, it is reasonable given the task at hand to formulate this as a deep learning problem.

That being said, it may be interesting to see how this pre-filtering approach performs alongside a crude filter that removes an equivalent number of landmarks at random. This would be expected to have a similar gain in processing speed performance, but at the cost of increased translational and rotational errors in pose estimation. It may also prove worthwhile to experiment with the sensitivity of MRG RO to a process of randomly masking-out portions of the scan as an extension to what has been presented in this contribution, which would give further insight into the significance of the advantage gained by this context-aware pre-filtering system.

3.3.5 Updated related work

The concept of pre-filtering radar scans has seen more recent advances since this manuscript was published. Most notably, Barnes et al. [73] exploited this idea of being able to mask out portions of the scan that were temporally inconsistent between viewpoints to improve both the processing speed and ego-motion estimation accuracy of RO, using an optimised ground truth pose supervisory signal. At runtime, these learnt masks are applied to the input, producing feature maps that contain only those power returns that are expected to be of use for ego-motion estimation.

Similarly to an approach that localises using a LiDAR intensity map [114], a correlation-based matching system is used to find the optimal pose. This combination of masking and correlating scans has proven to be beneficial in both speed and ego-motion estimation accuracy, trained and evaluated on 280 km of urban driving data from the *Oxford Radar Robotcar Dataset* [93], where it has outperformed all other methods to date [73]. While this method performs exceptionally well when trained on this dataset, as is common in many deep learning methods, it would likely not perform to the same standard in a different environment without retraining on hundreds of kilometres of representative data and a corresponding optimised ground truth signal.

On the contrary, our prior contribution focuses on a smaller subcomponent of the MRG RO system and is tasked with the simpler goal of binary mask generation rather than an end-to-end odometry prediction. As such, it can be trained on a single loop around the Oxford 10 km, and generate labels automatically using an unoptimised odometry source for weak supervision. While there are certain key differences between these methods, the notion of learning how to pre-filter radar scans by leveraging ego-motion through a scene is evidently effective, and continues to be an elegant mechanism for handling a radar's complex measurements.

3.4 Contribution review

This section provides a summary of both the novelty and significance of the work, expanding on the case presented in the original manuscript.

Novelty

Automatic annotation has been used successfully in previous applications in the context of mobile robotics, particularly in object detection [115], ephemerality mask prediction [116], and road marking segmentation [117]. Recent related work in MMW radar has also looked at automating dataset curation, specifically for data augmentation rather than annotation, so as to increase the number of samples available for training a neural network in an object classification task [118]. This particular system operated at a higher frequency of 300 GHz with a 20 GHz swept bandwidth, providing range resolution of 7.5 mm to capture more detail from objects in the scene [118].

At the time our contribution was published, this concept of using ego-motion to effectively distinguish between real scene objects and spurious returns in the context of radar was novel, although as already mentioned, it has been used particularly effectively in more recent work [73]. This is arguably due to two reasons: firstly, radar's use as an ego-motion estimation sensor had not progressed to the point of needing a mechanism to remove spurious returns until relatively recently. Secondly, the requirement for this sort of pre-filtering is sensor-specific; simply put, vision

and laser-based methods are not troubled by similar effects to an extent that warrants a specialised module to counter them. Feature extraction algorithms in those domains are particularly adept at selecting points that correspond to stable static features, unlike many of the radar landmark extraction algorithms that are vulnerable to the sensor's complexities. That being said, LO-Net [41] does have a masking component that aims to filter out LiDAR returns that correspond to dynamic objects, but this is not as significant an issue as the undesirable spurious returns are to radar-based applications, where we observed that it was possible to mask approximately half of the detected landmarks.

Significance

The ability to generate vast quantities of annotated data automatically has a number of benefits to our radar application. As mentioned in the results section, we saw a speed increase of almost two and a half times by masking these spurious returns, which brought the runtime of the original algorithm down to the point where processing each scan online became feasible. In addition, this method makes a contribution to the task of localisation, which shares much similarity with ego-motion estimation. Having a way to selectively process only those landmarks likely to correspond to static objects in the scene would arguably lead to an even greater performance boost, in the case of needing to match a live scan to a number of proximal node scans from a map instead of just a single one as in the odometry scenario. Another outcome of this contribution was the automatic annotation process itself. Labelling radar data manually is not only time-consuming, it is also relatively difficult to do to a high standard by hand, as it is not always clear which returns correspond to static portions of the scene when looking at a single scan.

However, this particular contribution is limited in its application to the wider ego-motion estimation context outside of radar, largely due to what was mentioned earlier, in that this issue of difficult-to-detect and widely-present spurious landmarks does not occur to the same extent in other sensor modalities. That being said, if we

want our robots to operate in places and conditions where radar is key, a method that enables us to filter out undesirable returns is always going to be beneficial.

Summary

Omitting spurious returns to lighten the processing burden on MRG RO successfully reduces the time taken to produce odometry estimates without significantly affecting the accuracy of the output. This addresses the first of the three areas identified for improvement in Section 2.7.2. However, if this mask prediction network were not functioning well, it could possibly omit landmarks that are critical for the stability of the ego-motion estimates, which would only be visible afterwards in the odometry metrics. In this state, MRG RO is still without a means to detect these failures, or those that occur in the baseline system when confronted with excessive scene symmetry, occlusions caused by surrounding objects, or rolling and pitching in tricky off-road conditions. Chapter 4 seeks to address this concern, and details a subsystem capable of identifying and mitigating odometry failures.

4

Introspective radar odometry

Contents

4.1 Overview of contribution 62
4.2 Integrated manuscript 63
4.3 Further insights 74
4.3.1 Why use a spectral correspondence technique for radar? 74
4.3.2 Effects of scene symmetry on correspondence success . . 75
4.3.3 Revisiting the Kalman filter component 76
4.4 Contribution review 82

4.1 Overview of contribution

Operating radar in challenging weather and lighting conditions where cameras and LiDAR fail has been discussed in Chapter 2, but there are still many ways in which MRG RO can fail — poor scene observations when tilting, occlusions, or excessive scene symmetry are known to be problematic. Although failure is not always avoidable, being able to detect instances of failure is vital for safe and reliable operation of such a system. This is not simply about failures that arise due to imperfect sensor data like spurious returns that were discussed in Chapter 3, as the radar may have successfully captured the surrounding scene but still not have enough information to avoid failure. Rather, the need here is for a subsystem that

allows us to "know when we don't know" when estimating ego-motion, by having a way of assessing and interpreting the state of the correspondence algorithm when matching two sets of landmarks from a pair of radar scans.

Prior to our contribution, related work has similarly sought to recognise and classify failures based on the shape of specific features. Stewart [119] sought to identify failure in a visual localisation system using these shape-detection techniques, which used the shapes of a custom feature based on their Normalised Information Distance traces to detect poor performance. Other related work in point-based introspection has seen efforts to predict failure by observing the alignment between point clouds after a solution has been proposed, where early attempts sought to quantify some form of post-alignment tolerance threshold to identify poor performance [120]. Another system presented by Nobili et al. [121] approached this problem by trying to anticipate failure rather than detecting it retrospectively, and looked at predicting when overlap between sets of points was critically low (and therefore likely to result in failure).

In this publication, we present a solution by adding introspection to the MRG RO system that builds on some of these previously identified themes [119]–[121]. By interpreting the magnitudes of the principal eigenvector elements in relation to one another (which individually can be interpreted as a confidence value in each landmark association), we show that it is possible to recognise scan matching failures and therefore classify the resulting relative pose estimates internally, without the need to compare them to another source of ego-motion estimation or motion model. To validate the benefits of introspection in this context, a fall-back motion model is introduced such that when failures are detected, the ego-motion estimates can continue to be provided based on the robot's dynamics and the most recent previously successful match, reducing failure count.

4.2 Integrated manuscript

Quantitatively, this contribution **reduced the frequency of failure by 24.7%** in conditions that are challenging for MRG RO. This manuscript was presented

at the 2019 IEEE Intelligent Transportation Systems Conference (ITSC) in Auckland, New Zealand.

One minor editorial correction is noted here that was overlooked in the published manuscript: when describing the operation of the radar, we define N as the number of power readings, or bins, in an azimuth. This N is not the same N used later on to represent the number of preceding eigenvectors to consider, which in hindsight should have been represented by an alternative symbol.

What Could Go Wrong? Introspective Radar Odometry in Challenging Environments

Roberto Aldera, Daniele De Martini, Matthew Gadd, Paul Newman
Oxford Robotics Institute, Dept. Engineering Science, University of Oxford, UK.
{roberto,daniele,mattgadd,pnewman}@robots.ox.ac.uk

***Abstract*— This paper is about detecting failures under uncertainty and improving the reliability of radar-only motion estimation. We use weak supervision together with inertial measurement fusion to train a classifier that exploits the principal eigenvector associated with our radar scan matching algorithm at run-time and produces a prior belief in the robot's motion estimate. This prior is used in a filtering framework to correct for vehicle motion estimates. We demonstrate the system on a challenging outdoor dataset, for which current radar motion estimation algorithms fail frequently. By knowing when failure is likely, we achieve qualitatively superior motion estimates and quantitatively fewer odometry failures. Specifically, we see $24.7\,\%$ fewer failures in motion estimation over the course of a $15.81\,\mathrm{km}$ drive through a difficult, mixed rural-and-urban scene, with lower RMSE in translational and rotational estimates during particularly challenging conditions.**

***Index Terms*— radar, sensing, ego-motion estimation, introspection, field robotics**

I. Introduction

Radar offers the field of robotics a new level of robust sensing. While many of the hurdles in navigation have been cleared using lasers or vision to great success, these tools can only take us so far. Autonomous machines that can function reliably in challenging unstructured environments need a sensor that thrives in all conditions – rain, snow, dust, fog, or direct sunlight. We also need our robots to see further if they are to travel safely at higher speeds or operate in wide open spaces where distinct features are few and far between. These requirements are met in radar, which provides a $360°$-view of the scene and detects targets up to $500\,\mathrm{m}$ away. However, radar data is highly complex, making any sort of inference challenging. Measurements contain significant components made up of multipath reflections, speckle noise, and other artefacts in addition to the radar's internal noise characteristics [1].

Despite this, recent work has presented precise radar-only motion estimation algorithms that consistently outperform visual odometry systems, even in conditions which are favourable to vision [2, 3]. Yet, as we present in this paper, the state-of-the-art Radar Odometry (RO) system with no means of introspection is still prone to a high failure rate. We can distinguish between a few common scenarios for failures:

- Operation in off-road environments where the robot's motion is not truly planar causes large swathes of the scene to disappear and reappear intermittently while the robot is moving [4].

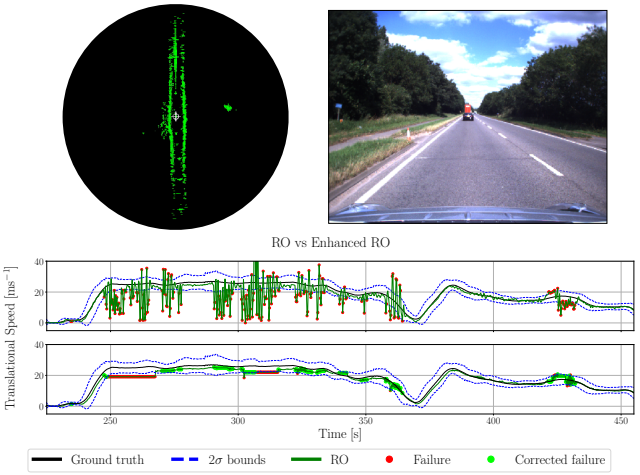


Fig. 1. RO has no means of detecting odometry failures. Here between the hedgerows the radar-only algorithm is starved of distinct features where it is difficult to estimate forward ego-motion between measurements. Our enhanced system addresses this and other common failure cases through introspection, allowing RO to function robustly under the uncertainty presented by challenging environments. Pose estimation is significantly improved, and here translational speeds are shown to closely track the baseline GPS measurements where previously RO had failed.

- Stationary or dynamic occluding obstacles (walls, buses, dense vegetation) encountered in the scene lead to sudden feature absence.
- A shortage of distinct features observable across wide baselines creates ambiguity, as shown between hedgerows in Figure 1.

In this paper, we present a system for both estimating confidence in and correcting the solution of the scan matching RO algorithm. At run-time, we use only the radar data itself to detect failures based on a weakly supervised classification procedure which is trained offline by data fusion. The classification output is passed to a motion correction system that performs an update based on the reliability of the odometry estimate.

We proceed by reviewing existing literature in Section II and discuss some preliminary concepts in Section III which will introduce notation and provide background on the relevant operational aspects of the RO system. Sections IV and V present necessary details for implementation before Section VI, where we report our results showing superior motion estimation performance under extremely challenging conditions.

II. Related works

Approaches to the correspondence problem using eigenvectors have been fundamental in the development of our RO system, first described in [2]. Central to this is the data association method that addresses the challenge of determining matches between sets of landmarks. Scott and Longuet-Higgins [5] present a general approach to cluster features based on the distances between them. Using an initial proximity matrix as input, the algorithm finds the clusters of features based on the eigenvectors of this matrix, ranked using the magnitude of their eigenvalues. An association matrix is computed to yield the strengths of the "bonds" between features, based on methods found in molecular physics. Building on this work, Shapiro and Brady [6] perform feature-based correspondence by incorporating shape information together with these techniques. Their algorithm is robust to translations, rotations, and scaling as it considers the structure within the given image, unlike [5] which fails when features in the first image are no longer neighbours in the next. Closest to RO's data association implementation is the spectral technique of Leordeanu and Hebert [7], which efficiently finds the consistent correspondences between two sets of features using the principal eigenvector of a constructed adjacency matrix. This matrix is built from a graph consisting of nodes representing each potential match, and weights on the respective links that represent the pairwise agreements between these correspondences. This elegant approach sees correct assignments establish strong links between each other and is inherently robust to outliers, while being considerably faster than existing methods.

The data association algorithm implemented in RO that establishes links between landmark sets has no means of introspection. That is to say, an odometry estimate based on the proposed correspondences is made regardless of how well the landmarks have actually been matched. It is suggested in [2] that the portion of landmarks that have been matched provides a measure of uncertainty, which begins to provide indication that a given solution might be unsuitable to use in the next odometry estimate. A more rigorous approach native to the correspondence algorithm itself is discussed in [7], which looks at the difference in magnitude of the first two largest eigenvalues, known as the *eigengap*. If a proposed set of assignments between the points is not particularly distinct, the eigenvalues of the two largest eigenvectors will be similar in magnitude, as they both correspond to similarly plausible solutions. In other words, no single solution emerges above the other possible assignments, as would be the case if a distinctive set of matches could not be found. The authors in [7] note that this occurs when there is excessive symmetry in the data, too many outliers, or high deformation noise. Further to this insight, Sakar and Boyer [8] provide an extensive study on the stability of the eigenvalues and the corresponding robustness of the eigenvectors, and apply this to monitor change in how features are organised in images. In the context of ranking citation information (or web pages), Ng et al. [9] analyse the stability of the eigenvector and provide a simple example that illustrates why a small eigengap corresponds to a less distinctive solution, i.e one that is sensitive to small perturbations in the graph.

To improve RO under failure, we turn to the Kalman Filter (KF), a widely-adopted algorithm that provides a means to estimate under uncertainty [10]. In [11], the authors describe a probabilistic test to detect failures based on the KF estimates, where the system model and the corresponding sensory data should always be statistically coherent. This coherence is modelled as a χ^2-distribution where a simple threshold can be used to detect failures in the sensor measurement stream. A positive failure detection causes the KF algorithm to reject the corresponding sensor measurement.

Our proposed method makes use of an external classifier that flags motion estimation failures prior to the motion correction stage. This addresses the limitations of the approach in [11], where consecutive failures (that lead to an accumulation of uncertainties) result in a system that will be more tolerant to taking on poor sensor measurements.

III. Preliminaries

We use a Frequency-Modulated Continuous-Wave (FMCW) scanning radar which rotates about its vertical axis while sensing the environment continuously through the transmission and reception of frequency-modulated radio waves. While rotating, the sensor inspects one angular portion (*azimuth*, α) of space at a time and receives a power signal that is a function of reflectivity, size, and orientation of objects at that specific azimuth and at a particular distance, ρ. The radar takes measurements along an azimuth at one of M discrete intervals and returns N power readings, or *bins*. We call one full rotation across all M azimuths a *scan* $\mathcal{S}$. Furthermore, let $\mathbf{s}(k) \in \mathbb{R}^{N\times 1}$ be the power-range readings at time step k, where $t(k) = t_k$ is the time value at k and $\alpha(k) \in A$ is the azimuth associated with the measurement. The element $s_i(k) \in \mathbf{s}(k)$ is the power return at the i-th range bin, with $i \in \{1, \ldots, N\}$; its measurement range is given by $\rho_i(k) = \beta(i - 0.5)$, where β is the range resolution of the radar.

Our work builds upon our odometry estimation pipeline first described in [2], which we summarise here. The first aim is to extract features from the sets of power-range spectra measured by the radar as it scans the environment, and describe them as a set of landmarks characterised by a position in space relative to the radar and a power value. This is nontrivial due to the complexity of the radar data itself, which is susceptible to multipath reflections, harmonics, speckle noise, and other hindering effects [1, 3].

Given two so-obtained landmark sets, $\mathbb{L}_1$ and $\mathbb{L}_2$, the next task is to associate each landmark from $\mathbb{L}_1$ with a single corresponding landmark in $\mathbb{L}_2$. As in [7], we pose this data association as a pairwise problem. This leverages the assumption that the set of distances from each point to its neighbours is unique to that point in all observations, regardless of the relative translation and rotation between scans. This allows the landmark to be matched to its counterpart in any other landmark set that contains sufficient overlap.

After extracting a set of possible unary candidates B of size W we build a pairwise-compatibility matrix C, which is a non-negative symmetric matrix of size $W \times W$ containing the score for each pair of proposed matches.

The vector $\mathbf{m} \in \{0,1\}^W$ is a solution of the data association problem, where $m_i = 1$ for a unary match $B\{i\}$ that is considered or $m_i = 0$ otherwise. The optimal solution, i.e. the solution that maximises the overall compatibility, can be computed as:

$$\mathbf{m}^* = \underset{\mathbf{m}\in\{0,1\}^W}{\arg\max}\ \mathbf{m}^T C \mathbf{m} \tag{1}$$

As $\mathbf{m}$ is discrete, the above maximisation is computationally difficult. So inline with [7], we relax this constraint to find the continuously-valued $\mathbf{u}^*$:

$$\mathbf{u}^* = \underset{\mathbf{u}\in[0,1]^W}{\arg\max}\ \mathbf{u}^T C \mathbf{u} \tag{2}$$

The elements in $\mathbf{u}^*$ are only relevant in their respective values to each other, so we can normalise $\mathbf{u}^*$ and from Rayleigh's quotient theorem we iterate to obtain the $\mathbf{u}^*$ that is the principal eigenvector of C [12]. The optimal solution $\mathbf{m}^*$ is approximated from $\mathbf{u}^*$ using a greedy algorithm, as described in more detail in [2]. Finally, the relative motion between the observations that yielded $\mathbb{L}_1$ and $\mathbb{L}_2$ is estimated using a singular-value decomposition algorithm [13]. In addition to the above description, we add a new component to the RO system to account for the motion during a scan that would otherwise cause significant distortions in landmark positions at higher speeds. By assuming constant velocity between motion estimates, landmarks are repositioned relative to the interpolated position of the robot at the time at which they were observed.

IV. Methodology

We begin in Sections IV-A and IV-B by describing how we use inertial measurements recorded alongside radar scans to automatically label a set of eigenvectors from RO as **good** or **bad**. Section IV-D discusses a classification procedure which makes use of the principal eigenvector from the motion estimation algorithm discussed in Section III, allowing us to predict the reliability of the odometry estimate at runtime introspectively without access to inertial measurements. Section IV-E concludes this section by discussing a motion-correcting KF implementation which uses these scores as priors for the motion of the vehicle and propagates the belief through the motion estimation pipeline.

A. Alignment of GPS and radar data

The GPS acceleration readings augmented with INS measurements are integrated twice to obtain a list of relative 6DoF poses. The 6DoF extrinsic calibration between the radar and GPS is used to express these measurements in the radar's frame of reference. As the radar data is interpreted in two dimensions on the xy-plane, the GPS poses need to be projected onto the ground plane [14]. Finally, both radar and GPS pose sources are differentiated to obtain frame-to-frame translational and rotational velocity measurements.

