Ego-Motion Estimation and Localization with Millimeter-Wave Scanning Radar

Thesis



Sarah Huiyi Cen

MSc by Research

Worcester College

University of Oxford

Supervisor: Paul Newman

March 4, 2019

Sarah Huiyi Cen
Worcester College

MSc by Research
Hilary 2019

Ego-Motion Estimation and Localization with Millimeter-Wave Scanning Radar

Abstract

In contrast to cameras, lidars, GPS, and proprioceptive sensors, radars are affordable and efficient systems that operate well under variable weather and lighting conditions, require no external infrastructure, and detect long-range objects. For these reasons, radar is a reliable and versatile sensor well suited for ego-motion estimation and localization. However, various radar artifacts and characteristics—including speckle noise, receiver saturation, multipath reflections, low spatial resolution, and slow update speeds—make it an especially challenging sensor with which to work.

In this thesis, we present algorithms for radar-only odometry and localization for mobile autonomous systems. We focus primarily on odometry and show that our approach can be adapted for localization. Using a frequency-modulated continuous-wave (FMCW) scanning radar, we first extract keypoints from radar scans in a manner that avoids redundant returns, accounts for high levels of noise, and reduces reliance on the amplitude of received power as an indicator of occupancy. To estimate relative motion from the keypoint sets, we then perform scan matching by greedily adding point correspondences based on unary descriptors and pairwise compatibility scores in a graph matching framework.

Our pipeline achieves reliable and accurate ego-motion estimation under challenging conditions and across a range of settings. The pipeline of our most recent paper requires only one input parameter and no outlier detection, no model-reliant motion filters, nor any additional sensors. Our system operates well even in the absence of a motion prior and is highly robust to radar artifacts. We demonstrate its ability to match the performance of visual odometry and GPS/INS, remaining precise and robust even under difficult conditions that cause the latter two systems to fail. Additionally, it adapts effortlessly to diverse environments, from urban UK to off-road Iceland, achieving an average scan matching accuracy of 5.20 cm and 0.0929 degrees when using GPS as ground truth.

Statement of Authorship

This thesis is submitted to the Department of Engineering Science, University of Oxford, in fulfillment 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.

Sarah Huiyi Cen, Worcester College

Funding

The work described in this thesis was funded by the Daniel M. Sachs Class of 1960 Graduating Scholarship.

Acknowledgements

I would first like to thank my advisor, Paul Newman, for his endless support and guidance, and my examiners, José Neira and Ingmar Posner, for their assistance and feedback. Thank you to everyone at the Oxford Robotics Institute and Worcester College for being so welcoming and kind. I am incredibly grateful to all—friends, colleagues, teammates, administrators, faculty—who made my time at Oxford unbelievably enjoyable and transformative.

As always, an inexpressible thank you to my parents and brother for being constant sources of strength and love.

Contents

1 The case for radar **1**

1.1 Motivation 1

1.2 Contributions 3

1.3 Thesis roadmap 6

2 Overview of scanning radar **7**

2.1 Introduction to radar 7

2.1.1 Fundamental principles 8

2.1.2 Basic design 10

2.1.3 The radar equation 15

2.2 FMCW scanning radar 17

2.2.1 Operating principle 19

2.2.2 Output 22

2.2.3 Capabilities 25

2.2.4 Sources of error 27

2.2.5 The Navtech CTS350-X radar 31

2.3 Summary 32

3 Literature review **34**

3.1 Uses of radar 34

3.1.1 Early applications 34

3.1.2 Radar for mobile robotics 38

3.2 Related work on odometry and localization 41

3.2.1 Keypoint extraction 41

3.2.2 Data association 45

3.2.3 Direct methods 51

3.3 Summary 54

4 Radar-only odometry: robustness under variable conditions **55**

4.1 Paper (S. H. Cen and P. Newman, ICRA 2018) 55

4.2 Discussion 64

4.2.1 Summary of contributions 64

4.2.2 Design choices 65

4.2.3 Next steps 71

5 Radar-only odometry: versatility and reliability across settings **73**

5.1 Paper (S. H. Cen and P. Newman, ICRA 2019) 73

5.2 Discussion 81

5.2.1 Summary of contributions 81

5.2.2 Design choices 82

5.3 Illustrative results 87

6 Discussion **93**

6.1 Keypoint extraction 93

6.1.1 Problem statement 93
6.1.2 Considerations and challenges 94
6.1.3 Analyzing our approach 95
6.2 Data association 99
6.2.1 Problem statement 99
6.2.2 Considerations and challenges 100
6.2.3 Analyzing our approach 102
6.3 System robustness 105
6.4 Understanding radar odometry 107
6.5 Localization 109
6.6 Summary 111

7 Conclusion and future work **112**

Chapter 1

The case for radar

1.1 Motivation

In mobile robotics, an autonomous agent's primary objective is to accurately, efficiently, and safely travel through its environment by properly sensing, understanding, and judging its surroundings. In order to do so, it must overcome several challenges. In particular, an autonomous agent must achieve robust localization and navigation despite changing conditions (e.g., lighting and weather) and moving objects (e.g., pedestrians and other vehicles).

Currently, most platforms employ lidar, vision, GPS, proprioceptive sensors, or a combination of these systems to observe and interpret their surroundings. However, while very fast and high-resolution, lidars are sensitive to weather conditions, especially rain and fog, and only register the first surface encountered. Vision systems are versatile and cheap but easily impaired by scene changes, like poor lighting or the sudden presence of snow. Both optical sensors only yield dependable results for short-range measurements. A typical GPS on its own guar-



Figure 1.1: Images showing assorted conditions and settings: outdoors and overcast (top left), outdoors and rainy (top right), outdoors and dark (bottom left), and indoors and crowded (bottom right). In these environments, many common sensor systems, including cameras and GPS, fail while radar remains consistent.

antees at best 1-2 m accuracy and frequently experiences reception difficulties near obstructions due to its reliance on an external infrastructure. Additionally, proprioceptive sensors, like wheel encoders and IMUs, suffer from significant drift among other detrimental effects.

In contrast, **radar** is a long-range, on-board system that performs well under variable lighting and atmospheric conditions both indoors and outdoors, making it highly reliable and consistent. Compared to other popular rangefinders, radar is rapidly becoming more affordable, compact, and data efficient. In addition to being practical, radar features multifaceted sensing capabilities. Due to its beam spread and long wavelength, radar receives multiple readings from a single

transmission (unlike lidar, which only returns the first detection per transmission). As a result, radar provides an information-rich grid representation of its world. The returns can encode various details about the radar's environment, including the distance, speed, orientation, and material of each object it detects.

For these reasons, radar is a promising sensor for numerous tasks in mobile robotics, from mapping and path planning to object detection and tracking. Because it detects stable, long-range features in the environment, radar is particularly well suited for odometry and localization, which are the focuses of this work. In this thesis, we seek to explore radar's capabilities via a radar-only system. We do so not to advocate for a uni-modal sensor system but to demonstrate the strengths of radar by featuring what it can achieve on its own. We show that this low-cost, data-inexpensive sensor matches the performance of state-of-the-art systems under more diverse, and often adverse, conditions.

1.2 Contributions

While radar offers many advantages over its alternatives, working with it is non-trivial. It has lower resolution and updates measurements less frequently than lidar and vision sensors. Its returns frequently contain "ghost" objects (i.e., false detections) and other unwanted artifacts. It is also affected by high levels of noise and interference. As a result, radar is an especially challenging sensor. However, for the same reasons, a pipeline that achieves high performance using only a radar in spite of its drawbacks may generalize well to other applications and sensing modalities that face the same or milder effects (e.g., sonar for underwater robotics