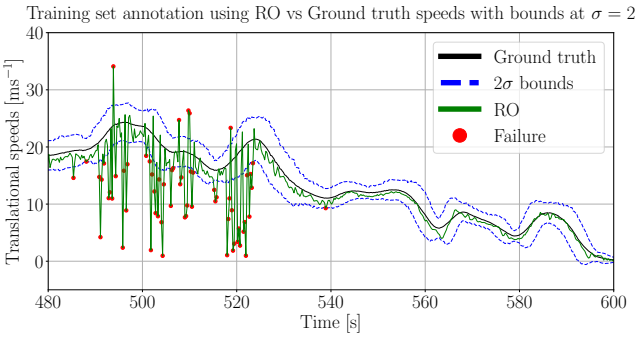


Fig. 2. Aligned instrumentation for GPS and radar sensors used in the training phase of this system. Here we show translational rates. The GPS rates are shown as a moving average, with bounds that lie two standard deviations from the mean. These thresholds are used to trigger labelling events – when the radar rate exceeds them, we produce a **bad** label.

B. Labelling odometry estimates

As a trade-off between quantity and quality of labelled data, we use a form of weak supervision known as *inaccurate supervision* [15], where information in the ground truth data presented to the model may suffer from errors. The GPS readings are preprocessed to allow direct comparison with our RO outputs. We take a moving average over the translational velocities and define an upper and lower bound two standard deviations either side of this. Any RO estimates that fall outside these bounds are labelled as **bad**, or **good** otherwise. Figure 2 shows the signals used in this comparison. After this automatic label generation, we use a median filter to improve training label consistency – during periods of failure these initial labels switch rapidly in and out of bounds, even though the associated principal eigenvectors all correspond to poor matches (evident around 520 s in Figure 2). We refer to these instances in Section IV-C as **brittle** estimates, and discuss further implications in Section IV-D. Note that this framework is flexible to using another baseline signal to produce labels – RTK-corrected GPS traces [16] could have been used as a substitute for our inertial measurements.

C. Eigenvector insights

Each element in the eigenvector can be interpreted as a contribution of a particular pairwise match between two landmarks to the overall proposed alignment. As RO's greedy algorithm from [2] selects matches based on the ordering of these elements, it is natural to want sort these in descending order of magnitude. In Figure 3, we observe that for well-posed data association problems, RO produces eigenvectors that have many similar element magnitudes. Conversely, when there is greater variance among the eigenvector elements, we generally observe a weak match that coincides with a **bad** odometry label. This is intuitive: all things being equal, the algorithm should depend relatively evenly on all available points. But if reliance on these landmarks is not particularly uniform, it indicates abnormally high uncertainty in the compatibility of two landmark sets. This could be a result of a comparison between landmark sets that simply

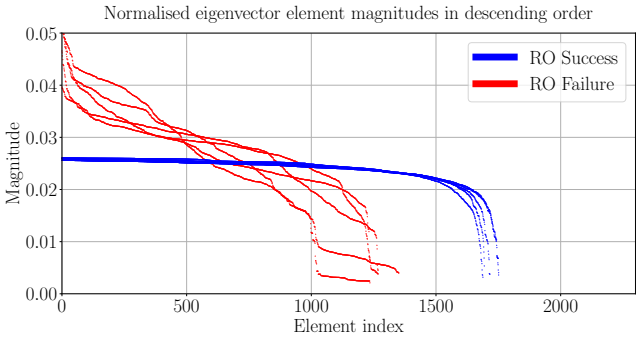


Fig. 3. A sample of eigenvectors produced during the data association step of the RO system. We sort each eigenvector's elements in descending order of magnitude. Odometry estimates labelled **good** and **bad** correspond to blue and red curves respectively. Here, we observe that **bad** eigenvectors typically start off steeper and vanish to zero quicker than their **good** counterparts – this is the decision boundary that the SVM classifier must learn.

do not match well, or more interestingly, it could indicate a **brittle** solution that is well matched in that it produces plausible odometry, but is likely to fail imminently. In our robotics application, these cases are highly susceptible to failure in the presence of any rolling, pitching, occlusion, or other environmental factors that could cause those crucial landmarks to be absent from the next measurement.

D. Training the classifier

Having discussed the origin of the principal eigenvector in Section III and provided insights into what it encodes in Section IV-D, we now turn to training of a classifier. As described in Section IV-B, we have annotated these eigenvectors as belonging to the **good** or **bad** class. To predict the class of a given eigenvector, we train a SVM classifier [17] with an RBF kernel which we found in practice to outperform other commonly used classifiers like random forests and Convolutional Neural Networks in this instance. However, the eigenvectors are not of consistent length as the compatibility matrix is populated by pairwise unary matches for candidate landmarks which vary in number from scan to scan. Along with sorting the elements in descending order, we found that performing a simple feature extraction step on the eigenvector before training improved performance. Other approaches included padding eigenvectors with zeros or naïvely generating a linearly spaced array of the elements to force them to a standard equal length. The selected features are now described and substantiated in detail.

1) **Magnitude of maximum eigenvector element**: the greedy algorithm responsible for approximating $\mathbf{m}^*$ from $\mathbf{u}^*$ uses the pairing related to the highest element magnitude of the eigenvector first:

$$F_1 = \max_i \mathbf{u}_i^*$$

This feature gives an idea as to how strongly the pairwise match of highest confidence contributes to a particular solution. Higher values tend to suppress other pairwise matches and result in fewer pairs being used in the odometry estimate.

2) **Number of candidates**: the number of possible matches, i.e. the cardinality of the unary matches set, also encodes information on the robustness of a solution based on how many pairs of landmarks are available to be associated.

$$F_2 = \|\mathbf{u}^*\| = \min(\|\mathbb{L}_1\|, \|\mathbb{L}_2\|)$$

3) **Area below the curve**: the area bounded by the eigenvector, defined as $\sum_i \mathbf{u}_i^*$, encodes the shape of the curve itself in a compact way. Note that the $\mathbf{u}^*$ vector is defined as a unit vector, such that:

$$F_3 = \sum_i \mathbf{u}_i^* \neq \sqrt{\sum_i \mathbf{u}_i^{*2}} \equiv 1$$

Since the area of this constrained shape is maximised for a broad rectangle created by relatively similar element magnitudes, overdependence on a smaller subset of matches with higher contributions will lead to a smaller area value.

In addition to these features from the present eigenvector, it is possible to add to the feature space the same three features for N previous odometry estimates, which allows the current eigenvector to be placed in context with what has happened in the recent past. This should aid inference – sudden fluctuations in feature values tend to correspond to **bad** RO estimates. Section V discusses the implications of these additional features and details performance in Figure 4 with different values of N.

Prior to training, features are scaled in a standardisation step to have zero mean and unit variance to handle the range of magnitudes and minimise the effects of outliers [18]. We elect to use the tried-and-tested SVM for classification as it is easy to train on our number of features and performs satisfactorily for our application. As with the motion correction module of the pipeline, the choice of classifier for this framework is also flexible. In practice, the SVM learns a class boundary that classifies the majority of curves that look like they lead to a poor solution as **bad**. This is not always the case: sometimes a match that is excessively dependent on a subset of landmarks, or that has other concerning signs already discussed, will in fact lead to a **good** RO estimate (and sometimes be labelled as such, being acceptable to the GPS baseline and slipping past the median filter). This is a significant insight – the classifier will detect these **brittle** solutions, which do not always coincide with **bad** ones. We desire this in the motion correction phase outlined in Section IV-E as it is preferable to defer to a model when high uncertainty in a solution is present, regardless of whether the proposed RO estimate would classify as **good** or **bad**.

Due to the nature of this failure detection problem, many more examples of **good** RO are available in the training data. We select *average precision* as our evaluation metric from the Precision-Recall curve as it gives the clearest insight into binary classifier performance on imbalanced datasets [19]. At run-time, we use the SVM to classify the latest eigenvector from the data association step and treat the

binary classification output as a prior belief in the reliability of the odometry estimate. This failure detector forms the core of our contribution – providing RO with a means of introspection.

E. Motion correction

A KF technique [10] is used to propagate the belief through the motion estimation pipeline. The system model is given by:

$$\begin{aligned} x_k &= A_k\, x_{k-1} + \varepsilon_k \\ z_k &= C_k\, x_k + \delta_k \end{aligned} \tag{3}$$

where x_k and z_k are the state vector and the sensor readings at the k-th instant respectively. We denote the motion and sensor models as A_k and C_k, with corresponding error terms ε_k and δ_k. We exclude the control input matrix, and we consider it as included in the ε_k error term. In this work, we select a constant velocity model for A_k (which could easily be substituted for a more complex model) and use C_k to sample the relative poses from RO directly.

While the prediction process is carried out at each step – yielding the motion estimate μ_m and motion covariance Σ_m – the sensor update is computed only if the eigenvector that produced the proposed match has not been flagged by the failure detector. In that case, the intermediate results μ_s and Σ_s are calculated.

A second stage determines if the proposed sensor update is consistent with the expected motion. Note that although the SVM classifier output is not dependent on the motion model, this additional test relies on it largely. We use the technique described in [11] to calculate a probabilistic index and define the scalar quantity λ as:

$$\lambda = (\mu_m - \mu_s)^T\, (\Sigma_s - \Sigma_m)^{-1}\, (\mu_m - \mu_s) \tag{4}$$

It has been demonstrated that λ follows a χ^2 probability distribution. We set a threshold k on the value such that if $\lambda \geq k$ we reject the sensor reading and update the model with μ_m and Σ_m instead of μ_s and Σ_s.

V. Experimental setup

Test data is collected on the *Oxford RobotCar* platform [20], with the addition of our radar sensor.

A. Hardware

We employ a CTS350-X Navtech FMCW scanning radar without Doppler information, mounted on top of the platform with an axis of rotation perpendicular to the driving surface. It is characterised by an operating frequency of 76 GHz to 77 GHz, capturing 400 azimuth readings per rotation at 4 Hz, with a beam spread of 2°. Each azimuthal slice contains 3768 range measurements at a bin resolution of 0.1752 m. For inertial and GPS measurements, the vehicle is equipped with a NovAtel SPAN-CPT Single Enclosure GNSS/INS Receiver[1] that provides fused GPS and INS measurements.

[1] https://www.novatel.com/products/span-gnss-inertial-systems/span-combined-systems/span-cpt/

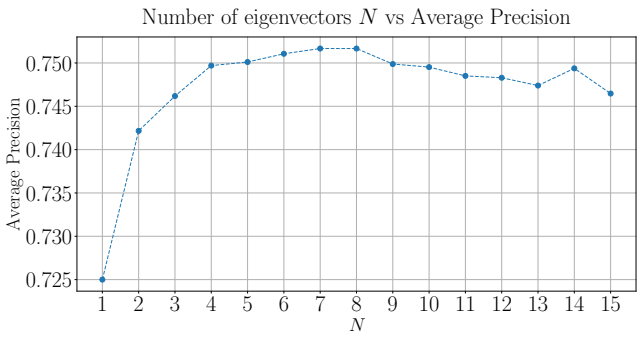


Fig. 4. Plot motivating the choice of hyperparameter N showing the effects on average precision for varying N values. As the number of preceding eigenvectors increases, the classifier performance also rises, until a maximum value where $N = 7$. These average precision values are taken from a mean of the scores obtained on the training set during cross validation using 5 folds.

B. Dataset curation

RO performs excellently in central Oxford where it estimates precise odometry with ease. However, as mentioned in Section I, failures are common when traversing uneven ground or driving between hedgerows at higher speeds. For this reason, we seek to train and evaluate our enhanced RO system on this challenging rural Oxfordshire data. For training, we use radar and GPS data from a single run of a 15.99 km route. At test time, we produce odometry estimates using only radar data as input on a different 15.81 km route in similar surroundings. Training and evaluation were performed in scikit-learn [21] prior to C++ implementation in the enhanced RO system.

C. Training details

We train on a rural Oxfordshire dataset consisting of 5956 radar scans captured during 15.99 km of driving. This produces 5955 labels, of which 21.24 % are labelled **bad** RO. After feature extraction, the dataset is shuffled and split by a 80 % to 20 % training-testing ratio. Using k-fold cross-validation, a grid search that maximised the classifier's average precision yielded a linear kernel with a C value of 1. A further search over values of hyperparameter N determined the optimal number of preceding eigenvectors to use including the latest candidate was 7, as shown in Figure 4. Over the training set, average precision was 75.3 %. The precision was 63.3 % with a recall of 95.5 %.

The testing set sees an average precision of 75.9 %, for a corresponding precision and recall of 62.3 % and 95.7 % respectively. Due to the nature in which RO fails, labels generated automatically using inaccurate supervision as mentioned in Section II tend to fluctuate rapidly between **good** and **bad** during a period of consecutive failures. As already discussed in Section IV-D, the SVM learns a decision boundary that will classify these **brittle** matches as **bad**. Figure 5 shows the effect of passing the labels through a median filter of width 5 where we enjoy a 8.93 % performance boost in our average precision metric.

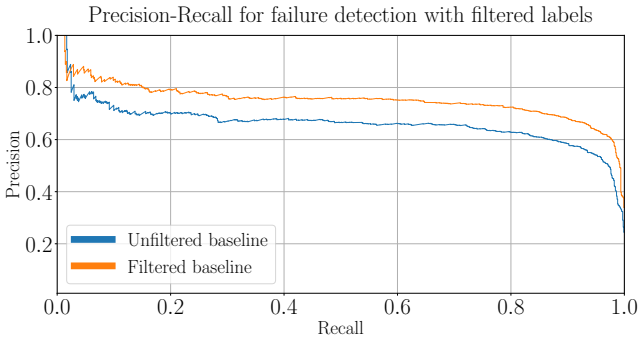


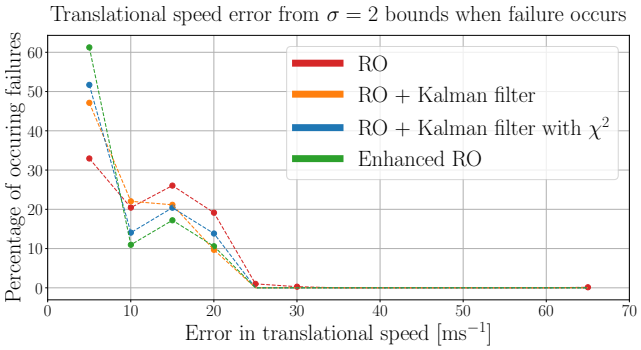
Fig. 5. Two Precision Recall curves are shown. Classifier performance using the labels generated automatically by comparing RO and GPS speeds is improved by passing these labels through a median filter. The orange curve shows this improvement, in line with what has been mentioned regarding the classifier's tendency to flag **brittle** RO as failure in Section IV-D.

VI. Results

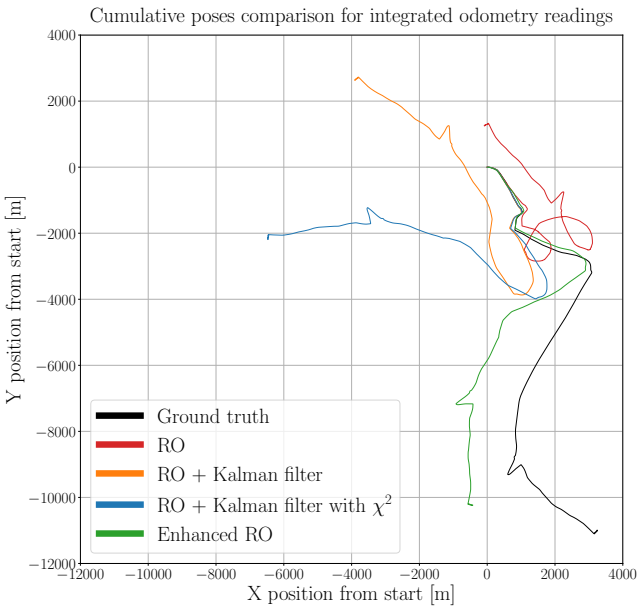
Figure 6(a) motivates our addition of each new section of the system and shows the key result of our overall introspective addition to RO. The combination of the failure detection and motion correction enhancements sees a reduction in the total number of failures over 6635 matches from 695 to 531 (24.7 %), where failures that are still present have reduced from a translational RMSE of $11.112\,\mathrm{m\,s^{-1}}$ to $8.256\,\mathrm{m\,s^{-1}}$ and rotational RMSE of $7.740\,^\circ\,\mathrm{s^{-1}}$ to $2.418\,^\circ\,\mathrm{s^{-1}}$. Table I gives a detailed overview of estimation performance improvements. As illustrated in Figure 6(b), the standard RO system is insufficient in this challenging setting, as the recovered global pose is subject to serious drift and some outright failures. The global pose available from our system more closely matches the shape traced by the GPS readings, which we show here for a qualitative comparison. Figure 7 gives insight into the system's improved performance in recovering translational speeds for each additional component, plotted alongside the GPS baseline and the corresponding bounds.

Despite significant improvement over standard RO performance, the enhanced system's ability to recover a motion estimate under large uncertainties is limited. This is especially prevalent during periods of prolonged failure, where the system is forced to depend excessively on the most recent reliable motion estimate. By deferring to a constant velocity model during periods of likely failure, more complex trajectories can stretch the motion correction process beyond its limits. This is evident in Figures 7(b) to 7(d) where a handful of failures are introduced.

By making use of the principal eigenvectors to detect failures at run-time, our complete system enhances the KF with χ^2 method by rejecting **bad** matches before they can be considered by the χ^2-test. Without this introspection, the filter's tolerance on uncertainty expands to the point of eventually accepting a **bad** estimate, evident in Figure 7(c) between 250 s to 300 s. Note that the addition of the failure detector does introduce a few minor additional failures, as it can be overly sensitive to **brittle** solutions. However, we assert that the system's ability to know when these failures are occurring could reduce future failures if we were able to take action more appropriate to the current state, perhaps in some form of feedback loop. We leave this for future work.



(a) Improvements to RO show that the large majority of remaining failures that do occur will be at lower magnitudes, with our complete system outperforming previous iterations.

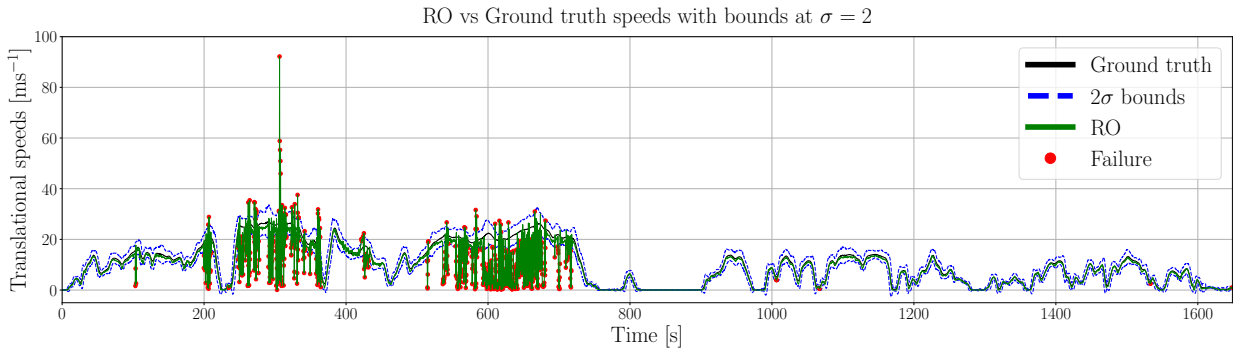


(b) Global pose comparison gives a qualitative result in cumulative odometry improvement in the proposed introspective system from standard RO. By knowing when to reject poor solutions, the enhanced RO system recovers more of the ground trace shape than the naïve KF with χ^2, despite the total number of failures being slightly higher.