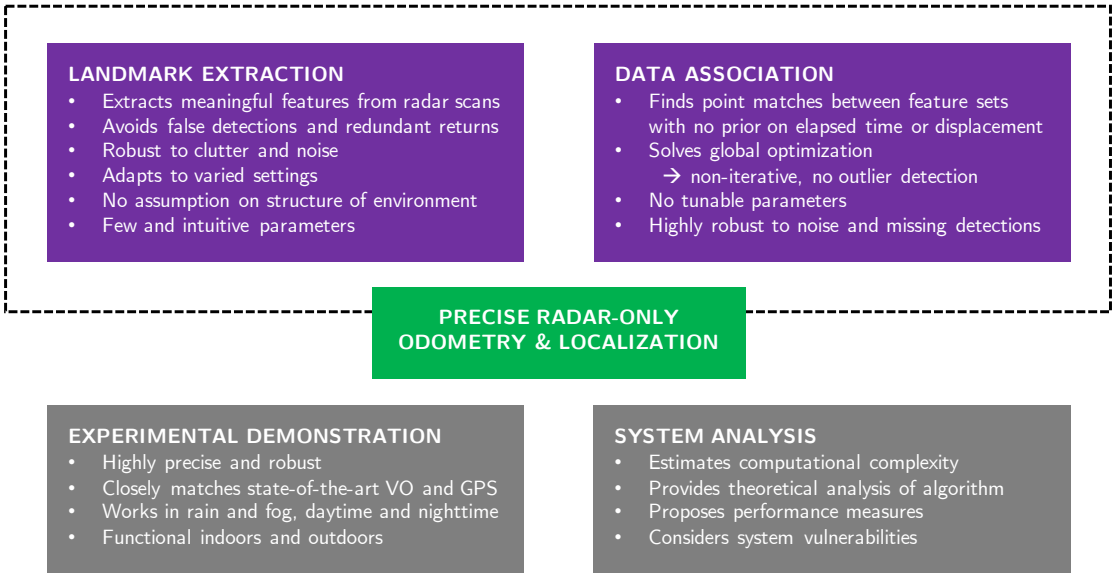


Figure 1.2: Summary of our contributions. The two algorithms in purple constitute the core of our radar-only odometry and localization system (green), which undergoes in-depth testing and analysis (grey) in this thesis.

as well as CT, MRI, and ultrasound systems for medical imaging [1]). Thus, our contributions are the application of radar to odometry and localization for mobile robotics with potential extensions to similar wave-based applications, as follows:

1. We demonstrate precise odometry and localization using only a millimeter-wave (MMW) frequency-modulated continuous-wave (FMCW) scanning radar. Two main advantages of our system are that it is: (1) robust to adverse conditions, and (2) designed to adapt to different environments with minimal prior knowledge and, accordingly, human intervention. It does not rely on other sensors, map creation, model-reliant filters, or outlier detection.

2. We present an efficient keypoint extraction method that reliably identifies meaningful features in the raw radar scan while avoiding false detections and minimizing redundant returns. The algorithm is designed to be robust to clutter and high levels of noise, and it performs well across varied settings,

requiring few and intuitive tunable parameters. Our discussion presents a useful analysis of the information encoded in radar scans.

3. We present a geometrically-based, non-iterative data association algorithm that provides point correspondences between two keypoint sets. Our method is capable of matching sets that are displaced by any translation or rotation as long as they observe a sufficient number of common keypoints. This property means that we require no prior on the temporal or spatial relationship between sets. Although it matches individual keypoints, our algorithm solves a global optimization problem and is, accordingly, undeterred by high levels of measurement noise and frequent missing detections. We show that our system requires no parameter tuning or outlier detection.

4. We experimentally verify that our system produces precise radar-only odometry and localization under circumstances that cause other popular sensors to fail, like changing environments (e.g., urban, indoor, natural) or undergoing variable lighting and weather conditions (e.g., dark, sunny, rainy). We show that the radar closely matches the performance of vision systems and GPS even under conditions that are ideal for the latter two sensors.

5. We provide an in-depth analysis of our system, including computational complexity, vulnerabilities, and performance metrics. We discuss ways in which the system can be improved and extended to other tasks as well as important considerations for the use of scanning radars in mobile robotics.

1.3 Thesis roadmap

The remainder of this thesis is organized as follows.

Chapter 2 introduces important concepts and principles for understanding our sensor, the FMCW scanning radar. It discusses the operation, capabilities, and sources of error of the sensor.

Chapter 3 summarizes the related work on odometry and localization, drawing from literature on radar, lidar, and vision systems. It focuses on indirect methods (e.g., keypoint extraction followed by data association) and direct methods (e.g., raw scan matching or end-to-end deep neural networks). It then explains how our work contributes to the existing literature.

Chapter 4 presents our publication on precise radar-only ego-motion estimation under diverse and challenging conditions. It includes details on our proposed algorithms as well as experimental results.

Chapter 5 presents our latest publication on radar-only ego-motion estimation across diverse environments. As in Chapter 4, it includes algorithmic details as well as results in addition to an in-depth system analysis.

Chapter 6 provides a discussion on our system's advantages and flaws as well as important considerations for radar odometry.

Chapter 7 summarizes our contributions and suggests future work.

Chapter 2

Overview of scanning radar

2.1 Introduction to radar

Radar falls under the class of sensors known as *active remote sensing rangefinders*, which are devices that (1) determine a target's range by actively transmitting a signal in its direction and detecting the echo that is reflected from it; and (2) measure the distance, or *range*, between the sensor and a target object without coming into physical contact with the object [2]. Radar can be distinguished from similar systems, such as lidar [3] and sonar [4], by the nature of the transmitted signals, which are radio, laser, and sound waves, respectively. Compared to other rangefinders, radar stands out due to its abilities to measure distances at long ranges with relatively high accuracy, observe its surroundings even under variable lighting and weather conditions, and capture both the locations and velocities of objects. In addition, radar is becoming increasingly affordable, and it lacks the environmental or health hazards accompanying other sensor systems.

Historically, the primary uses of radar have been target detection and tracking,

Table 2.1: Definitions and classifications of relevant sensing terminology

Term	Definition
Remote sensor	Acquires data without physical contact
Active	Influences the environment in order to gather data, and, in doing so, requires an external energy source
Monostatic	Uses same antenna for transmitting and receiving
Bistatic	Transmits and receives from separate locations
Passive	Obtains measurements by responding to inputs from the environment and thus relies on natural energy sources
Rangefinder	Determines the relative distance of target objects
Electromagnetic	Uses electromagnetic waves as the energy source
Radar	Emits radio waves
Lidar	Emits laser waves
Mechanical	Uses mechanical waves (e.g., sound) as the energy source
Scanning	Surveys the environment by sequentially changing its field-of-view, thus obtaining spatial information from the accumulation of many measurements over time
Imaging	Collects measurements into a spatial array containing observations that are captured simultaneously; observes environment within a specified field-of-view

meteorological analysis, and terrain mapping for military and industrial applications [5]. In mobile autonomy, radar is generally used for obstacle avoidance (e.g., during lane changing). Doppler radars can also measure the relative speed of approaching objects, like other vehicles or pedestrians. In our work, a scanning radar is mounted on a moving vehicle and used to survey the entire environment, including buildings, cars, and humans as well as the terrain and vegetation.

2.1.1 Fundamental principles

Radar is an electromagnetic sensor that determines the location and reflective properties of objects in its surroundings [2]. As shown in Figure 2.1, it does so by radiating electromagnetic energy in the form of radio waves outwards. When the radar's transmitted wave encounters an object in the environment, it is reflected