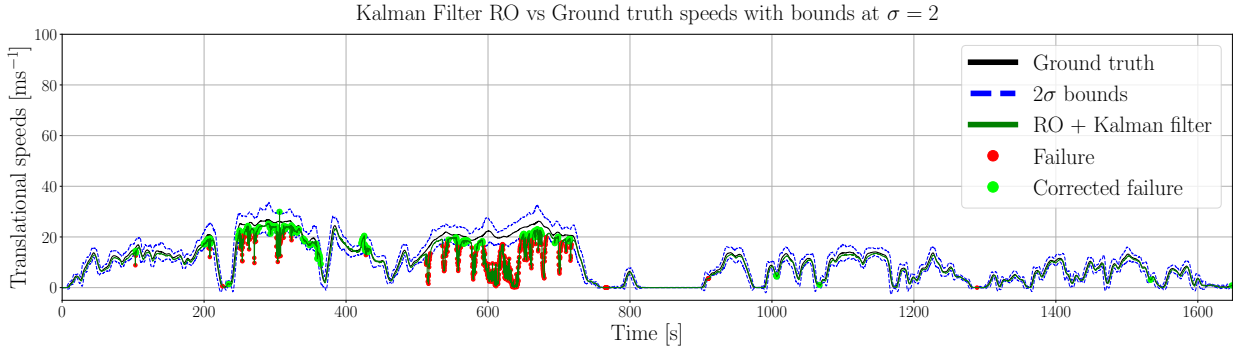
Fig. 6. Comparison between experiments showing quantitative reductions in the distributions of failure case magnitudes and qualitative improvements in the recovered global pose.

VII. Conclusions

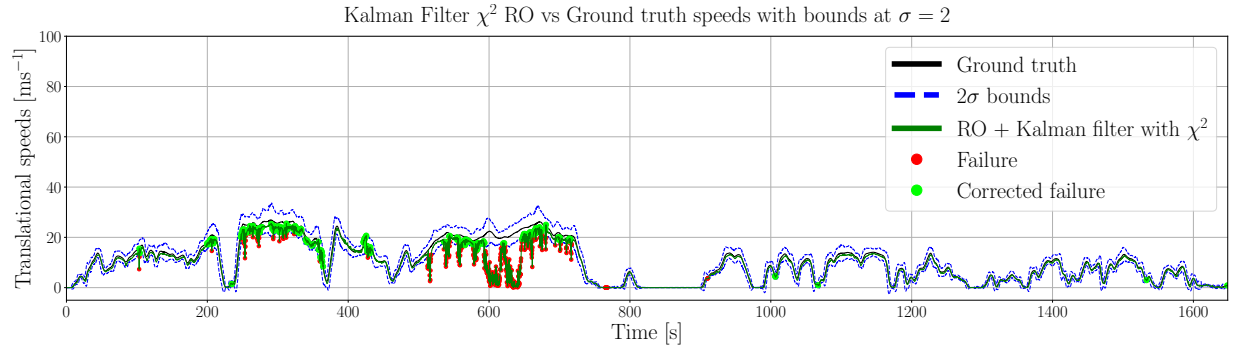
We have presented in this paper a system for online introspective analysis and correction of a radar-only motion estimation scan matching algorithm. The analysis is bootstrapped in a weakly supervised fashion using data fusion, but relies solely on the principal eigenvector associated with the radar motion estimate at run-time. We have shown that the resulting classifier can be used to maintain an evolving prior belief in the motion that the robot undergoes which is exploited by a statistical filter to optimise the quality of the recovered motion. We demonstrated and evaluated our system on an outdoor dataset which challenges the



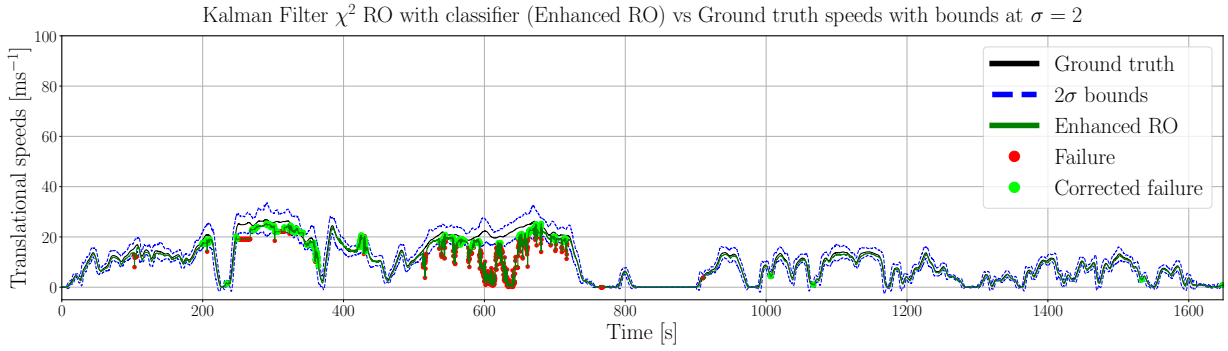
(a)



(b)



(c)



(d)

Fig. 7. Comparison of experiments for performance in estimating translational speed over 6635 matched frames. The standard RO system in Figure 7(a) sees 695 failures. Adding a KF in Figure 7(b) brings this down to 560 failures, with an addition of the χ^2-test in Figure 7(c) reducing this further to 442 detections. Finally, the incorporation of introspection in the form of the classifier output shown in Figure 7(d) adds a small percentage of failures (1.34 %), but these have a lower RMSE during failure and improve performance under prolonged periods of difficulty.

TABLE I

METRICS FOR SYSTEMS UNDER FAILURE COMPARED TO GROUND TRUTH

System	Failure [%]	Translation* [$\mathrm{m\,s^{-1}}$]		Rotation* [$^\circ\,\mathrm{s^{-1}}$]	
		Failure	Overall	Failure	Overall
RO	10.473	11.112	4.755	7.740	3.899
RO, KF	8.440	8.739	3.567	2.271	2.826
RO, KF, χ^2	6.661	9.167	3.263	2.580	2.791
Enhanced RO	8.003	8.257	3.284	2.418	2.794

*RMSE values while under failure, and on the overall performance for this particularly challenging dataset. Please see [2, 3, 4] for performance metrics on other less challenging datasets. We remind the reader that datasets used in this paper were intentionally selected to expose where our previously existing RO system fails and highlight the benefits of the introspection addition.

current state-of-the-art in radar-only motion estimation and produced qualitatively and quantitatively superior RO ego-motion estimation.

ACKNOWLEDGEMENTS

Roberto Aldera and Daniele De Martini are supported by the UK Engineering and Physical Sciences Research Council (EPSRC) programme grant EP/M019918/1. Matthew Gadd is supported by Innovate UK under CAV2 – Stream 1 CRD (DRIVEN). Paul Newman is supported by EPSRC Leadership Fellowship Grant EP/J012017/1. The authors would also like to thank Paul Murcutt for his invaluable software engineering support.

REFERENCES

[1] E. J. M. Adams, J. Mullane and B. Vo, *Robot Navigation and Mapping with Radar*. Artech House, 2012.

[2] S. Cen and P. Newman, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[3] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "Fast radar motion estimation with a learnt focus of attention using weak supervision," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada*, 2019.

[4] S. Cen and P. Newman, "Radar-only ego-motion estimation in difficult settings via graph matching," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada*, 2019.

[5] G. L. Scott and H. C. Longuet-Higgins, "Feature grouping by "relocalisation" of eigenvectors of the proximity matrix." in *BMVC*, 1990, pp. 1–6.

[6] L. S. Shapiro and J. M. Brady, "Feature-based correspondence: an eigenvector approach," *Image and vision computing*, vol. 10, no. 5, pp. 283–288, 1992.

[7] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2. IEEE, 2005, pp. 1482–1489.

[8] S. Sarkar and K. L. Boyer, "Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors," *Computer vision and image understanding*, vol. 71, no. 1, pp. 110–136, 1998.

[9] A. Y. Ng, A. X. Zheng, and M. I. Jordan, "Link analysis, eigenvectors and stability," in *International Joint Conference on Artificial Intelligence*, vol. 17, no. 1. Lawrence Erlbaum Associates Ltd, 2001, pp. 903–910.

[10] G. Welch and G. Bishop, "An introduction to the kalman filter," Chapel Hill, NC, USA, Tech. Rep., 1995.

[11] B. D. Brumback and M. Srinath, "A chi-square test for fault-detection in kalman filters," *Automatic Control, IEEE Transactions on*, vol. 32, pp. 552 – 554, 07 1987.

[12] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.

[13] J. H. Challis, "A procedure for determining rigid body transformation parameters," *Journal of biomechanics*, vol. 28, no. 6, pp. 733–737, 1995.

[14] C. Karney, "Geographiclib," *online at http://geographiclib. sourceforge. net*, 2015.

[15] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2017.

[16] G. S. Gurusinghe, T. Nakatsuji, Y. Azuta, P. Ranjitkar, and Y. Tanaboriboon, "Multiple car-following data with real-time kinematic global positioning system," *Transportation Research Record*, vol. 1802, no. 1, pp. 166–180, 2002.

[17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[18] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc., 2017.

[19] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[20] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	What could go wrong? Introspective radar odometry in challenging environments
Publication Status	Published
Publication Details	**R. Aldera**, D. De Martini, M. Gadd, and P. Newman, "What could go wrong? Introspective radar odometry in challenging environments", in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2019, pp. 2835–2842.

Student Confirmation

Student Name:	Roberto Aldera		
Contribution to the Paper	- Developed theory behind introspective component to use match scores as an indication of overall success - Selected data required for testing failures where RO system fails frequently - Implemented classifier + model corrections - Performed all experiments, reducing failure frequency by almost 25% - Wrote manuscript and created video		
Signature: Roberto Aldera		Date	27/09/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Paul Newman		
Supervisor comments: The contributions are as above and Roberto drove it all.		
Signature 	Date	28/9/2021

This completed form should be included in the thesis, at the end of the relevant chapter.

4.3 Further insights

This section looks at additional considerations that were not presented in the manuscript for publication. The choice of data association technique is discussed in the context of radar, and a simple example is presented to show the effects of scene symmetry on performance.

4.3.1 Why use a spectral correspondence technique for radar?

As mentioned in Section 2.6.1, extracting landmarks from radar data presents unique challenges not encountered in vision or LiDAR systems. Radar-specific techniques like CFAR have been successful in extracting landmarks from radar signals in the past, but have been shown to perform poorly in an application where the radar is mounted on a moving platform, where they fail to capture the salient features in a scene and often produce redundant landmarks [2], [91]. The purpose-built landmark extraction method outlined by Cen and Newman [2] performs well on each scan, but is still prone to detecting landmarks that might not persist between scans. Established point-based data association methods like ICP often used in LiDAR applications are difficult to use for radar, as landmarks tend to be more unstable and are often unsuitable for navigation [3].

The spectral correspondence technique presented in [103] addresses these concerns, exploiting the geometric relations between landmarks to find associations, making it particularly robust to the outliers that are often present in radar landmark sets, due to all of the difficulties encountered in feature extraction while in motion. Connections between correct assignments will tend to establish a highly connected cluster, while the links between incorrect matches tend to occur less frequently in an unstructured way, and so do not form these strong clusters. One clear vulnerability of this method in the context of scan matching is when there is significant symmetry in the scene, making it difficult to determine the correct correspondences. This particular weakness is not specific to this method, and often occurs for related reasons when operating in environments where uniform geometry can fail to sufficiently constrain ego-motion estimates, such as in corridors [66], [121].

4.3.2 Effects of scene symmetry on correspondence success

In order for the spectral correspondence technique presented in [103] to identify the correct associations successfully, a certain degree of asymmetry is required in the point set data. Figure 4.1 illustrates four examples, where the level of asymmetry is gradually increased. The left column contains the layout of the point sets, where the primary points in blue have been shifted slightly to generate the secondary set. On the right, the principle eigenvector elements are shown, where each instance corresponds to a relationship between a point in the primary set and one from the secondary set. These are plotted in their unsorted format in red, and then are sorted by magnitude in blue as they would be when we want to pick the best n matches, which correspond to the n elements with the largest magnitudes.

Ignoring the one-to-one mapping constraint for simplicity, this illustrative example contains 4 points which can be referred to by their indices $i = \{0, 1, 2, 3\}$, where there are 16 possible pairings to evaluate. Here, the elements of the unsorted eigenvector correspond to the relationships between points in the primary and secondary sets with indices:

$$(0,0), (0,1), (0,2), (0,3)..., (3,0), (3,1), (3,2), (3,3)$$

where the correct associations to find would be:

$$(0,0), (1,1), (2,2), (3,3)$$

in this instance. For clarity, those correct correspondences occur at $n = [0, 5, 10, 15]$ which should correspond to elements from the unsorted eigenvector which have the four largest magnitudes. In practice, the list of 16 possible matches would not be neatly aligned, but this is done to avoid a convoluted example.

Initially, when the points are arranged in a square as in the first row in Figure 4.1, it is impossible to distinguish any difference between the proposed matches, as all points are equally compatible with each other. This is catastrophic for our use-case, and represents the worst possible scenario where failure would be inevitable (except for a lucky guess).

The second and third row show how a gradual perturbation of just one of the four points begins to increase the difference in magnitude of the assignments we know to be correct, and greater deformity here leads to a clearer separation between the four best assignments. Lastly, for a distinctly asymmetric set of points, the separation is at its best, which makes for an easy scenario for this technique to pick out those strong associations.

Although only a simplified example, this shows how in scenes where there is a high degree of symmetry relative to the number of observed features, it is challenging to associate landmarks correctly. But as in this contribution, the possibility of incorrect assignments occurring is detectable when inspecting the principal eigenvector elements.

4.3.3 Revisiting the Kalman filter component

Each system evaluated in the results section of the original manuscript builds on the preceding system. Initially we start with our MRG RO, then add a Kalman filter component, which is further extended with a χ^2-test [122], before finally adding the introspection contribution as a final layer to produce "Enhanced RO". Although our primary focus is on the introspection component and not the implementation of the Kalman filter, it is worth expanding on this area in greater detail as it provides insight into the design decisions that arise.

The selection of a constant-velocity motion model for A_k is motivated by the choice to keep this component of the system relatively simple, and avoid the improvements in odometry performance from the introspection subsystem being hidden by the gains from an expertly-crafted filter. However, even a simple filter requires tuning in the form of selecting appropriate process and measurement noise covariances, which is known to be a challenging task [123], [124]. In practice, we did not explore the effects of tuning these parameters, and instead set both of them to the identity matrix, which corresponds to weighting the constant-velocity model and MRG RO estimates equally. While there are established techniques available

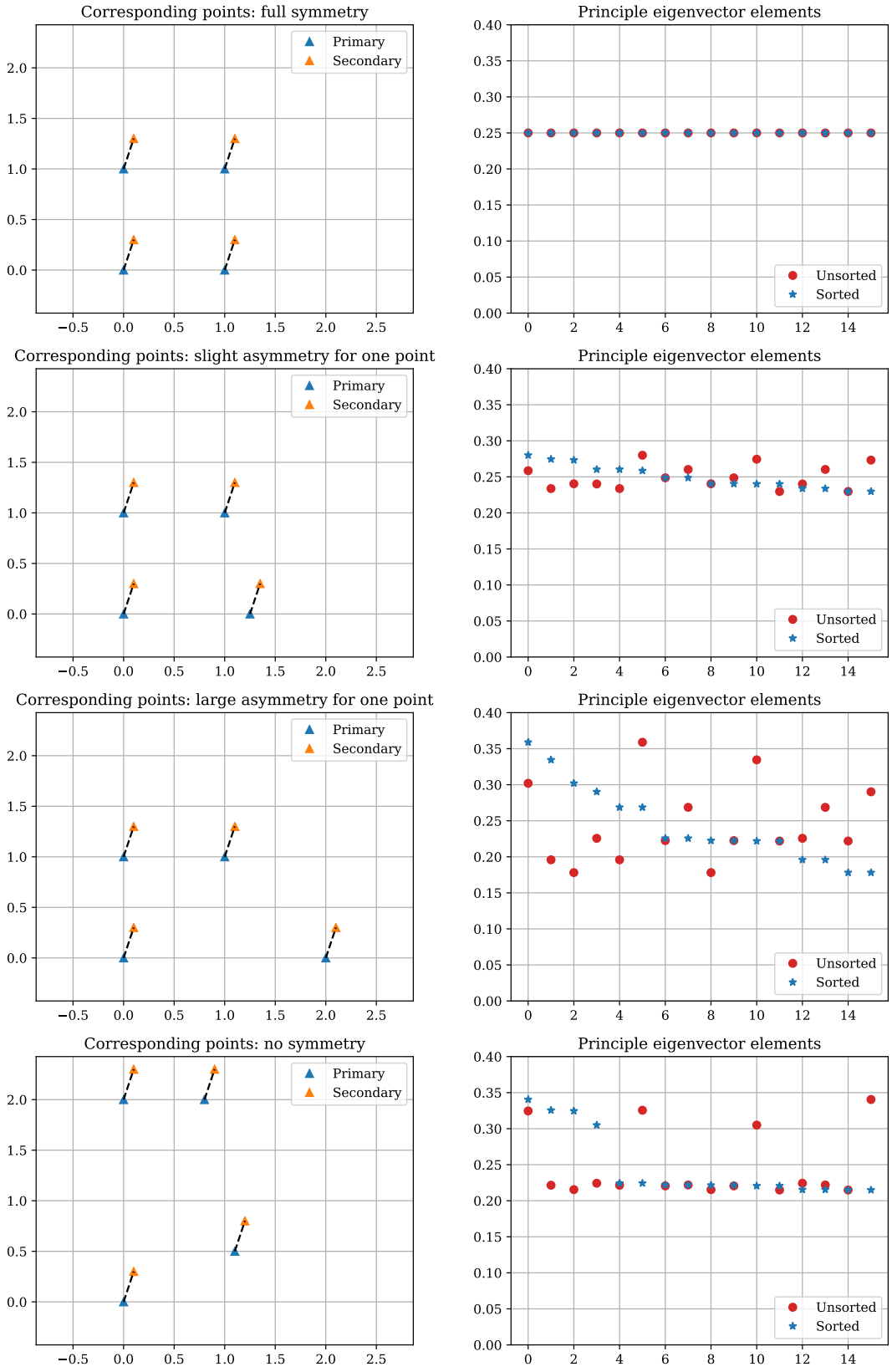


Figure 4.1: Without sufficient asymmetry in the point sets, the spectral correspondence technique struggles to find the four correct associations at $n = [0, 5, 10, 15]$ from the plots on the right. These top-four values become clearer as symmetry diminishes.

for tuning these parameters appropriately, our particular application raises some interesting difficulties that require careful consideration.

We intentionally evaluate performance on data where MRG RO fails extensively (in more than 10% of its pose estimates), and due to the nature of the problem these failures are often concentrated on subsections of the route, rather than being spread out evenly between successful instances. This makes the selection of these covariance matrices challenging — trusting the MRG RO estimates too much leads to poor performance when odometry is unreliable, but going too far the other way and trusting the model estimates too much (so as to aid in sections where MRG RO fails) tends to oversimplify the motion where odometry estimates are in fact reliable. This difficulty unfortunately means that the final tuning that performs well on one dataset will not necessarily transfer successfully to another. The performance depends largely on the frequency of failure in a given sequence for MRG RO, and requires more thought aside from the usual fine-tuning of parameters.

To illustrate some of what has been said, it is possible to re-run our Enhanced RO system with a new set of covariance matrices for the model (Σ_m) and the "sensor" (Σ_s), where varying the weighting between the constant-velocity motion model and our MRG RO system produces slight differences in the final result. The KITTI odometry metrics are used in place of failure counts and RMSE as in the original work to evaluate performance, as they offer a more odometry-specific set of metrics on which to compare each system in line with established literature [44].

Initially we look at the effect of weighting both sources equally, alongside trusting one over the other in turn. These three configurations indicate that for this particular dataset and system setup, the MRG RO estimates are usually more reliable than our simple constant-velocity model, which is unsurprising. Figure 4.2 shows how the segment errors in translation and rotation improve as Σ_s is decreased; in other words, the MRG RO estimates are treated as being more reliable as if they had less uncertainty.

However, there is a limit to this trend, shown in Figure 4.3, where continuing to decrease Σ_s starts to raise the translational and rotational errors again, indicating

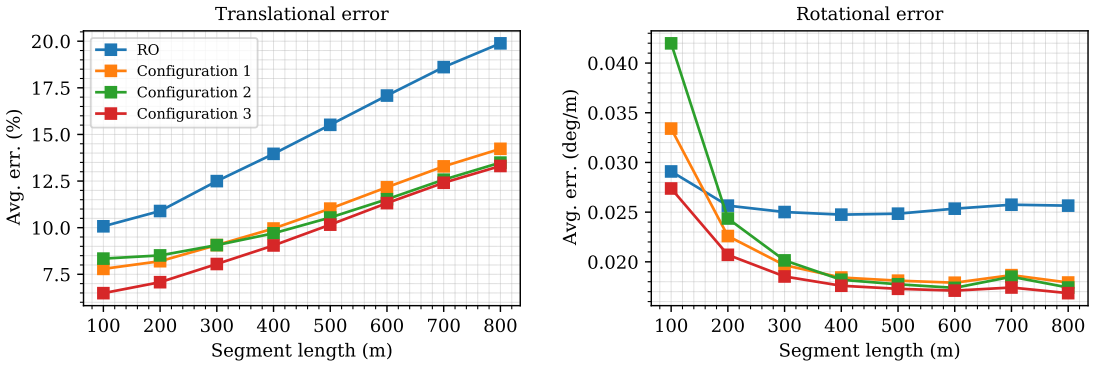


Figure 4.2: Trajectory metrics for MRG RO (labelled "RO") and the first three configurations described in Table 4.1 explore the effects of weighting Σ_m and Σ_s equally, as well as skewing these weights one way or the other. In this instance, these error metrics show that a lower weighting on the uncertainty of the MRG RO estimates results in lower odometry errors, as is the case for configuration 3 in red.