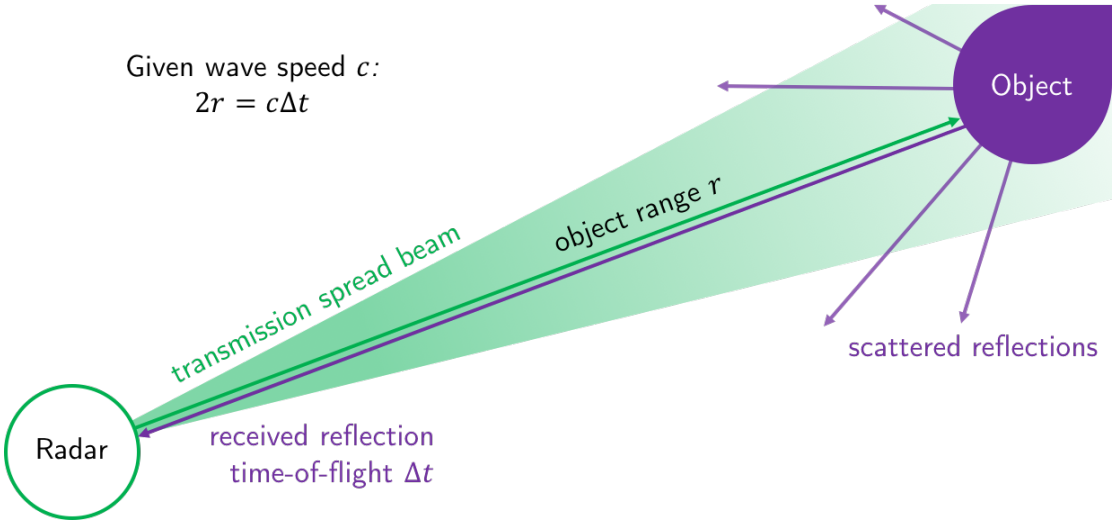


Figure 2.1: Fundamental operating principle of radar. The transmission, for which the power density decreases with distance, has a spread beam visualized by the green cone. When the radio wave transmissions (green arrow) travelling at speed of light c encounter an object at range r, they scatter (purple arrows). Some reflected waves, or echoes, return to the radar after time-of-flight Δt.

away, and some fraction of the reflected waves, or *echoes*, return to the receiver.

The received signal can be compared to the transmitted signal to compute the range of the object that reflected the wave (i.e., the *reflector*). Since electromagnetic waves travel at the speed of light c in a vacuum, a reflector's relative distance r can be computed as a function of the wave's *time-of-flight* Δt, the length of time that the transmitted signal took to travel the distance to the reflector and back to the radar $2r$. This relationship, $r = \frac{c\Delta t}{2}$, is the operating principle of *pulse radars* [2]. The received signal—which has a different shape than the transmission due to multiple effects, including pulse spreading—encodes additional characteristics of the reflector, such as its size, material, and orientation. As described in Section 2.2.1, a radar can also be designed to detect an object's relative speed.

All electromagnetic waves are specified by *wavelength* λ, which is inversely related to *frequency* $f = c/\lambda$, where c is the speed of light. Radio waves have wavelengths 10^8 to 10^{-4} m and frequencies 3 GHz to 3 THz [2]. Along this portion

of the electromagnetic spectrum, waves are not scattered by small atmospheric particles (e.g., precipitation), which allows radar to operate under variable weather conditions, although precipitation might cause atmospheric attenuation.

The specific transmission frequency band is selected based on the radar's intended purpose. While low-frequency radars are better suited for long-range applications, high-frequency radars can locate objects with greater accuracy (i.e., have higher range resolution). Currently, radars mounted on vehicles are primarily used for short-range applications (e.g., collision avoidance). However, long-range radar is gaining traction in mobile autonomy as an alternative to lidar and vision sensors for localization and mapping. For these applications, the range-accuracy trade-off becomes an important consideration. In general, greater bandwidth improves range accuracy, signal wavelength determines the types of materials observable by radar, and antenna design affects beamwidth, gain, and sensitivity.

2.1.2 Basic design

Radar contains four main modules: transmitter, receiver, antenna, and duplexer. An example block diagram for a monostatic radar is given in Figure 2.2.

Transmitter

The *transmitter*, which is normally a voltage-controlled oscillator and power amplifier, generates the waveform that the radar radiates (i.e., transmits) outwards. The signal is specified by its average and peak powers as well as by its waveform. The choice of waveform (e.g., pulse or continuous, fixed or frequency-modulated) depends on the desired application. For instance, only certain waveforms can be

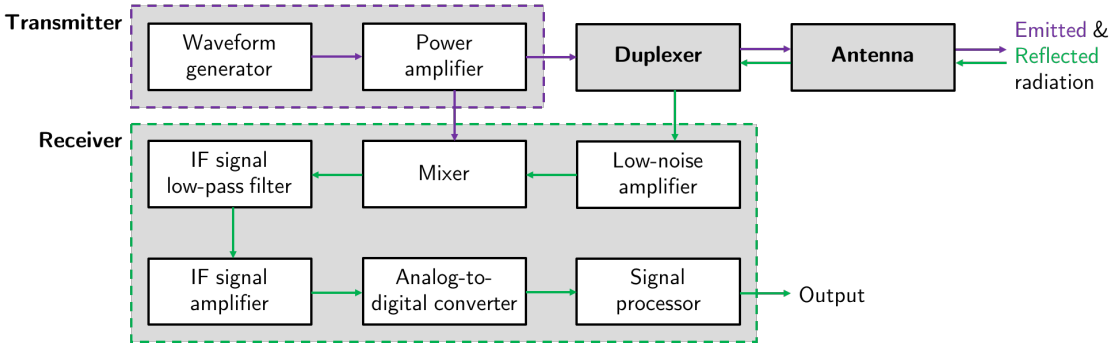
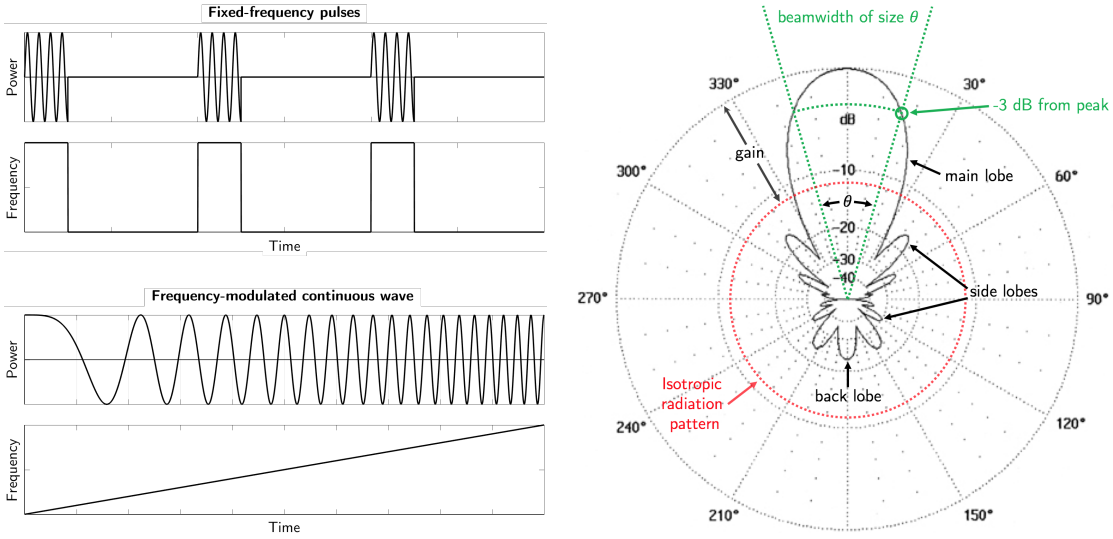


Figure 2.2: Basic block diagram for a monostatic (i.e., single-antenna) radar. The antenna interfaces with the environment by emitting and receiving radiation. The duplexer allows a single antenna to both transmit and receive. The transmitter consists of a waveform generator (e.g., voltage controlled oscillator) and amplifier. The receiver includes two amplifiers, a low-pass filter, and an analog-to-digital converter. It also contains the mixer, which compares the transmitted and received waves to extract the IF signal. The range, velocity, and other relevant information of a target are extracted by the signal processor.

used to solve for the speed of a reflector. Two example waveforms are shown in Figure 2.3a. Other important considerations for the transmitter are its stability, operating frequency, and bandwidth. The latter two determine several key radar characteristics, including the radar's wave propagation properties (e.g., penetration, attenuation, or absorption), range resolution, maximum range, and measurement update speed, as explained in Section 2.1.3.