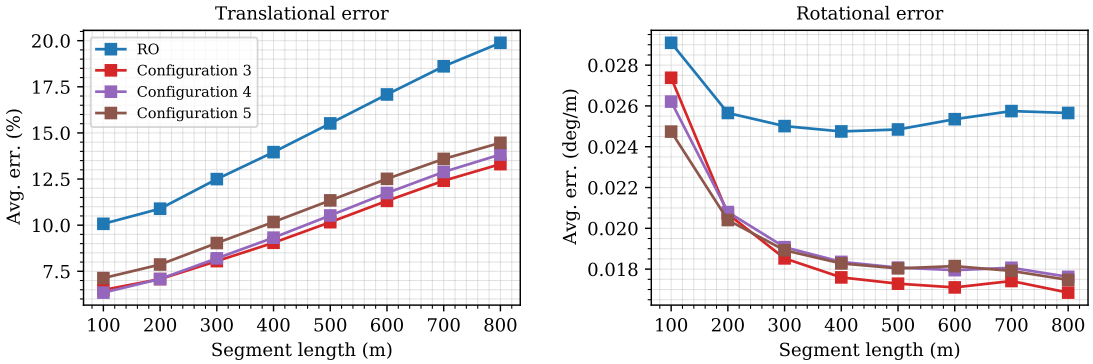


Figure 4.3: Following on from Figure 4.2, if the uncertainty on MRG RO is lowered further, the errors in both translation and rotation begin to increase. Configuration 3 appears to indicate a suitable weighting in this particular case, having performed best overall among the configurations assessed.

that too much confidence is being placed on the MRG RO estimates. Table 4.1 shows the average segment errors for translation and rotation for each configuration as well as MRG RO. It appears that in this instance it would be sensible to use a lower Σ_s value, but the choice of a particular value is likely going to vary based on how frequently MRG RO is encountering difficulties that cause failure.

It is also possible to assess each system's performance qualitatively by studying a subsection of this evaluation sequence. In Figure 4.4, the translational speed estimates from configurations 2, 3, and 5 are plotted alongside our ground truth signal and MRG RO. These three configurations give a sense of the odometry

Table 4.1: Error metrics for different values of Σ_m and Σ_s on a challenging dataset for radar — the same sequence used in the original evaluation in [4].

System	Σ_m	Σ_s	Translation [%]	Rotation [$^\circ\,\mathrm{m}^{-1}$]
MRG RO	-	-	14.8150	0.0258
Configuration 1	1.0	1.0	10.7154	0.02083
Configuration 2	0.5	1.0	10.4689	0.02197
Configuration 3	1.0	0.5	**9.7326**	**0.01911**
Configuration 4	1.0	0.3	9.9922	0.01952
Configuration 5	1.0	0.1	10.0763	0.01924

performance as the filter is tuned, showing how trusting one source over the other has mixed effects on the final outcome depending on how challenging a particular sequence is for MRG RO.

For configuration 2 in green, the speed estimates tend to be nearer to the ground truth signal when MRG RO is failing between 708 s and 718 s. This can be attributed to the model having more influence, which is desirable during challenging sections. However, this leads to poorer performance when MRG RO recovers, shown by how the estimates lag behind the ground truth signal, which is closely approximated by MRG RO once it has recovered.

On the contrary, configuration 5 in brown takes the opposite stance of trusting MRG RO significantly more than the motion model. During the tricky 10 s section it largely follows the erroneous MRG RO which deviates from the ground truth estimates, but then shows excellent performance after MRG RO stabilises and is again closest to ground truth.

Lastly, configuration 3 in red shows how by finding a balance between these two alternatives, the system is able to handle the periods of MRG RO failure and success better than when using a more extreme setting. Estimates from this configuration tend to find themselves between those of the other two, resulting in them being closer to ground truth overall. This illustrates the sensitive nature of tuning these parameters, where settings that favour one source over the other will perform differently depending on how many challenging sections there are in a particular sequence.

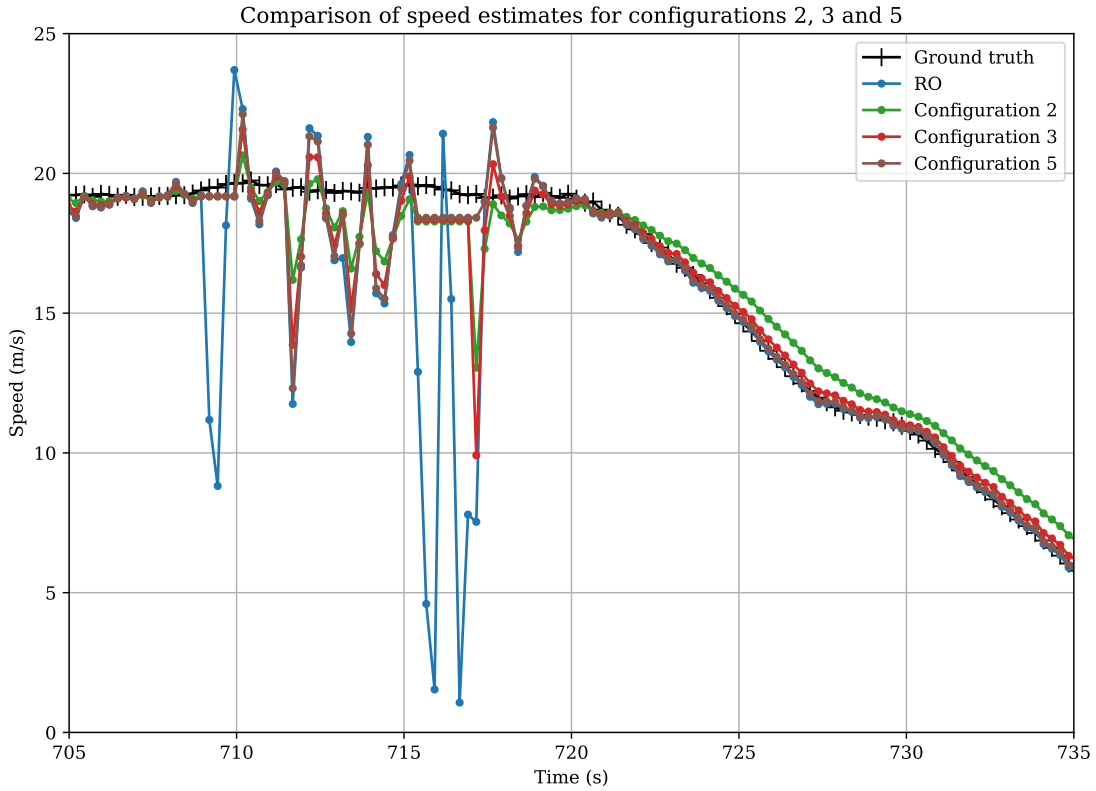


Figure 4.4: By looking at a 30 s snippet of the evaluation sequence, it is possible to get a sense of how the system performs under different configurations. This section includes a period where MRG RO is failing, as well as a stretch where estimates are tracking the ground truth signal reliably. This gives an indication of why configuration 3 in red performs best — when MRG RO is failing, it is relatively conservative (compared to configuration 5 in brown), but once MRG RO has recovered it is still able to track these estimates as desired, unlike configuration 3 in green which is overly sceptical of the odometry estimates from MRG RO.

Perhaps a more suitable approach would be to adjust these parameters dynamically, using a different weighting based on the uncertainty in the most recent estimate, as mentioned by Welch and Bishop [124]. The method described in Section 2.6.5 of this thesis could provide uncertainties for the MRG RO estimates that can be used to update Σ_s rather than keeping it static. To a lesser extent, it may also be worthwhile to update Σ_m during operation, as these model uncertainties may fluctuate dynamically based on the robot's motion during a traversal that might include varied acceleration and cornering [124]. Further investigation into the effects of these additions is left for future work as it is deemed to be beyond the scope of the contribution presented here, but is noted for completeness.

4.4 Contribution review

This section provides a summary of the contribution's novelty and significance and serves to complement the case that the original manuscript presented.

Novelty

Prior to this contribution, MRG RO was not equipped to detect or manage failure cases when they arose. Our intention was to establish a way to recognise a poor scan match result by assessing the overall distribution of corresponding principal eigenvector elements, which provide an indication of the scan match reliability. This particular matching algorithm presented by Leordeanu and Hebert [103] was selected for use in MRG RO due to its robustness to outliers, natural ordering of preferential matches, and overall speed. In their spectral correspondence work, the authors mention that the elements of the principle eigenvector can be interpreted individually as a form of match confidence, and discuss its robustness to small unstructured perturbations as others had identified before [125].

Our particular contribution provides a way to interpret these earlier insights in a new way that is meaningful in the context of scan match health, which had not been exploited in the past. Previously, the notion of using the ratio of selected matches to total observed landmarks had been suggested as a simple indicator of scan match quality, which provides a high-level overview of what might be happening internally [2]. However, by comparing the relative magnitudes of the elements corresponding to each match, it becomes possible for a classifier to distinguish between scan matches that contain some strong and some weak associations (which is good), and those that are either overly dependent on a small subset, or equally confident in all matches. The latter of these occurs when there is a failure to commit to any of the associations as being more trustworthy than their counterparts, illustrated earlier in Figure 4.1.

This data-driven approach opts to learn what failures look like in feature space, rather than requiring careful hand-tuning of parameters to set an appropriate detection threshold manually. Additionally, we found through experimentation that

in this particular context, we could leverage a number of principle eigenvectors from the recent past to infer sudden shifts that may occur under scan match failure. Prior to these insights, only individual matches were considered for their isolated confidence values, rather than the whole set or a recent sequence of sets that indicate overall success [103], [126].

Significance

The primary gain of being able to internally recognise the health of a scan match is itself a significant enhancement, particularly in that it offers a way to determine the reliability of the result directly. To quantify this in our application, we have shown that as part of a larger subsystem, this contribution reduces failure by around 25 % in environments that are particularly challenging for radar. This outperforms a simpler approach of deferring to a Kalman filter [122] to determine whether the system model or measurement should be propagated, as it adds a mechanism for identifying poor scan matches without relying on secondary information like the final pose estimate or overall match ratio [2].

Aside from improving MRG RO's performance in challenging conditions, this new capability of assessing scan match health could be useful in the context of localisation too, where there is benefit to being able to determine how well a live scan matches to a scan captured on a previous traversal of the same location. The system would need to be reconfigured slightly, as the use of the previous N scan matches only makes sense in the context of odometry, although this addition is not strictly required and merely provided a boost in the original implementation [4]. Although this thread is a potential avenue of interest to build on the significance of this contribution, it is left for future work so that the focus on ego-motion estimation development is maintained.

Summary

This contribution targeted the lack of awareness present in our MRG RO system, which had no way to discern whether or not it was experiencing failure. Once these failure cases have been flagged, they can be overridden by estimates from

a motion model that acts as a simple fallback system to handle difficult sections of a traversal. The contribution in Chapter 5 also appeals to the advantage of using a motion model in ego-motion estimation, although instead of imposing it under failure as a last resort, we present a way to refine the subset of candidates that go on to produce these odometry estimates.

5

Leveraging a motion model to increase ego-motion accuracy

Contents

5.1 Overview of contribution 85
5.2 Integrated manuscript 86
5.3 Further insights 97
5.3.1 Selecting matches using bounds for θ-values 97
5.3.2 Attempting to refine match selection 99
5.3.3 Relevance to alternative RO systems 99
5.4 Contribution review 100

5.1 Overview of contribution

Our MRG RO pipeline is versatile: it matches any two radar scans that contain sufficient overlap and estimates a relative pose between them, making it adept at both odometry and localisation tasks. Both of these problems involve matching scans, either between two consecutive observations, or between the current observation and a scan from a map node. Although this generic functionality is an asset, there remains room for improvement in ego-motion estimation accuracy if we specialise the system for odometry. In this case, the current scan is always matched to

the previous scan, captured from the same vehicle moving along a continuous path, which provides a strong prior.

Similar approaches in the past have sought to incorporate a motion model to refine the final ego-motion estimate, as in the work of Almeida et al. [127], which improved a LiDAR-based system's performance by enforcing compliance with a vehicle motion model. The contribution of Scaramuzza et al. [128] in the vision context avoided imposing these motion constraints directly, seeking rather to remove outliers that did not conform to the chosen model's range of motion. This approach is closer to the work presented here, but it was specifically implemented for features captured in 3D with a camera for a VO system.

Other related work by Anderson and Barfoot [36] improved the performance of a RANSAC system by accounting for the motion distortion present in their 2 Hz LiDAR to boost feature matching success. This made use of a constant velocity motion model, and is another example where including additional information on the specific system's range of motion provides a useful constraint to reduce estimation errors.

In this chapter's contribution, we leverage the underlying motion model that physically constrains how a robot is able to move through its environment in $\mathbb{SE}(2)$, so as to reduce the final pose estimate error. This is achieved by assessing whether each individual match produces a plausible set of constant-curvature motion parameters when constrained with a constant-curvature model, before proceeding to find the final pose solution with only the conforming candidates. However, this is accomplished strictly by *leveraging* the motion model, and not by directly imposing it on the output estimates, as would be the case if we were seeking simply to smooth the ego-motion results. Instead, our approach targets the sources of error arising from these outliers.

5.2 Integrated manuscript

The theory and experimental validation are detailed in the manuscript that follows, where the results achieved saw a **45.94% reduction in relative translational**

error over the baseline MRG RO performance. This work has been submitted for review to IEEE Robotics and Automation Letters (RA-L) journal and the 2022 IEEE International Conference on Intelligent Robots and Systems (IROS), which is scheduled to take place in Kyoto, Japan.

What Goes Around: Leveraging a Constant-curvature Motion Constraint in Radar Odometry

Roberto Aldera[1], Matthew Gadd[1,†], Daniele De Martini[1,†], Paul Newman[1]

***Abstract*—This paper presents a method that leverages vehicle motion constraints to refine data associations in a point-based radar odometry system. By using the strong prior on how a non-holonomic robot is constrained to move smoothly through its environment, we develop the necessary framework to estimate ego-motion from a single landmark association rather than considering all of these correspondences at once. This allows for informed outlier detection of poor matches that are a dominant source of pose estimate error. By refining the subset of matched landmarks, we see an absolute decrease of $2.15\,\%$ (from $4.68\,\%$ to $2.53\,\%$) in translational error, approximately halving the error in odometry (reducing by $45.94\,\%$) than when using the full set of correspondences. This contribution is relevant to other point-based odometry implementations that rely on a range sensor and provides a lightweight and interpretable means of incorporating vehicle dynamics for ego-motion estimation.**

***Index Terms*—radar, sensing, radar odometry, ego-motion estimation, motion constraints, field robotics**

I. Introduction

EGO-MOTION estimation using radar-only methods has gained much attention recently in the robotics community. As a sensor that functions consistently in all weather and lighting conditions, radar has become an attractive alternative to traditional vision and laser-based approaches, which are of limited use in poor conditions like fog, rain, snow, direct sunlight and shadows, dust, or darkness. In addition to functioning reliably in diverse environmental conditions, radar has a significantly greater sensing range than its traditional counterparts. This can be of considerable benefit in enabling earlier vehicle and obstacle detection for improved safety or more robust autonomous navigation for mobile robots in environments where stable static features are sparse. For experimental evidence comparing vision and radar in the ego-motion estimation task, see [1, 2].

Building on the foundational Radar Odometry (RO) system developed in [1], this work seeks to leverage a strong prior that previously has not been exploited in this context. Namely, by acknowledging that the robot's motion can be approximated by an underlying model with certain constraints, we aim to improve the accuracy of ego-motion estimates against some ground truth signal. This is achieved by developing a model to estimate ego-motion from a single landmark match and expressing this result as a parameter pair. These parameters fully describe an arc of constant curvature traversed to generate this local observation, allowing each landmark association's odometry estimate to be assessed directly. Once each match is in this form, it becomes easier to identify associations responsible for introducing significant error into the final estimate, which can naturally be refined based on the best available matches as shown in Fig. 1. Examples of these poor matches produced by the baseline RO system are seen in Fig. 2, which provides a clearer zoomed-in view.

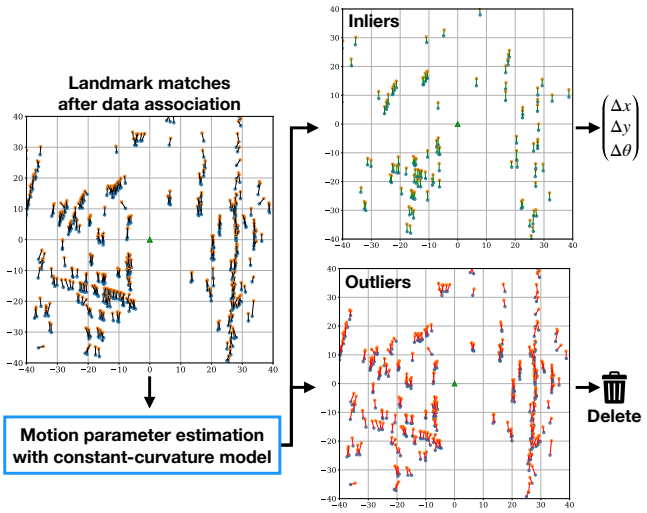


Figure 1. The system estimates motion parameters that result from each landmark match. These parameters provide an indication of match quality and can be classified as part of the inlier set to keep or the outlier set marked for deletion. Odometry accuracy improves when considering only matches corresponding to more plausible ego-motion, as defined by the parameters when under a constant-curvature constraint.

Key contributions in this work are as follows:

- The derivation of a model to solve for relative motion in $\mathbb{SE}(2)$ from a single landmark association, as observed with a range-bearing sensor
- Comparison of performance between the proposed method, a simpler baseline, and the original system without this additional refinement
- Validation of the theoretical contributions on real-world data in complex urban environments [3] alongside a comparison with other RO systems

We proceed by reviewing relevant literature in Sec. II and cover background material in Sec. III, before detailing the method in Sec. IV which illustrates how motion estimates

Manuscript received: February 22, 2022; Revised: May 24, 2022; Accepted: June 16, 2022.
This paper was recommended for publication by Editor Sven Behnke upon evaluation of the Associate Editor and Reviewers' comments.
We are grateful to our partners at Navtech Radar, the Assuring Autonomy International Programme – a partnership between Lloyd's Register Foundation and the University of York – and EPSRC Programme Grant "From Sensing to Collaboration" (EP/V000748/1).
[1]Authors are with the Mobile Robotics Group (MRG), University of Oxford {roberto,mattgadd,daniele,pnewman}@robots.ox.ac.uk
[†]Equal contribution.
Digital Object Identifier (DOI): see top of this page.

are derived from a single landmark observed twice. Sec. V presents the experimental setup used to validate the theoretical contribution of this work, with subsequent results and conclusions covered in Sec. VI and Sec. VII respectively.

II. Related Work

The richness of radar data itself has meant a relatively diverse set of approaches to ego-motion estimation and localisation has been developed. Our system is based on the RO pipeline of [1] and successive related contributions [4, 5]. Learnt approaches to the odometry task have included an end-to-end method from Barnes et al. [6], which uses the cross-correlation between scans to predict pose after removing distractor features, as well as a point-based method that provides greater interpretability [7]. Burnett et al. looked at the effects of failing to correct errors in landmark positions resulting from motion distortion at higher speeds [8] and later presented a hybridised RO framework [9] that combines probabilistic trajectory estimation with learnt features to benefit from both classical and data-driven approaches.

Other work that has made use of radar as a navigation sensor includes that by Park et al. [10] which applies the Fourier-Mellin Transform to log-polar images computed from downsampled Cartesian images, and Kung et al. [11] which uses a normal distribution transform typically applied to 2D and 3D LiDAR. Adolfsson et al. [12] employ filtering to retain the strongest azimuthal returns and compute a sparse set of oriented surface points, while Hong et al. [2] use vision-based features and graph-matching in a radar context.

However, previous work on radar-only ego-motion estimation has directly attempted to leverage the dynamic constraints of the system in operation. This is a natural consideration, as exploiting the underlying smoothness in real trajectories traversed by sensors mounted on non-holonomic robots has proven useful in a wide range of contexts in the past. Using cubic B-splines, Lovegrove et al. [13] looked at leveraging the torque-minimal motion of their visual-intertial system approximated by a parameterised spline, built using relative poses as control points to describe a continuous pose estimate on an interval between discrete poses. This effectively constrains the underlying motion but is not directly applicable to online ego-motion estimation when future pose estimates are unavailable.

Tong et al. [14] developed Gaussian Process Gauss-Newton as a non-parametric state estimation framework where similarly to [13], the continuous-time reality of robot motion is acknowledged when processing measurements. Although our work does not directly impose a continuous-time framework, we do leverage the continuous nature of our robot's trajectory when assuming that it moves smoothly between poses.

As a range-based sensor, radar is in some respects similar to 2D LiDAR, making the work in [15] of interest where authors used a non-holonomic extended Kalman Filter motion model to refine laser scan match estimates. Earlier, the authors in [16] proposed a Visual Odometry (VO) system that solved for Ackerman model parameters directly from feature pixel displacement using a ground-facing camera. However, these methods *enforce* a motion constraint to either directly estimate or smooth the odometry instead of *leveraging* the underlying model to isolate poorer matches as a refinement step.

Scaramuzza et al. [17] present 1-point RANSAC which draws on this idea of exploiting a vehicle's motion constraints in an odometry system and is perhaps the closest work to our contribution, but is only applicable in the vision context. They show how outliers can be discarded in monocular odometry by using a restrictive motion model. However, the details between their work and that which is presented here differ significantly due to the sensor measurement model – radar is a range sensor that measures distances directly, unlike the vision context. We solve this particular problem for range-azimuth sensors, with the full derivation in the appendix. Furthermore, [17] do not use a baseline odometry system to propose their initial correspondences, as we do. Lastly, empirically we found the histogram voting used in [17] performs poorly in this domain and so present a new selection approach using the lower and upper quantiles of the estimated motion parameter. More details of these differences follow in Sec. IV.

III. Preliminaries

We use a Navtech FMCW scanning radar which rotates about its vertical axis and continuously senses the environment using frequency-modulated radio waves. Measurements are captured along an azimuth (α) at one of M discrete angular positions and return a single power reading (a function of reflectivity, size, and orientation of objects) for N range *bins* corresponding to distances, r. We refer to one full rotation through all M azimuths as a *scan* $\mathcal{S}$. Furthermore, let $\mathbf{s}(k) \in \mathbb{R}^{N\times 1}$ be the power-range readings at time step k, where $t(k) = t_k$ is the time value at k and $\alpha(k) \in A$ is the azimuth associated with the measurement. The element $s_i(k) \in \mathbf{s}(k)$ is the power return at the i-th range bin, with $i \in \{1, \ldots, N\}$; its measurement range is given by $r_i(k) = \beta(i - 0.5)$, where β is the range resolution.

Our contribution builds upon the RO pipeline first described in [1], which we summarise here. Firstly, landmark features are extracted from each azimuth's power-range spectrum captured by the radar as it scans the environment. Although this process is not detailed here, it is well known to be difficult in practice due to the complexity of radar measurements themselves, which may often be obscured by multipath reflections, harmonics, speckle noise, and other hindering effects [18].

To compute odometry from the extracted features in two successive landmark sets $\mathbb{L}_{t_k}$ and $\mathbb{L}_{t_{k-1}}$, landmarks in $\mathbb{L}_{t_k}$ must be associated with a single corresponding landmark in $\mathbb{L}_{t_{k-1}}$. We pose this data association as a pairwise problem as in [19], postulating that the distance from each point to its neighbours as observed in one instance uniquely describes that point in any proximal observation, regardless of the relative translation and rotation between observations. This assumption allows landmarks to be matched to counterparts in any other landmark set containing sufficient overlap. Once unary candidates have created preliminary associations and final matches selected using the largest elements from the principal eigenvector of the pairwise compatibility matrix (see [1, 5] for more detail), a final pose estimate G_{t_{k-1},t_k} is computed by a

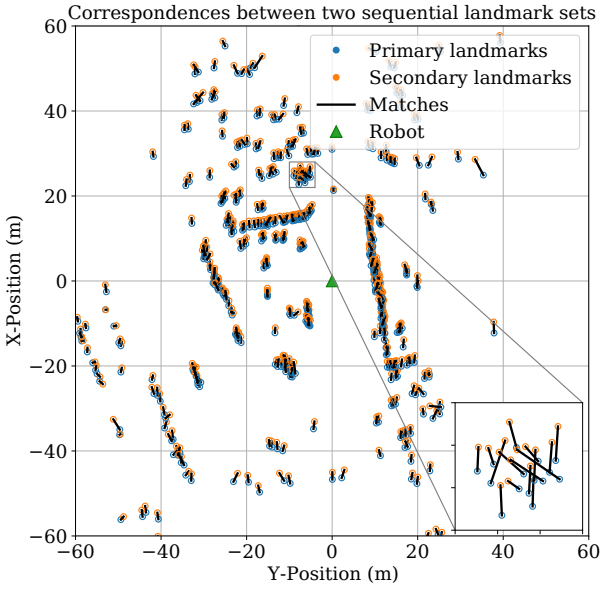


Figure 2. Two successive sets of associated radar landmarks are shown from a scan match, cropped from the full radar range of 165 m to 60 m for clearer visualisation. After observing the orange landmarks, the robot shown at the origin as a green triangle moved straight ahead by roughly 1.4 m and then observed the blue landmarks. Although most matches indicate this motion relatively clearly, there are few stark examples where associated points imply contradictory relative motions, which contribute disproportionately to ego-motion estimate error.

Singular Value Decomposition (SVD) algorithm [20]. For a graphical representation of this RO pipeline, please refer to Fig. 5 on page 5 of Cen and Newman [1].

Our contribution interfaces with the system components described above, taking these associated landmarks as inputs. We seek to refine the association and pose estimation process by considering the physical constraints of the robot platform and leveraging this additional information to improve odometry accuracy.

Fig. 2 illustrates the problem we have identified where it is clear that despite the best efforts in landmark extraction and data association, poor matches will persist. Those shown here are from the baseline system after the data association step, which is tasked with selecting the best candidates. In this approach, a point is associated with the most similar available point based on distances and angles to other points (i.e the scene geometry), which does not consider vehicle kinematic constraints. At first it appears as if an outlier rejection scheme like RANSAC would explicitly remove these outliers, however we show in Sec. VI that a standard application of RANSAC only marginally improves ego-motion estimation accuracy, as it also does not exploit domain knowledge of how the vehicle moves. The effect of failing to remove these poor matches is shown in Fig. 3, where implausible lateral motion is estimated due to poor correspondences.

IV. Method

This section details approaches to mitigating problematic matches in Fig. 2, outlines the proposed model for leveraging a constant-curvature constraint, and describes pose estimation from a refined subset of candidates. We assume the motion of our vehicle is described by a non-holonomic kinematic motion model, appropriate to two-wheel drive urban vehicles which traverse local arcs of constant curvature between two successive measurements, detailed in Sec. IV-B.

A. Outlier rejection baseline

Before leveraging a constant-curvature constraint, we first consider a sensible starting approach that provides a suitable baseline. One instinctive way to improve pose estimate accuracy is to identify and discard poor associations. Matches contributing disproportionately to the final pose error can be flagged as outliers by a simple RANSAC [21] implementation before the SVD calculation. By considering landmark set $\mathbb{L}_{t_{k-1}}$ in the reference frame of $\mathbb{L}_{t_k}$ (which amounts to transforming $\mathbb{L}_{t_{k-1}}$ into the same frame using the computed pose estimate G_{t_{k-1},t_k}), associations can be assessed individually. Practically, RANSAC is run by randomly selecting two matched landmarks and computing a pose estimate with the SVD from this minimal subset. All other candidates are then classified as inliers or outliers based on their range-compensated distance away from their associated partner when under the proposed pose estimate. The sampled model that produces the most inliers after a fixed number of iterations yields the best proposal. These inliers are passed to the SVD method for a final pose estimate. This serves as a baseline to compare our method against, which takes a more informed approach that does not use RANSAC.

B. Constant-curvature motion constraints

Inspired by [17], we seek to describe ego-motion using only a single landmark association. However, in contrast to [17], our formulation does not use pixel coordinates and epipolar geometry required in vision, but instead develops the corresponding theory for range sensors that provide range and azimuth readings. To this end, we present a similar yet novel expression in eq. (A.7) in the appendix for the constraint parameter to be estimated, described below.

When matching two landmark sets to compute a relative pose in $\mathbb{SE}(2)$, at least two pairs of associated landmarks need to be available to constrain the solution. With only a single pair, the relative motion cannot be fully resolved as there would be infinitely many poses on a circle around that landmark that satisfy the single constraint. However, by leveraging a constant-curvature constraint to describe the local robot motion between poses, it is possible to resolve relative motion using only one landmark association pair. This model serves as a suitable approximation for the motion of the vehicle in the experimental dataset (see Sec. VI) which was captured in an urban environment where wheel slippage is minimal. Additionally, the short 0.25 s duration between successive scan measurements means that the curvature of the

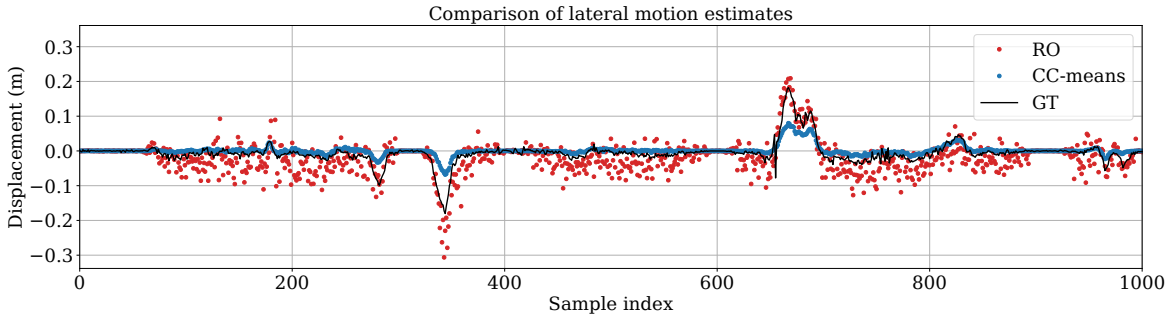


Figure 3. Without leveraging any knowledge of platform constraints, pose estimates from RO in red tend to produce relatively poor performance in lateral motion estimation, particularly evident when driving straight, where estimates deviate by up to as much as 10 cm. Our contribution shown in blue produces smoother estimates in these Δy-values which are a nearer approximation of the ground truth signal in black.

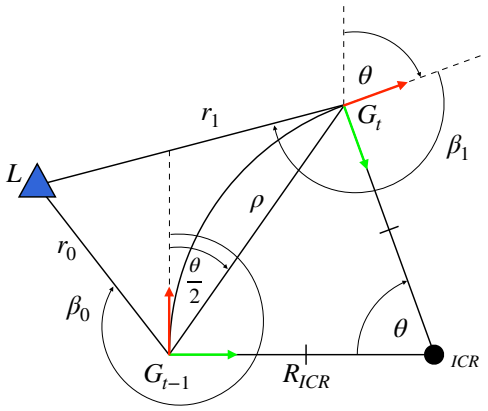


Figure 4. Associated observations of a landmark L from the robot at pose G_{t-1} and then at G_t can be used to describe the local relative motion as a circular arc.

traversed path can be assumed to be constant. Fig. 4 shows two successive robot poses where an arc has been traversed along the circumference of a circle centred at the Instantaneous Centre of Rotation (ICR) of radius R_{ICR}, through an angle of θ. The constant curvature constraint here is applied only between these two poses at which consecutive scans are collected, and does not consider a window of scans or poses before this current estimate. The final pose is a distance ρ from the starting position, with a yaw offset of this same θ. Relative to the starting pose, the final pose is at an angle of $\theta/2$ in this case, observing that both poses are at a distance R_{ICR} from the ICR, forming an isosceles triangle containing θ and two other equal angles. For straight trajectories, the robot can be described as moving along a circle of infinite radius.

C. Pose estimation from a single landmark association

A single landmark observed from the starting pose is at a range r_0 and bearing β_0. At the next instance, once the robot has reached the final pose, that same landmark is observed again, now at a range r_1 and bearing β_1. These four parameters allow us to express the robot's motion along the arc that subtends θ from the ICR where θ and R_{ICR} are expressed as our constant-curvature motion parameters:

$$\theta = 2\arctan\left(\frac{r_0/r_1 \sin\beta_0 - \sin\beta_1}{r_0/r_1 \cos\beta_0 + \cos\beta_1}\right) \quad (1)$$

$$R_{ICR} = \frac{r_1 \sin(\beta_0 - \beta_1 - \theta)}{2\sin(\theta/2)\sin(\theta/2 - \beta_0)} \quad (2)$$

And from these parameters, Δx, and Δy are simply

$$\Delta x = \rho\cos(\theta/2) \quad (3)$$

$$\Delta y = \rho\sin(\theta/2) \quad (4)$$

where $\rho = 2R_{ICR}\sin\theta/2$ and $\Delta\theta$ is the same as its constant-curvature parameter counterpart. The reader is directed to the appendix for a full derivation. Note that, for the edge case when driving in a perfectly straight line, $\theta = 0$ and R_{ICR} is infinite, as mentioned in Sec. IV-B. Although theoretically possible, this tends only to happen in practice when $\beta_0 = \beta_1$ and $r_0 = r_1$ which occurs when the robot is stationary, where we set $\Delta x = \Delta y = \theta = 0$. Fig. 5 attempts to illustrate intuitively how the combination of a single associated landmark and this constant-curvature motion constraint is all that is required to solve for a relative pose estimate.

D. Final pose estimation using a conforming subset

Once each associated landmark match is expressed in terms of θ and R_{ICR}, those that yield θ-values near the median generally correspond to matches that have captured the underlying ego-motion. Associations that produce more extreme θ-values usually correspond to dynamic objects or spurious radar power returns that were similar enough between two observations to pass through the initial data association algorithm. Unlike in [17], we found in this context that simply using the median θ association to calculate the final pose estimate led to inconsistent estimation performance. It appears in [17] that θ-values are discretised, with the middle of the most populated bin taken as the solution. In practice, we found in early experiments that the histogram approach performed poorly in this context. We instead choose to retain a conforming subset and avoid any discretisation errors which may arise. We also experimented with considering the R_{ICR} values when evaluating candidates but found they did not give a strong indication of a candidate's credibility. Based on these observations, we select a subset of matches when calculating

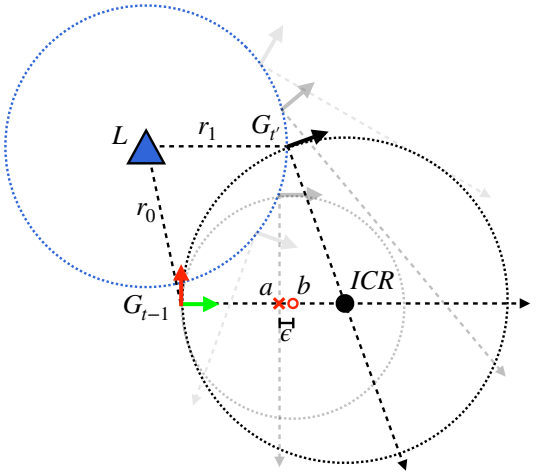


Figure 5. Once a landmark L has been observed initially from pose G_{t-1} to be at a certain range and bearing angle, the task is now to find where $G_{t'}$ is relative to G_{t-1} given a new range and bearing to that same associated landmark from the new pose. This diagram illustrates that although there are infinitely many possible $G_{t'}$ poses that satisfy the second range and angle measurement around a locus, which is a circle centred on L of radius r_1, only one of them will satisfy the constant-curvature motion constraint too. Notice how for the pose slightly below the correct $G_{t'}$ where a faded proposed circle has been drawn, there is an error ϵ between the point a at which their extended axes meet and the centre of that circle b. These points must coincide if this $G_{t'}$ guess were to fully satisfy both the landmark observation and the constant-curvature motion constraints.

the final pose estimate, which can be done by sorting all θ-values in ascending order and taking an inter-quantile set, where specific quantiles to use are left as a hyperparameter. The final pose estimate is produced by passing every match in the selected subset to the SVD algorithm as before.

Fig. 6 shows how when these θ-values are sorted they tend to form a general consensus, with outliers being pushed to the extremities. However, we found that estimates improve if we instead produced Δx, Δy, and $\Delta\theta$ estimates for individual matches in the subset and selected the mean of each of these parameters as our final pose estimate. Note that although we rely on our motion model to assess the reliability of each landmark association, the final estimate is not forced to conform to these constraints and instead is free to describe more complex and realistic robot motions.

V. Experimental Setup

To evaluate our proposed system, we utilise the *Oxford Radar RobotCar Dataset* [3] as its optimised ground truth signal is required to assess if RO accuracy has been improved. Our experimental data consists of the first five sequences of the dataset, totalling approximately $50\,\mathrm{km}$ of driving. Landmarks are extracted and associated as previously described in [1] prior to our proposed refinement process. The use of this RO framework is convenient as it allows us to build on a familiar system, but we emphasise that our contribution is transferable to other point-based odometry implementations that use measurements from a range sensor.

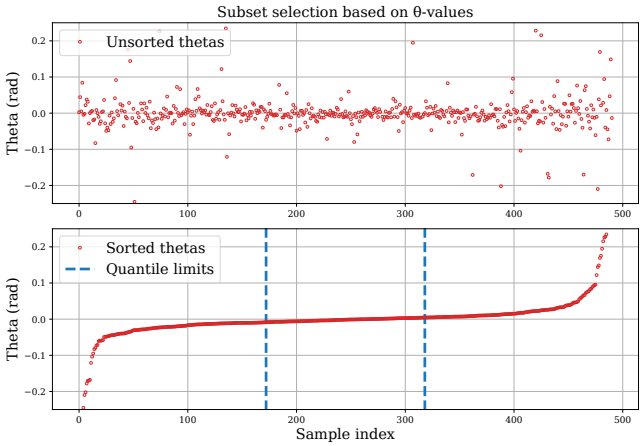


Figure 6. After calculating a θ-value for each individual match between landmark sets from adjacent scans, we can sort them to get an indication on what the majority of matches are estimating. Poor matches tend to produce more extreme θ-values that do not align with the consensus. By taking a subset between two quantiles, we are able to exclude these less reliable (and in some instances more obviously erroneous) matches when computing the final pose estimate.

A. RANSAC baseline

As discussed in Sec. IV-A, we implement a rudimentary RANSAC-based outlier rejection system as a baseline. This provides a reference point in assessing performance when matches have undergone some form of refinement but without leveraging a constant-curvature constraint.

B. Selecting a subset using constant-curvature parameters

When determining which matches to consider in the final pose estimation step described in Sec. IV-D, we found using $\{0.35, 0.65\}$ as the quantile limits yields satisfactory performance for this dataset, slightly more so than using the first and third quartiles. These may need to be adjusted when operating in settings with a different distribution of misleading associations, like in open rural environments where there may be fewer reflections or scenarios with an increased number of dynamic objects in the scene.

C. Processing the selected subset to produce pose estimates

After landmark associations have been reduced to a conforming subset, Sec. IV-D described two methods to compute a final pose estimate. Final poses are estimated using both the SVD method as well as the simpler approach of taking the means of Δx, Δy, and $\Delta\theta$ to illustrate relative performance.

D. Performance metrics

For evaluation we follow the KITTI odometry benchmark [22] which evaluates the distance-normalised translational and rotational error over all possible fixed segment lengths ranging from $\{100\,\mathrm{m}, 200\,\mathrm{m}, ..., 800\,\mathrm{m}\}$.

This approach is commonly used when assessing odometry performance, as it provides a balanced metric that unlike endpoint errors is not overly sensitive to *when* in the sequence an

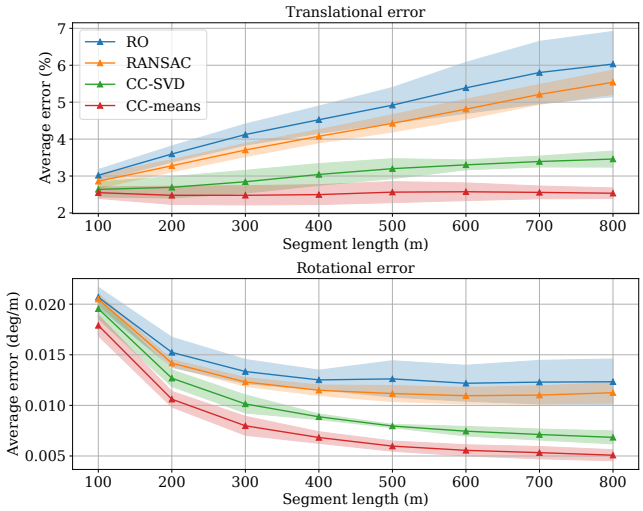


Figure 7. Mean segment errors with corresponding standard deviation envelopes over the course of the first five sequences of the Oxford 10 km route showing translational and rotational error performance for all four methods.

Table I
ERROR METRICS FOR EGO-MOTION ESTIMATION METHODS AVERAGED OVER ALL TESTED SEQUENCES

Method	Tr. [%]	Rot. [°/m]
RO (full matches)	4.6755	0.0139
RO w/ RANSAC	4.2379	0.0129
RO w/ CC*, SVD	3.0717	0.0101
RO w/ CC*, means	**2.5287**	**0.0082**

*Constant-curvature method

error occurs, i.e. an error in a rotational estimate near the start would tend to produce higher end-point errors than if occurs near the end. Our evaluation is based on the implementation provided by the STARS laboratory[1].

VI. RESULTS

Over the five 10 km test sequences, trajectory metrics are summarised in Tab. I for each method's performance. Tab. II details the performance of the four different systems on each of these five sequences to provide a more in-depth set of results.

As expected, because of its widespread application to outlier removal, the naïve RANSAC method performs better than when using the full set of matches. However, this improvement is marginal. Therefore, the implausible associations in Fig. 2 are evidently not dealt with effectively by standard RANSAC. We expect that this is because the initial match selection performed by RO has already selected matches based on their plausibility from a geometric standpoint, which is similar to what this baseline method is doing. However, when leveraging the constant-curvature constraint to select matches, we see a much larger drop in both translational and rotational error. For a direct comparison between methods, the SVD algorithm is used to compute poses on the refined subset and demonstrates the quantitative benefit of removing spurious matches.

Further to this, estimating pose from a single landmark association allows the constant-curvature motion parameters to be converted into their $\mathbb{SE}(2)$ form for every inlier, where we can find the mean of Δx, Δy, and $\Delta\theta$. This outperforms the SVD approach, as it reduces the impact of relatively poor associations more so than the SVD algorithm, which assigns all members of the selected subset an equal weighting.

Taking a closer look at the estimates, we see that the lateral error that is present when we originally use the full set of matches (i.e. *RO (full matches)* in Tabs. I and II) is erratic while our proposed system is steady in its estimates, particularly in areas when there should be smooth forward motion i.e. no lateral displacement, shown in Fig. 3. During sharp turns, there is some underestimation of the lateral motion likely due to the applied quantile limits selecting a narrower set of associations as compared to more common gradual turns or straight trajectories. While being a limitation of our system, this method still tracks the ground truth signal *far* closer than the baseline system (see Tabs. I and II and Fig. 7), which *itself* sometimes over- or underestimates the true motion and presents highly deviant estimates when the vehicle is travelling straight forwards. Stated plainly, the ego-motion estimates that our proposed system produces are more plausible given the real constraints that the vehicle is under – i.e. cars do not suddenly move sideways.

Aggregating over all segment lengths for each sequence, we see from Tab. II translational error in the range of 2.25 % to 2.84 % as compared to 4.26 % to 5.55 % for the baseline RO system, a reduction in error in the range 51 % to 53 %. Similar gains are made for rotational errors.

When looking across all five sequences, our contribution (*CC-means*) has been able to reduce absolute error in translation by approximately 2.15 %, almost halving the original error in RO. Rotational error was just $0.0082\,^\circ\,\mathrm{m}^{-1}$, down from $0.0139\,^\circ\,\mathrm{m}^{-1}$ when using all matches.

Fig. 8 (best viewed in colour) shows the overall trajectories from each method compared with ground truth, which provides more of a qualitative result. Here, the reduction in drift over kilometres of driving can be exemplified by paying attention to the square loop located at approximate coordinates $(200\,\mathrm{m};\ 400\,\mathrm{m})$ in the ground truth trace shown in black. This section of the route occurs near to the end of the 10 km sequence, where the majority of any accumulated error is observable. At this point, our proposed method has drifted approximately 140 m with a few degrees of yaw error, while the original system is more than 450 m off-course and has a yaw estimate more or less perpendicular to the ground truth.

Making direct comparisons with other RO systems is difficult as they publish their results on different sets of either 7 or 8 of the 32 sequences from the *Oxford Radar Robotcar Dataset* [3], as there is no universally recognised "test set". However, the translational error can provide a loose comparison metric, despite being reported on different sequences from the dataset. In achieving 2.53 %, our proposed system brings the baseline RO [1] on which it is applied (having been relegated with 4.68 %) back into

[1]Evaluation metrics implementation available here: https://github.com/utiasSTARS/pyslam

Table II
ERROR METRICS FOR EGO-MOTION ESTIMATION METHODS OVER ALL EXPERIENCES FROM 2019-01-10

Sequence →	11-46-21		12-32-52		14-02-34		14-50-05		15-19-41	
Method ↓	Tr. [%]	Rot. [°/m]	Tr. [%]	Rot. [°/m]	Tr. [%]	Rot. [°/m]	Tr. [%]	Rot. [°/m]	Tr. [%]	Rot. [°/m]
RO (full matches)	4.54	0.013	5.55	0.017	4.26	0.013	4.44	0.014	4.61	0.013
RO w/ RANSAC	4.50	0.013	4.20	0.012	4.09	0.013	4.34	0.013	4.05	0.012
RO w/ CC*, SVD	2.96	0.010	3.15	0.010	3.32	0.011	2.92	0.010	3.32	0.011
RO w/ CC*, means	**2.35**	**0.008**	**2.57**	**0.009**	**2.63**	**0.008**	**2.25**	**0.007**	**2.84**	**0.009**

*Constant-curvature method



Figure 8. The ground traces for the four methods from the 2019-01-10-15-19-41 sequence show accumulated odometry performance over time as compared to optimised ground truth. This figure is best viewed in colour.

competition with more recently developed top-performing RO systems, which include: RadarSLAM [2] (3.11 % to 2.21 %), Barnes' Under The Radar [7] (2.06 %), HERO [9] (1.99 %), CFEAR [12] (1.76 %), and Barnes' Masking by Moving [6] (2.78 % to 1.16 %). This is achieved without any use of learning techniques that introduce additional complications, like those required in Under The Radar [7], HERO [9], and Masking by Moving [6]. Furthermore, our new method is applicable to any point-based RO pipeline (e.g. CFEAR [12]), where it should improve match refinement while introducing relatively negligible additional processing overhead.

VII. CONCLUSION

This work shows that, by considering the physical constraints on how a robot is able to move through its environment, we can drastically reduce both translational and rotational error in RO. Crudely filtering out matches based on the overall consensus of the group is potentially helpful, but ultimately a naïve approach to this problem. Instead, if we consider matched landmarks based on how an appropriately-constrained robot would have had to move in order to generate those associated candidates, the match refinement task becomes far more natural and yields compelling improvements in ego-motion estimation at a low cost.

Specifically, this approach has the advantage of being extremely light-weight: finding the two parameters that describe ego-motion for each match requires two simple equations, before a sorting and subset-selection step refines the set of matches. Solving for the final pose estimate then just involves taking a mean for Δx, Δy, and $\Delta\theta$ which is also inexpensive.

Our contribution makes use of a strong prior on how our vehicle is constrained to move. By folding this information into our system at the match-selection level, we see reduced lateral errors introduced by poor matches that adversely affect odometry estimates. This is achieved without the need to collect hundreds of kilometres of training data which is usually required to learn how a vehicle is likely to move.

Lastly, our method is portable: any point-based odometry system that uses a range sensor (radar, LiDAR) to measure ranges and bearings to landmarks in the environment, captured from a non-holonomic vehicle, should benefit from our contribution presented here.

REFERENCES

[1] S. Cen and P. Newman, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[2] Z. Hong, Y. Petillot, and S. Wang, "Radarslam: Radar based large-scale SLAM in all weathers," in *International Conference on Intelligent Robots and Systems*. IEEE/RSJ, 2020, pp. 5164–5170.

[3] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The Oxford radar robotcar dataset: A radar extension to the Oxford robotcar dataset," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[4] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "Fast radar motion estimation with a learnt focus of attention using weak supervision," in *International Conference on Robotics and Automation (ICRA)*, 2019.

[5] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "What could go wrong? Introspective radar odometry in challenging environments," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019.

[6] D. Barnes, R. Weston, and I. Posner, "Masking by moving: Learning distraction-free radar odometry from pose information," in *Conference on Robot Learning*, 2020.

[7] D. Barnes and I. Posner, "Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[8] K. Burnett, A. P. Schoellig, and T. D. Barfoot, "Do we need to compensate for motion distortion and doppler effects in spinning radar navigation?" *Robotics and Automation Letters*, vol. 6, no. 2, pp. 771–778, 2021.

[9] K. Burnett, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Radar odometry combining probabilistic estimation and unsupervised feature learning," in *Robotics: Science and Systems*, 2021.

[10] Y. S. Park, Y. S. Shin, and A. Kim, "Pharao: Direct radar odometry using phase correlation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[11] P. C. Kung, C. C. Wang, and W. C. Lin, "A normal distribution transform-based radar odometry designed for scanning and automotive radars," in *International Conference on Robotics and Automation*. IEEE, 2021, pp. 14 417–14 423.

[12] D. Adolfsson, M. Magnusson, A. Alhashimi, A. J. Lilienthal, and H. Andreasson, "CFEAR radarodometry — conservative filtering for efficient and accurate radar odometry," in *International Conference on Intelligent Robots and Systems*. IEEE/RSJ, 2021, pp. 5462–5469.
[13] S. Lovegrove, A. Patron-Perez, and G. Sibley, "Spline fusion: A continuous-time representation for visual-inertial fusion with application to rolling shutter cameras." *British Machine Vision Conference*, pp. 93.1–93.12, 2013.
[14] C. H. Tong, P. Furgale, and T. D. Barfoot, "Gaussian process gauss–newton for non-parametric simultaneous localization and mapping," *The International Journal of Robotics Research*, vol. 32, no. 5, 2013.
[15] J. Almeida and V. M. Santos, "Real time egomotion of a nonholonomic vehicle using lidar measurements," *Journal of Field Robotics*, vol. 30, no. 1, 2013.
[16] N. Nourani-Vatani, J. Roberts, and M. V. Srinivasan, "Practical visual odometry for car-like vehicles," in *IEEE International Conference on Robotics and Automation*, 2009.
[17] D. Scaramuzza, F. Fraundorfer, and R. Siegwart, "Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC," in *IEEE International conference on robotics and automation*, 2009.
[18] E. J. M. Adams, J. Mullane and B. Vo, *Robot Navigation and Mapping with Radar*. Artech House, 2012.
[19] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *IEEE International Conference on Computer Vision*, vol. 2, 2005.
[20] J. H. Challis, "A procedure for determining rigid body transformation parameters," *Journal of biomechanics*, vol. 28, no. 6, 1995.
[21] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, 1981.
[22] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

APPENDIX

As referenced in Sec. IV-C, given the ranges r_0, r_1 and bearings β_0, β_1 we can find θ and R_{ICR}. The diagram has the landmark on the left of both relative poses to avoid overcrowding of symbols on the right-hand side, but this derivation is clearer if we avoid reflex angles. As such, we define angles $\phi_0 = \beta_0 - 2\pi$ and $\phi_1 = \beta_1 - 2\pi$ before getting started. Firstly, we express γ in terms of ϕ_1 and θ as

$$\gamma = \phi_1 - \theta \tag{A.1}$$

Then find α by summing angles in the constructed triangle $\Delta P G_{t-1} G_t$ and rearranging:

$$\begin{aligned} \pi &= \alpha + \gamma + {}^{\theta}/_{2} \\ \therefore \alpha &= \pi - \phi_1 + {}^{\theta}/_{2} \end{aligned} \tag{A.2}$$

Then, by the sine rule, we can express

$$\begin{aligned} \frac{r_0}{\sin\alpha} &= \frac{r_1}{\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \\ \frac{r_0}{\sin\left(\pi - (\phi_1 - {}^{\theta}/_{2})\right)} &= \frac{r_1}{\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \\ \frac{r_0}{r_1} &= \frac{\sin\left(\phi_1 - {}^{\theta}/_{2}\right)}{\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \end{aligned}$$

Expanding using the trigonometric identity $\sin(A + B) = \sin A \cos B + \cos A \sin B$ gives

$$\frac{r_0}{r_1} = \frac{\sin(\phi_1)\cos\left(-{}^{\theta}/_{2}\right) + \cos(\phi_1)\sin\left(-{}^{\theta}/_{2}\right)}{\sin\left({}^{\theta}/_{2}\right)\cos(\phi_0) + \cos\left({}^{\theta}/_{2}\right)\sin(\phi_0)}$$

and dividing all terms through by $\cos\left({}^{\theta}/_{2}\right)$

$$\begin{aligned} {}^{r_0}/_{r_1}\left(\tan\left({}^{\theta}/_{2}\right)\cos(\phi_0) + \sin(\phi_0)\right) \\ = \sin(\phi_1) - \cos(\phi_1)\tan\left({}^{\theta}/_{2}\right) \end{aligned}$$

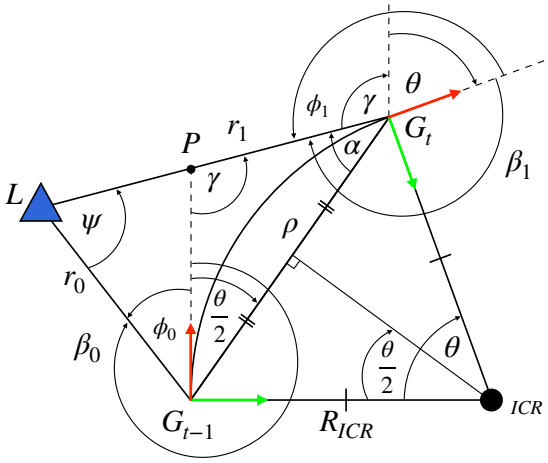


Figure 9. As in Fig. 4, the robot observes a landmark L from two successive poses. Included here are additional labels required to step through the derivation of the constant-curvature motion equations.

then moving θ-terms to the left-hand side

$$\begin{aligned} {}^{r_0}/_{r_1}\tan\left({}^{\theta}/_{2}\right)\cos(\phi_0) + \cos(\phi_1)\tan\left({}^{\theta}/_{2}\right) \\ = \sin(\phi_1) - {}^{r_0}/_{r_1}\sin(\phi_0) \end{aligned}$$

and finally rearranging to express θ as

$$\theta = 2\arctan\left(\frac{\sin\phi_1 - {}^{r_0}/_{r_1}\sin\phi_0}{{}^{r_0}/_{r_1}\cos\phi_0 + \cos\phi_1}\right) \tag{A.3}$$

The diagram in Fig. 9 shows that we can express Δx and Δy in terms of the trigonometric components of ρ, which is computable given ranges r_0, r_1 and angles $\phi_0 = -\beta_0$ and $\phi_1 = -\beta_1$ as before. Firstly, we find angle ψ between the start and end pose from the landmark's position at L by summing angles in the triangle $\Delta L G_{t-1} G_t$ and using Eq. (A.2) to give

$$\begin{aligned} \pi &= \alpha + \psi + \phi_0 + {}^{\theta}/_{2} \\ \therefore \psi &= \phi_1 - \phi_0 - \theta \end{aligned} \tag{A.4}$$

and relating ρ to R_{ICR} as

$$\rho = 2R_{ICR}\sin\left({}^{\theta}/_{2}\right) \tag{A.5}$$

Then using the sine rule, we have

$$\begin{aligned} \frac{\rho}{\sin\psi} &= \frac{r_1}{\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \\ \rho &= \frac{r_1\sin\psi}{\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \end{aligned}$$

and substituting in Eqs. (A.4) and (A.5) to solve for R_{ICR}

$$\begin{aligned} 2R_{ICR}\sin\left({}^{\theta}/_{2}\right) &= \frac{r_1\sin\psi}{\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \\ \therefore R_{ICR} &= \frac{r_1\sin(\phi_1 - \phi_0 - \theta)}{2\sin\left({}^{\theta}/_{2}\right)\sin\left(\phi_0 + {}^{\theta}/_{2}\right)} \end{aligned} \tag{A.6}$$

Lastly, expressing θ and R_{ICR} in terms of our bearings β_0 and β_1 after the above derivation yields:

$$\theta = 2\arctan\left(\frac{{}^{r_0}/_{r_1}\sin\beta_0 - \sin\beta_1}{{}^{r_0}/_{r_1}\cos\beta_0 + \cos\beta_1}\right) \tag{A.7}$$

$$R_{ICR} = \frac{r_1\sin(\beta_0 - \beta_1 - \theta)}{2\sin\left({}^{\theta}/_{2}\right)\sin\left({}^{\theta}/_{2} - \beta_0\right)} \tag{A.8}$$

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	What Goes Around: Leveraging a constant-curvature motion constraint in radar odometry
Publication Status	Under review
Publication Details	R. Aldera, M. Gadd, D. De Martini, and P. Newman, "What Goes Around: Leveraging a constant-curvature motion constraint in radar odometry", 2021, Submitted for review.

Student Confirmation

Student Name:	Roberto Aldera		
Contribution to the Paper	- Developed theory to determine pose from single landmark in SE(2) - Implemented benchmarks and final proposed system - Performed all experiments, reducing translational errors by over 45% - Wrote manuscript that is currently under review		
Signature: Roberto Aldera		Date	27/09/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Paul Newman		
Supervisor comments: The contributions are as above and Roberto drove it all.		
Signature 	Date	28/9/2021

This completed form should be included in the thesis, at the end of the relevant chapter.

5.3 Further insights

In extension of the contribution presented in the manuscript, this section includes additional detail that provides more insight into the new approach.

5.3.1 Selecting matches using bounds for θ-values

Identifying matches that are plausible is relatively straightforward to do when considering a plot of the sorted θ-values that correspond to each association. A poor match tends to produce a θ-value outside the general consensus, and when sorted, these naturally appear at the extreme ends of the plot as shown in Figures 5.1 and 5.2. Conversely, the matches that tend to encode the underlying ego-motion accurately generally agree with each other. These occupy the midsection of the plot of θ-values, and which we want to classify as inliers.

One way to do this would be to include all θ-values within one standard deviation of the mean, excluding the more extreme values from the inlier subset. However, this method tends to include some poor matches that ought to be filtered out, as the standard deviation is dependent on the magnitude of those extreme outliers, which often causes it to inflate and increase the selection bounds.

A more robust selection process that is successful in addressing these leniencies involves using quantiles to determine these bounds. In practice, for this urban environment, we found that using quantile limits $Q_{lim} = \{0.35, 0.65\}$ performed well, as mentioned in the manuscript. These bounds are not sensitive to extreme θ-values, and capture the desired subset despite variations in the vehicle's motion, from scenarios where θ is near zero (travelling straight), to those where θ has significant magnitude (turning around a tight corner).

Although these quantiles are more robust than the standard deviation limits, they can at times appear to be oversensitive. Figures 5.1 and 5.2 show sorted θ-values along with the bounds discussed above, for travelling straight and making a turn respectively. It is not always straightforward to draw the line between θ-values that are plausible and those that are introducing significant error, which can lead to useful information being discarded.

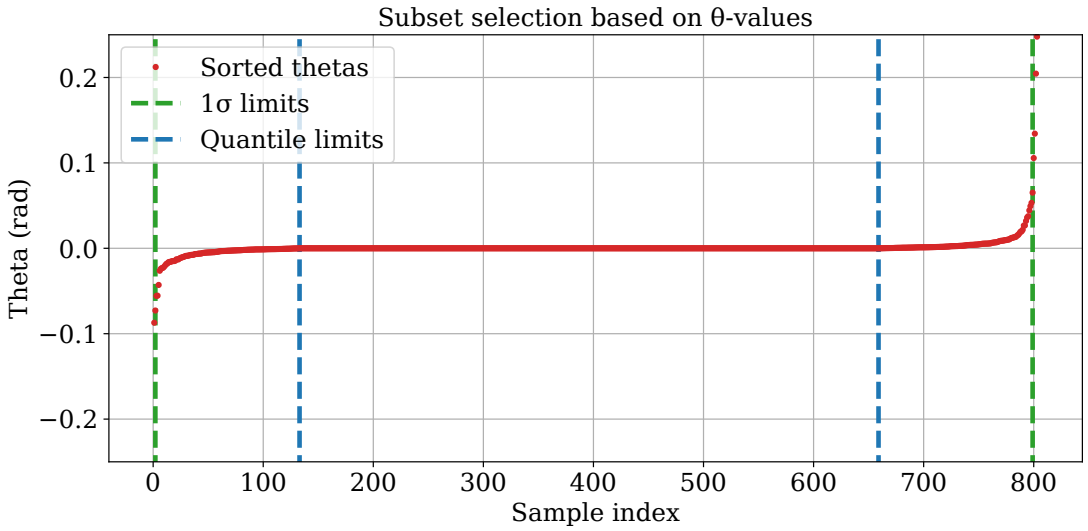


Figure 5.1: Sorted θ-values while travelling straight show a dominant consensus around $\theta = 0$, with a few matches producing deviating values at the extremities of the graph. The bounds set by the quantile limits are stricter than those set by the standard deviation, which in this case is clearly passing a few matches that have a θ-value that is significantly further from zero, which will ultimately introduce additional error into the final odometry estimate.

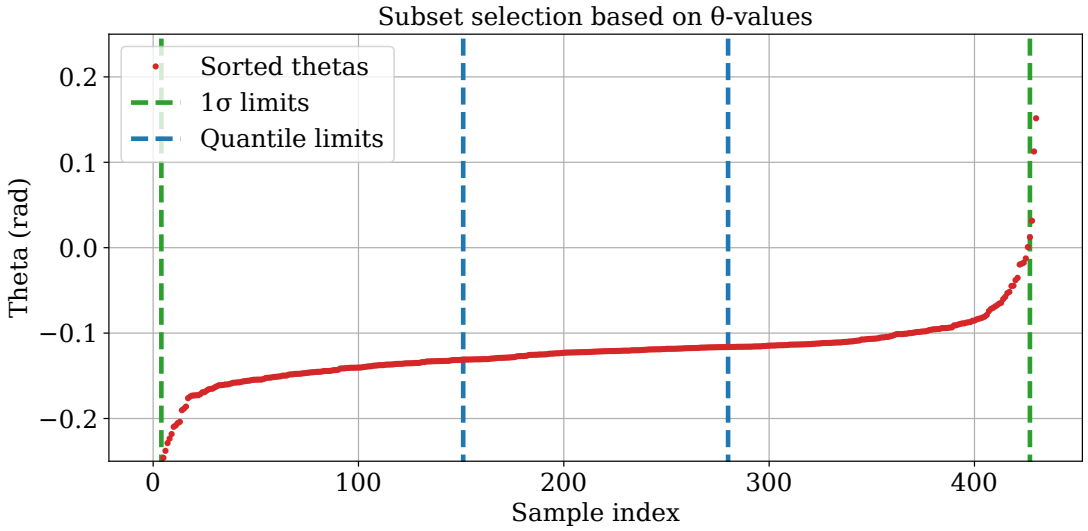


Figure 5.2: Sorted θ-values while turning a corner are shown in this plot, which has almost half of the total matches in Figure 5.1. Again, the limits set by the quantiles are outperforming those determined by the standard deviation. However, it is not immediately obvious how wide the ideal bounds should be, as there is a gradual slope in the θ-values here. The quantiles perform well, but instances like this motivate the case for a more sophisticated system that may be sensitive to those matches that have been excluded by the sometimes overly-cautious quantile bounds.

5.3.2 Attempting to refine match selection

A possible extension here could be to solve this selection problem with a trained classifier that picks out θ-values that appear to be plausible based on thousands of training instances. This selection method could learn to infer the bounds in each instance, or alternatively, to predict a score for each θ-value to determine whether or not it should be included in the subset.

Attempts to teach a network how to select θ-values that would outperform the quantile method have proved more challenging than anticipated, even with more than 32000 training samples captured over approximately 50 km of driving. This classifier-based selection policy is ultimately unsuccessful when pitted against the simpler approach described in the manuscript. In particular, these models failed to handle instances where the θ-values are predominantly non-zero when the vehicle is turning, preferring to favour those values that lie closer to zero. Weighting the sampler during training to sample instances in proportion with their θ-value slightly improved ego-motion estimate performance, but not enough to compete with the quantile bounds method, or even the baseline MRG RO itself. There may be other techniques that clear these challenges in future work which were not discovered here, but for the time being, using these quantile bounds performs best.

5.3.3 Relevance to alternative RO systems

This contribution proves to be very effective in reducing odometry error in MRG RO, as evidenced in the metrics presented in the results section. By refining the set of matched landmarks used to infer ego-motion estimation, we see the performance of this earlier system approach that of more recent point-based RO implementations like HERO [16], CFEAR [91], RadarSLAM [18], and Barnes' keypoint method [89], as mentioned in the original manuscript.

With these results in mind, it is important to note that this contribution is not specifically designed for use in MRG RO alone, which is outperformed by more recent RO implementations like those mentioned above. As a refinement step in a point-based RO pipeline, it should improve other point-based RO implementations

which are not making use of this information explicitly. Given the simplicity of this outlier detection subsystem, the additional processing overhead incurred is minimal in comparison with that associated with the other components in an odometry pipeline, making it an attractive option for use in future RO implementations.

5.4 Contribution review

As in Chapters 3 and 4, we now turn to summarising the novelty and significance of this particular contribution, elaborating on the original manuscript.

Novelty

By considering the underlying motion constraints of the model, this method presents a way to solve for relative motion from a single landmark match. This allows each landmark to be queried for a relative pose estimate, which previously required at least two matches in this $\mathbb{SE}(2)$ context. As mentioned in the manuscript, the concept of estimating motion from a single landmark by leveraging a vehicle model is not novel in and of itself, having been presented in the vision context by Scaramuzza et al. [128]. However, the derivation and method of extracting motion parameters is new for the context of a 2D range sensor, and differs substantially in the formulation of the parameter equations from those presented in [128], where they relate to processing 3D features from the visual scene.

Enforcing a motion model to improve ego-motion estimation accuracy with a 2D range sensor is beneficial, notably in the previously mentioned contribution of Almeida et al. [127], where the non-holonomic vehicle constraint is incorporated with an Extended Kalman Filter model. However, the key difference here is that our contribution seeks to take these constraints into consideration, rather than imposing them directly. By considering each match in the context provided by the motion model, it is possible to detect those matches that are unreliable, enabling us to identify and omit them before estimating ego-motion. This is in some way similar to Hong et al. [18], where the authors select correspondences based on whether the difference in the distance between two landmarks that have been associated across

scans is below some maximum threshold. However, instead of employing this general geometric constraint, which is similarly leveraged in our baseline MRG RO system, our approach appeals specifically to the non-holonomic vehicle's range of motion.

Another minor difference between our method and the one presented in [128] is the way in which the final estimate is computed. Instead of taking the match corresponding to the median θ-value, we found that it was more performant to average each of x, y, and θ and to use the mean values to make up the final ego-motion estimate.

Significance

This contribution shows that detecting and removing poor landmark matches that have made it through the two-part data association stage almost halves the translational estimate error. This is a significant improvement for a system that was already competitive with VO in favourable conditions [2], and adds relatively little computational burden. These improvements in ego-motion estimation error metrics bring our baseline MRG RO system's performance nearer to that of more complex methods, that require additional computational resources and training data to implement their deep learning solutions [16], [73], [89].

Additionally, this method is not radar-specific, and could lead to improvements in LiDAR ego-motion estimation too, where there may be similarly erroneous matches that are difficult to detect and introduce small-scale errors, as observed in our original radar use-case. Although this appears promising, we leave this for future work as it is beyond our particular scope of radar-only ego-motion estimation.

Summary

This contribution addresses the lack of specialisation in our baseline MRG RO system, seeking to refine ego-motion estimation accuracy by leveraging a simple motion model to identify sources of estimate error. Although this comes at the cost of losing MRG RO's generality, the reduction in error metrics by almost half of the baseline system's performance makes a strong case for the inclusion of such a component when applying this radar scan-matching system to the odometry task.

This rounds off our final core contribution to MRG RO, which addresses the last identified area in need of improvement summarised in Section 2.7.2. We now move on to Chapter 6 to discuss these contributions further, and conclude the thesis.

6 Conclusion

Contents

6.1 Summary of contributions 103
6.2 Future work 105
6.2.1 Velocity-aware automatic annotation 105
6.2.2 Pre-filtering radar scans for other tasks and settings . . 106
6.2.3 Reducing the sensitivity of the introspection system . . 106
6.2.4 Integrating an external odometry fallback 107
6.2.5 Introspection for topometric localisation 107
6.2.6 Sparse odometry with match refinement 108
6.2.7 Using individual match quality in a learnt approach . . 108
6.3 Closing remarks 108

6.1 Summary of contributions

Providing mobile robots with the ability to estimate ego-motion using only radar allows them to navigate autonomously in environmental conditions and locations that would otherwise not be possible with more commonly used sensors. The core contributions presented in this thesis provide new ways to augment our baseline MRG RO system so that it runs faster, is capable of introspection and therefore fails less, and is more accurate. This section summarises each of the core contributions and discusses what has been achieved.

The problems presented by the inherent complexity of radar measurements naturally appeal to some of the strengths of learning-based methods which are able to handle these ambiguities. By taking advantage of the robot's ego-motion through a scene it becomes relatively easy to isolate and label the static parts of the environment. The spurious radar returns tend to move erratically which makes them easier to distinguish from stable features. As detailed in Chapter 3, this allows us to obtain a vast quantity of automatically annotated data and avoid the high cost of hand-labelling, simply by overlaying sequential scans that have been shifted by the relative pose between them and accumulating their power returns in a global metric grid. At runtime, processing only those portions of a scan that are likely to contain stable features significantly reduces the overall number of landmarks that are detected and therefore have to be processed by the rest of the pipeline. In our application, this more than halved the overall processing time as compared to the baseline MRG RO system, without adversely affecting the ego-motion estimation accuracy.

Addressing the complexity present in the radar measurement stream is shown to have clear benefits, but does not account for instances where ego-motion estimation is made difficult by factors in the environment. It is vital to know when a system is failing if it is to be used in safety-critical tasks as is often the case in mobile robotics applications. As presented in Chapter 4, understanding how confident the matching process is in its solution makes it possible to interpret the likelihood of failure having occurred, or to predict its imminence. A trained classifier can successfully identify failure based on signals inherent to the data association step, and either flag these to some form of central control system or, as in Chapter 4's contribution, the motion prediction model can cover for instances in which MRG RO estimates are temporarily unreliable. This addition reduces the failure rate of our baseline MRG RO system by nearly 25 %, and provides a way to handle instances where radar's use is otherwise limited.

Although the primary contribution of Chapter 4 centred on a novel introspection technique, much of the quantitative improvement in the final results can be

attributed to the addition of a constant-velocity motion model. Using knowledge of how the system is constrained between scans that are consecutive and proximal in time is not factored into the baseline MRG RO implementation, which explains how even a simple motion model could improve ego-motion estimation performance. However, rather than impose these constraints directly on the odometry estimates to enforce smoothness, the work presented in Chapter 5 seeks to appeal to them and attempt to remove individual matches that would violate these limits. This method successfully identifies those matches that would otherwise introduce significant error, allowing them to be removed prior to estimating ego-motion. By refining the set of matches with this process, the final translational error is almost halved.

Together, these contributions demonstrate how leveraging the wider system-level context of the tasks at hand can lead us to improve the speed, reliability, and accuracy significantly in a radar-only odometry pipeline.

6.2 Future work

The contributions summarised above present solutions to key aspects of radar-only egomotion estimation, but also reveal where there is room for further improvement going forward. Some of these instances are relatively well-scoped minor amendments, while others may inspire further research on radar's use as a navigation sensor in mobile robotics.

6.2.1 Velocity-aware automatic annotation

Faced with a trade-off between the annotation cost and annotation accuracy, data generated via automatic data annotation will always contain some degree of imperfection. In our pre-filtering contribution, we utilise a weakly-supervised approach where these slight inaccuracies are expected and permissible. However, there are some situations in which the method could be improved. All frames in the training dataset are used during automatic annotation, with no distinction made between periods where the robot is in motion and when it is stationary. This results in reflections and other artefacts that are persistent during stationary

periods being labelled together with the true static objects, manifesting as false positives. Addressing this oversimplified assumption would be possible by removing frames from the dataset where the relative pose between frames was below some small threshold, which may lead to a small improvement in mask inference, and subsequent gains in speed and prediction accuracy performance.

6.2.2 Pre-filtering radar scans for other tasks and settings

Although the pre-filtering method was used to speed up ego-motion estimation, it could be extended to other tasks where radar complexities degrade performance, such as localisation, object detection, scene segmentation, or path planning. A natural extension would be to validate this approach in domains other than the urban setting of central Oxford, such as off-road or other rural environments. However, one observation on this point is that the presence of multipath reflections which are particularly prevalent in urban settings because of buildings, buses, and other nearby objects are less likely to appear in open spaces or more rural scenes. This reduces the likelihood of encountering these radar complexities that we aim to mitigate, which raises two considerations.

Firstly, this could present difficulties in data curation where sufficient examples of spurious returns may be hard to come by, resulting in a class imbalance that is problematic for training success. Secondly, there may be limited benefit to operating this filtering component in scenes where these artefacts are so infrequently encountered, and where most of the detected landmarks are likely to be true positives. That being said, this extension may still improve the system's utility in settings where the scene structure is perhaps not as uniform, and where the pre-filter would benefit from improved generalisation.

6.2.3 Reducing the sensitivity of the introspection system

Finding instances in which MRG RO fails consistently (but not catastrophically) is not easy and makes data acquisition for the development and evaluation of the introspection system relatively challenging. It would be interesting to see if

prediction performance and the subsequent odometry estimates could be improved by collecting more data in places where MRG RO is prone to failure — particularly where this failure is intermittent. This could address one of the limitations of the introspection system's classifier, which is occasionally oversensitive to "brittle" scan matches and as a result will in some instances introduce new failures, as is seen in the manuscript's results section.

6.2.4 Integrating an external odometry fallback

When failure is detected by the introspection component, the Kalman filter detailed in the original work is solely responsible for producing ego-motion estimates for the duration of odometry failure. This presents another limitation, as the model's ability to estimate ego-motion accurately deteriorates with each successive instance where there is no sensor update. Using an external odometry source such as VO may handle these periods of prolonged failure better, but would of course be reliant on there being favourable environmental conditions for vision. Another option would be to incorporate an inertial sensor which may be able to cover these periods of temporary MRG RO failure.

6.2.5 Introspection for topometric localisation

As outlined in Section 4.4, another application of the introspective component would be in a task like topometric localisation, which also involves matching landmarks in radar scans. Usually, these systems obtain the nearest subset of nodes in a map and compute relative pose between stored radar scans and the live scan. However, they have no way of quantifying how strong the correspondences are between any two sets, and so they usually default to the node that appears to be nearest in metric space. Perhaps, with a system that could quantify match confidence, using the node corresponding to the highest-scoring scan match would improve localisation performance, as well as give an indication of whether that best-performing match should be trusted. This is a relatively natural extension, where

the gains already made in an odometry application are relatively transferable to the localisation domain.

6.2.6 Sparse odometry with match refinement

The ability to detect and ignore poor matches in Chapter 5 has not yet been fully exploited — it would be interesting to cut down on some of the time-consuming steps in the MRG RO pipeline that usually serve to maintain accuracy in the odometry estimates, and instead rely more so on leveraging the motion model to do this refinement prior to final pose estimation. This should provide a compelling boost in speed without compromising on ego-motion estimation accuracy, much like the outcome of Peretroukhin and Kelly's DPC-Net [35], but without the complexities that come with this sort of deep learning method.

6.2.7 Using individual match quality in a learnt approach

Using the framework in Chapter 5 allows each individual match to be assessed based on two motion parameters. This could be useful in a data-driven approach as part of a loss function, where a system could be trained to do tasks based on how well it manipulates these parameters. One radar-related instance in which this may prove worthwhile would be in devising a means to learn small corrections to landmark positions. There are a number of factors that can cause inaccuracies in range measurements of landmarks, some of which are described and specifically targeted in [47]. It may be feasible to train a network to correct individual landmarks based on how plausible the resulting ego-motion they would produce appears to be, which would form a natural extension to this otherwise learning-free contribution.

6.3 Closing remarks

This thesis presents new contributions that appeal to using a systems-driven approach in improving a radar-only ego-motion estimation system. These advances all build on our MRG RO pipeline [2], a radar-only system that can operate in environmental conditions where traditional sensors fail. They significantly extend

its utility in the mobile robotics context where improving speed, reliability, and accuracy are critical to real-world operation. Further applications and avenues for future work described above are now open for these methods to see use in other RO frameworks and related systems. With these advances that capitalise on useful systems-level information that was previously mere unharnessed data, we take another step towards seeing pervasive autonomous robots in our society do the things we can't, don't want to, or just aren't very good at.

References

[1] S. Kyberd, J. Attias, P. Murcutt, C. Prahacs, M. Towlson, S. Venn, A. Vasconcelos, M. Gadd, D. De Martini, and P. Newman, "The Hulk: Design and development of a weather-proof vehicle for long-term autonomy in outdoor environments", in *Field and Service Robotics*, Springer, 2021, pp. 101–114.

[2] S. H. Cen and P. Newman, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions", in *International Conference on Robotics and Automation*, IEEE, 2018, pp. 1–8.

[3] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "Fast radar motion estimation with a learnt focus of attention using weak supervision", in *International Conference on Robotics and Automation*, IEEE, 2019, pp. 1190–1196.

[4] ——, "What could go wrong? Introspective radar odometry in challenging environments", in *Intelligent Transportation Systems Conference*, IEEE, 2019, pp. 2835–2842.

[5] R. Aldera, M. Gadd, D. De Martini, and P. Newman, "What goes around: Leveraging a constant-curvature motion constraint in radar odometry", 2021, Submitted for review.

[6] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to autonomous mobile robots*. MIT press, 2011.

[7] D. Scaramuzza and F. Fraundorfer, "Visual odometry Part I: The first 30 years and fundamentals", *Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[8] D. R. Andrews, "Ultrasonics and acoustics", in *Encyclopedia of Physical Science and Technology*, R. A. Meyers, Ed., Third Edition, New York: Academic Press, 2003, pp. 269–287. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/B0122274105008000`.

[9] P. F. McManamon, *LiDAR Technologies and Systems*. SPIE, 2019.

[10] D. Wu, J. Chae, A Lambert, and F. Zhang, "Characteristics of CALIOP attenuated backscatter noise: Implication for cloud/aerosol detection", *Atmospheric Chemistry and Physics*, vol. 11, no. 6, pp. 2641–2654, 2011.

[11] C. Linegar, "Vision-only localisation under extreme appearance change", PhD thesis, University of Oxford, 2016.

[12] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere", *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[13] R. H. Rasshofer, M Spies, and H Spies, "Influences of weather phenomena on automotive laser radar systems", *Advances in Radio Science: ARS*, 2011.

[14] M. Bijelic, T. Gruber, and W. Ritter, "A benchmark for lidar sensors in fog: Is detection breaking down?", in *Intelligent Vehicles Symposium*, IEEE, 2018, pp. 760–767.

[15] M. Kutila, P. Pyykönen, H. Holzhüter, M. Colomb, and P. Duthon, "Automotive lidar performance verification in fog and rain", in *International Conference on Intelligent Transportation Systems*, IEEE, 2018, pp. 1695–1701.

[16] K. Burnett, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Radar odometry combining probabilistic estimation and unsupervised feature learning", in *Robotics: Science and Systems*, 2021.

[17] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception in bad weather", in *International Conference on Robotics and Automation*, IEEE, 2021, pp. 1–7.

[18] Z. Hong, Y. Petillot, and S. Wang, "Radarslam: Radar based large-scale SLAM in all weathers", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2020, pp. 5164–5170.

[19] H. Porav, T. Bruls, and P. Newman, "I can see clearly now: Image restoration via de-raining", in *International Conference on Robotics and Automation*, IEEE, 2019, pp. 7087–7093.

[20] B. Yamauchi, "Fusing ultra-wideband radar and lidar for small UGV navigation in all-weather conditions", in *Unmanned Systems Technology XII*, 2010.

[21] Y. S. Park, J. Kim, and A. Kim, "Radar localization and mapping for indoor disaster environments via multi-modal registration to prior lidar map", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2019, pp. 1307–1314.

[22] M. Maimone, Y. Cheng, and L. Matthies, "Two years of visual odometry on the Mars exploration rovers", *Journal of Field Robotics*, vol. 24, no. 3, pp. 169–186, 2007.

[23] S. Chadwick, W. Maddern, and P. Newman, "Distant vehicle detection using radar and vision", in *International Conference on Robotics and Automation*, IEEE, 2019, pp. 8311–8317.

[24] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry", in *Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2004.

[25] F. Fraundorfer and D. Scaramuzza, "Visual odometry Part II: Matching, robustness, optimization, and applications", *Robotics & Automation Magazine*, vol. 19, no. 2, pp. 78–90, 2012.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[27] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection", in *European Conference on Computer Vision*, Springer, 2006, pp. 430–443.

[28] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)", *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[29] C. G. Harris and J. Pike, "3D positional integration from image sequences", *Image and Vision Computing*, vol. 6, no. 2, pp. 87–90, 1988.

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

[31] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates", in *International Conference on Robotics and Automation*, IEEE, 2006, pp. 2262–2269.

[32] P. Newman, D. Cole, and K. Ho, "Outdoor SLAM using visual appearance and laser ranging", in *International Conference on Robotics and Automation*, IEEE, 2006, pp. 1180–1187.

[33] S. Wang, R. Clark, H. Wen, and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks", in *International Conference on Robotics and Automation*, IEEE, 2017, pp. 2043–2050.

[34] R. Li, S. Wang, Z. Long, and D. Gu, "UndeepVO: Monocular visual odometry through unsupervised deep learning", in *International Conference on Robotics and Automation*, IEEE, 2018, pp. 7286–7291.

[35] V. Peretroukhin and J. Kelly, "DPC-net: Deep pose correction for visual localization", *Robotics and Automation Letters*, vol. 3, no. 3, pp. 2424–2431, 2017.

[36] S. Anderson and T. D. Barfoot, "RANSAC for motion-distorted 3D visual sensors", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2013, pp. 2093–2099.

[37] T. D. Barfoot, C. McManus, S. Anderson, H. Dong, E. Beerepoot, C. H. Tong, P. Furgale, J. D. Gammell, and J. Enright, "Into darkness: Visual navigation based on a lidar-intensity-image pipeline", in *Robotics research*, Springer, 2016, pp. 487–504.

[38] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time", in *Robotics: Science and Systems*, vol. 2, 2014.

[39] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes", in *Sensor fusion IV: control paradigms and data structures*, International Society for Optics and Photonics, vol. 1611, 1992, pp. 586–606.

[40] S. Hong, H. Ko, and J. Kim, "VICP: Velocity updating iterative closest point algorithm", in *International Conference on Robotics and Automation*, IEEE, 2010, pp. 1893–1898.

[41] Q. Li, S. Chen, C. Wang, X. Li, C. Wen, M. Cheng, and J. Li, "Lo-net: Deep real-time lidar odometry", in *Conference on Computer Vision and Pattern Recognition*, IEEE/CVF, 2019, pp. 8473–8482.

[42] Y. Cho, G. Kim, and A. Kim, "Unsupervised geometry-aware deep lidar odometry", in *International Conference on Robotics and Automation*, IEEE, 2020, pp. 2145–2152.

[43] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast", in *International Conference on Robotics and Automation*, IEEE, 2015, pp. 2174–2181.

[44] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite", in *Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 3354–3361.

[45] E. J. M. Adams J. Mullane and B. Vo, *Robot Navigation and Mapping with Radar*. Artech House, 2012.

[46] M. I. Skolnik, "Introduction to radar", *Radar handbook*, vol. 2, 1962.

[47] K. Burnett, A. P. Schoellig, and T. D. Barfoot, "Do we need to compensate for motion distortion and doppler effects in spinning radar navigation?", *Robotics and Automation Letters*, vol. 6, no. 2, pp. 771–778, 2021.

[48] R. J. James, "A history of radar", *IEE Review*, vol. 35, no. 9, pp. 343–349, 1989.

[49] Weapons and Systems Engineering Deptartment, United States Naval Academy. (1999). Chapter 15 guidance and control, [Online]. Available: `https://fas.org/man/dod-101/navy/docs/fun/part15.htm` (visited on 07/19/2021).

[50] J. Arlin Cooper and L. Thomas James, "Location estimation and guidance using radar imagery", *Transactions on Aerospace and Electronic Systems*, no. 6, pp. 1383–1388, 1975.

[51] R. G. Caves, P. J. Harley, and S. Quegan, "Matching map features to synthetic aperture radar (SAR) images using template matching", *Transactions on Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 680–685, 1992.

[52] R. Y. Wong, "Sequential scene matching using edge features", *Transactions on Aerospace and Electronic Systems*, no. 1, pp. 128–140, 1978.

[53] L. Novak, "Correlation algorithms for radar map matching", *Transactions on Aerospace and Electronic Systems*, no. 4, pp. 641–648, 1978.

[54] T. Y. Tang, D. De Martini, D. Barnes, and P. Newman, "RSL-net: Localising in satellite images from a radar on the ground", *Robotics and Automation Letters*, vol. 5, no. 2, pp. 1087–1094, 2020.

[55] T. Y. Tang, D. De Martini, S. Wu, and P. Newman, "Self-supervised learning for using overhead imagery as maps in outdoor range sensor localization", *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1488–1509, 2021.

[56] H. F. Durrant-Whyte, "An autonomous guided vehicle for cargo handling applications", *The International Journal of Robotics Research*, vol. 15, no. 5, pp. 407–440, 1996.

[57] S. Clark and H. Durrant-Whyte, "Autonomous land vehicle navigation using millimeter wave radar", in *International Conference on Robotics and Automation*, IEEE, vol. 4, 1998, pp. 3697–3702.

[58] M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem", *Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001.

[59] S. Clark and G. Dissanayake, "Simultaneous localisation and map building using millimetre wave radar to extract natural features", in *International Conference on Robotics and Automation*, IEEE, vol. 2, 1999, pp. 1316–1321.

[60] E. Jose and M. D. Adams, "An augmented state SLAM formulation for multiple line-of-sight features with millimetre wave radar", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2005, pp. 3087–3092.

[61] J. Callmer, D. Törnqvist, F. Gustafsson, H. Svensson, and P. Carlbom, "Radar SLAM using visual features", *EURASIP Journal on Advances in Signal Processing*, 2011.

[62] J. W. Marck, A. Mohamoud, E. vd Houwen, and R. van Heijster, "Indoor radar SLAM a radar application for vision and GPS denied environments", in *European Radar Conference*, IEEE, 2013, pp. 471–474.

[63] F. Schuster, C. G. Keller, M. Rapp, M. Haueis, and C. Curio, "Landmark based radar SLAM using graph optimization", in *International Conference on Intelligent Transportation Systems*, IEEE, 2016, pp. 2559–2564.

[64] F. Lu and E. Milios, "Robot pose estimation in unknown environments by matching 2D range scans", *Journal of Intelligent and Robotic systems*, vol. 18, no. 3, pp. 249–275, 1997.

[65] F. Schuster, M. Wörner, C. G. Keller, M. Haueis, and C. Curio, "Robust localization based on radar signal clustering", in *Intelligent Vehicles Symposium*, IEEE, 2016, pp. 839–844.

[66] E. Ward and J. Folkesson, "Vehicle localization with low cost radar sensors", in *Intelligent Vehicles Symposium*, IEEE, 2016, pp. 864–870.

[67] M. Chandran and P. Newman, "Motion estimation from map quality with millimeter wave radar", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2006, pp. 808–813.

[68] D. Vivet, P. Checchin, and R. Chapuis, "Radar-only localization and mapping for ground vehicle at high speed and for riverside boat", in *International Conference on Robotics and Automation*, IEEE, 2012, pp. 2618–2624.

[69] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration", *Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, 1996.

[70] F. Gérossier, P. Checchin, C. Blanc, R. Chapuis, and L. Trassoudaine, "Trajectory-oriented EKF-SLAM using the Fourier-Mellin transform applied to microwave radar images", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2009, pp. 4925–4930.

[71] P. Checchin, F. Gérossier, C. Blanc, R. Chapuis, and L. Trassoudaine, "Radar scan matching SLAM using the Fourier-Mellin transform", in *Field and Service Robotics*, Springer, 2010, pp. 151–161.

[72] R. Goecke, A. Asthana, N. Pettersson, and L. Petersson, "Visual vehicle egomotion estimation using the Fourier-Mellin transform", in *Intelligent Vehicles Symposium*, IEEE, 2007, pp. 450–455.

[73] D. Barnes, R. Weston, and I. Posner, "Masking by moving: Learning distraction-free radar odometry from pose information", in *Conference on Robot Learning*, PMLR, 2020, pp. 303–316.

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

[75] ——, "Instantaneous ego-motion estimation using multiple doppler radars", in *International Conference on Robotics and Automation*, IEEE, 2014, pp. 1592–1597.

[76] M. Cornick, J. Koechling, B. Stanley, and B. Zhang, "Localizing ground penetrating radar: A step toward robust autonomous ground vehicle localization", *Journal of Field Robotics*, vol. 33, no. 1, pp. 82–102, 2016.

[77] D. De Martini, M. Gadd, and P. Newman, "kRadar++: Coarse-to-fine FMCW scanning radar localisation", *Sensors*, vol. 20, no. 21, p. 6002, 2020.

[78] Ş. Săftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, "Kidnapped radar: Topological radar localisation using rotationally-invariant metric learning", in *International Conference on Robotics and Automation*, IEEE, 2020, pp. 4358–4364.

[79] M. Gadd, D. De Martini, and P. Newman, "Look around you: Sequence-based radar place recognition with learned rotational invariance", in *Position, Location and Navigation Symposium*, IEEE/ION, 2020, pp. 270–276.

[80] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, "Mulran: Multimodal range dataset for urban place recognition", in *International Conference on Robotics and Automation*, IEEE, 2020, pp. 6246–6253.

[81] A. Foessel, S. Chheda, and D. Apostolopoulos, "Short-range millimeter-wave radar perception in a polar environment", 1999.

[82] G. Reina, J. Underwood, G. Brooker, and H. Durrant-Whyte, "Radar-based perception for autonomous outdoor vehicles", *Journal of Field Robotics*, vol. 28, no. 6, pp. 894–913, 2011.

[83] P. Kaul, D. De Martini, M. Gadd, and P. Newman, "RSS-Net: Weakly-supervised multi-class semantic segmentation with FMCW radar", in *Intelligent Vehicles Symposium*, IEEE, 2020, pp. 431–436.

[84] R. Weston, S. Cen, P. Newman, and I. Posner, "Probably unknown: Deep inverse sensor modelling radar", in *International Conference on Robotics and Automation*, IEEE, 2019, pp. 5446–5452.

[85] 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.

[86] D. Williams, D. De Martini, M. Gadd, L. Marchegiani, and P. Newman, "Keep off the grass: Permissible driving routes from radar with weak audio supervision", in *International Conference on Intelligent Transportation Systems*, IEEE, 2020, pp. 1–6.

[87] I. Cvišić, J. Ćesić, I. Marković, and I. Petrović, "SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles", *Journal of Field Robotics*, vol. 35, no. 4, pp. 578–595, 2018.

[88] S. H. Cen and P. Newman, "Radar-only ego-motion estimation in difficult settings via graph matching", in *International Conference on Robotics and Automation*, IEEE, 2019, pp. 298–304.

[89] D. Barnes and I. Posner, "Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar", in *International Conference on Robotics and Automation*, IEEE, 2020, pp. 9484–9490.

[90] Y. S. Park, Y. Shin, and A. Kim, "PhaRaO: Direct radar odometry using phase correlation", in *International Conference on Robotics and Automation*, IEEE, 2020, pp. 2617–2623.

[91] D. Adolfsson, M. Magnusson, A. Alhashimi, A. J. Lilienthal, and H. Andreasson, "CFEAR radarodometry — conservative filtering for efficient and accurate radar odometry", in *International Conference on Intelligent Robots and Systems*, IEEE/RSJ, 2021, pp. 5462–5469.

[92] P. C. Kung, C. C. Wang, and W. C. Lin, "A normal distribution transform-based radar odometry designed for scanning and automotive radars", in *International Conference on Robotics and Automation*, IEEE, 2021, pp. 14 417–14 423.

[93] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The Oxford radar robotcar dataset: A radar extension to the Oxford robotcar dataset", in *International Conference on Robotics and Automation*, IEEE, 2020, pp. 6433–6438.

[94] M. A. Khalighi and M. H. Bastani, "Adaptive CFAR processor for nonhomogeneous environments", *Transactions on Aerospace and Electronic Systems*, vol. 36, no. 3, pp. 889–897, 2000.

[95] P. P. Gandhi and S. A. Kassam, "Optimality of the cell averaging CFAR detector", *Transactions on Information Theory*, vol. 40, no. 4, pp. 1226–1228, 1994.

[96] B. Magaz, A. Belouchrani, *et al.*, "Automatic threshold selection in OS-CFAR radar detection using information theoretic criteria", *Progress In Electromagnetics Research*, vol. 30, pp. 157–175, 2011.

[97] A. Alhashimi, D. Adolfsson, M. Magnusson, H. Andreasson, and A. Lilienthal, "BFAR — bounded false alarm rate detector for improved radar odometry estimation", *arXiv preprint arXiv:2109.09669*, 2021.

[98] D. Barnes, "Leveraging domain knowledge for self-supervision in scalable robot learning", PhD thesis, University of Oxford, 2020.

[99] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar", in *Radar Conference*, 2017, pp. 1703–1708.

[100] S. Chen and H. Wang, "SAR target recognition based on deep learning", in *International Conference on Data Science and Advanced Analytics*, IEEE, 2014, pp. 541–547.

[101] H. Wang, S. Chen, F. Xu, and Y. Q. Jin, "Application of deep-learning algorithms to MSTAR data", in *International Geoscience and Remote Sensing Symposium*, IEEE, 2015, pp. 3743–3745.

[102] B. Xue and N. Tong, "DIOD: Fast and efficient weakly semi-supervised deep complex ISAR object detection", *Transactions on Cybernetics*, no. 99, pp. 1–13, 2018.

[103] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints", in *International Conference on Computer Vision*, IEEE Computer Society, vol. 2, 2005, pp. 1482–1489.

[104] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF", in *International conference on computer vision*, IEEE, 2011, pp. 2564–2571.

[105] L. S. Shapiro and J. M. Brady, "Feature-based correspondence: An eigenvector approach", *Image and vision computing*, vol. 10, no. 5, pp. 283–288, 1992.

[106] G. L. Scott and H. C. Longuet-Higgins, "Feature grouping by "relocalisation" of eigenvectors of the proximity matrix.", in *BMVC*, Citeseer, 1990, pp. 1–6.

[107] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.

[108] J. H. Challis, "A procedure for determining rigid body transformation parameters", *Journal of Biomechanics*, vol. 28, no. 6, pp. 733–737, 1995.

[109] S Cen, "Ego-motion estimation and localization with millimeter-wave scanning radar", Master's thesis, University of Oxford, 2019.

[110] W. Churchill, "Experience based navigation: Theory, practice and implementation", PhD thesis, University of Oxford, 2012.

[111] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The Oxford robotcar dataset", *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[112] J. Illingworth and J. Kittler, "A survey of the Hough transform", *Computer vision, graphics, and image processing*, vol. 44, no. 1, pp. 87–116, 1988.

[113] J. Skingley and A. Rye, "The Hough transform applied to SAR images for thin line detection", *Pattern Recognition Letters*, vol. 6, no. 1, pp. 61–67, 1987.

[114] I. A. Barsan, S. Wang, A. Pokrovsky, and R. Urtasun, "Learning to localize using a lidar intensity map", in *Conference on Robot Learning*, PMLR, 2018, pp. 605–616.

[115] S. Hoermann, P. Henzler, M. Bach, and K. Dietmayer, "Object detection on dynamic occupancy grid maps using deep learning and automatic label generation", in *Intelligent Vehicles Symposium*, IEEE, 2018, pp. 826–833.

[116] D. Barnes, W. Maddern, G. Pascoe, and I. Posner, "Driven to distraction: Self-supervised distractor learning for robust monocular visual odometry in urban environments", in *International Conference on Robotics and Automation*, IEEE, 2018, pp. 1894–1900.

[117] T. Bruls, W. Maddern, A. A. Morye, and P. Newman, "Mark yourself: Road marking segmentation via weakly-supervised annotations from multimodal data", in *International Conference on Robotics and Automation*, IEEE, 2018, pp. 1863–1870.

[118] M. Sheeny, A. Wallace, and S. Wang, "Radio: Parameterized generative radar data augmentation for small datasets", *Applied Sciences*, vol. 10, no. 11, p. 3861, 2020.

[119] A. D. Stewart, "Localisation using the appearance of prior structure", PhD thesis, University of Oxford, 2014.

[120] J. Guehring, "Reliable 3D surface acquisition, registration and validation using statistical error models", in *International Conference on 3-D Digital Imaging and Modeling*, IEEE, 2001, pp. 224–231.

[121] S. Nobili, G. Tinchev, and M. Fallon, "Predicting alignment risk to prevent localization failure", in *International Conference on Robotics and Automation*, IEEE, 2018, pp. 1003–1010.

[122] B Brumback and M Srinath, "A chi-square test for fault-detection in Kalman filters", *Transactions on Automatic Control*, vol. 32, no. 6, pp. 552–554, 1987.

[123] R. Mehra, "On the identification of variances and adaptive Kalman filtering", *Transactions on Automatic Control*, vol. 15, no. 2, pp. 175–184, 1970.

[124] G. Welch, G. Bishop, *et al.*, "An introduction to the Kalman filter", 1995.

[125] S. Sarkar and K. L. Boyer, "Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors", *Computer Vision and Image Understanding*, vol. 71, no. 1, pp. 110–136, 1998.

[126] G. K. Tam, Z. Q. Cheng, Y. K. Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X. F. Sun, and P. L. Rosin, "Registration of 3D point clouds and meshes: A survey from rigid to nonrigid", *Transactions on Visualization and Computer Graphics*, vol. 19, no. 7, pp. 1199–1217, 2012.

[127] J. Almeida and V. M. Santos, "Real time egomotion of a nonholonomic vehicle using lidar measurements", *Journal of Field Robotics*, vol. 30, no. 1, pp. 129–141, 2013.

[128] D. Scaramuzza, F. Fraundorfer, and R. Siegwart, "Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC", in *International Conference on Robotics and Automation*, IEEE, 2009, pp. 4293–4299.

Abbreviations

CFAR Constant False Alarm Rate. 34, 35, 74

FMCW Frequency-Modulated Continuous-Wave. 1, 20, 21

FMT Fourier-Mellin Transform. 31, 32

GPR Ground-penetrating Radar. 31

ICP Iterative Closest Point. 19, 30, 74

LiDAR Light Detection And Ranging. 5, 7, 8, 10, 11, 13, 14, 18, 19, 23, 30, 31, 32, 58, 59, 62, 74, 86, 101

MMW millimetre-wave. 1, 10, 20, 23, 59

RANSAC Random Sample Consensus. 17

RCS Radar Cross Section. 20, 30, 37

RO Radar Odometry. 2, 3, 5, 7, 32, 33, 34, 39, 40, 41, 42, 43, 44, 45, 55, 56, 57, 58, 61, 62, 63, 76, 78, 79, 78, 79, 80, 82, 83, 85, 86, 99, 100, 101, 103, 104, 106, 107, 108

SAR Synthetic Aperture Radar. 29

SFM structure from motion. 16

SLAM Simultaneous Localisation and Mapping. 11, 30

SVD singular value decomposition. 17, 40

ToF Time-of-Flight. 2, 23

VO Visual Odometry. 13, 16, 17, 18, 19, 32, 41, 54, 55, 56, 86, 101, 107