Receiver

The *receiver* must be both sensitive enough to detect targets with low echo signals yet robust enough to avoid high levels of noise and distortion. Although high sensitivity ensures that the receiver observes the desired signal, it also increases the likelihood of false detections due to clutter. In the context of mobile autonomy, it is essential that the radar not return false negatives when detecting important objects, like humans, cyclists, vehicles, and obstacles. On the other hand, hypersensitivity can result in clutter (e.g., vegetation) distracting from true detections



(a) (b)

Figure 2.3: (a) Examples of waveforms that can be used for transmission: on top, fixed-frequency pulses and, on bottom, a frequency-modulated continuous wave. Each is shown by its power-time and frequency-time plots. (b) An example of a logarithmic polar plot (modified from [6]) showing the transmission radiation pattern for a highly directional antenna. The shape outlined by the solid black line shows the transmission power in the labeled directions for an antenna located at the origin. This plot only shows the azimuth-plane slice that contains the peak power point of a full 3D radiation pattern that is not necessarily symmetric. The main, back, and side lobes of this slice are indicated, as is the main beam of beamwidth θ. As shown, the gain can be found by comparing the transmitted pattern's peak power to that of an isotropic pattern of the same total power.

or even causing malfunctions (e.g., saturation).

This trade-off is determined by the receiver's dynamic range, defined as the ratio of the maximum to minimum signal powers over which the receiver operates properly. A large dynamic range guarantees high sensitivity (i.e., detections of signals produced by true reflectors), but a small one minimizes distortion (i.e., nonlinear effects, like saturation due to strong reflections). Another important consideration is the receiver bandwidth. The receiver experiences both internal and external sources of noise, which are undesirable and can be mitigated by narrowing the bandwidth. However, doing so can reduce the radar's range resolution, maximum range and velocity, signal-to-noise ratio, and robustness to interference.

Antenna

In a monostatic radar, the *antenna* interfaces with both the transmitter and receiver, which is desirable in applications requiring a compact radar. The antenna directs the transmission into space while also directing the incoming echoes to the receiver. It typically radiates the transmission in a highly directional, narrow beam such that the radar observes a specific angular region at a time, and the width of this beam determines the radar's angular resolution. In mobile autonomy, higher angular resolution is desirable because it provides more fine-grained information about the environment.

The amount of radiation that is transmitted in the desired direction is specified by the antenna's gain (explained in Section 2.1.3) and efficiency. The shape of the transmission is termed the radiation pattern, for which an example is shown in Figure 2.3b. It consists of the *main beam*, which contains the majority of the radiated energy, and *sidelobes*, which are leaked energy of lower intensity in unintended directions. The more narrow the main beam, the more accurately radar can locate its targets. Since all targets that are detected are believed to lie in the main beam, sidelobes introduce ambiguity by allowing for reflectors outside the main beam to be detected, resulting in false positives.

The transmission abilities (e.g., gain) are also related to the antenna's *effective aperture* or area. In particular, a higher effective aperture improves the antenna's ability to receive power and capture weak echos, but it is limited by size constraints. Thus, in general, a monostatic antenna's geometry depends on its dual transmitting-receiving roles.

Table 2.2: Radar design parameters and key considerations

Design parameter	Considerations
Transmission frequency	Wave propagation properties
	Atmospheric absorption and attenuation
	Range resolution
	Maximum range
	Antenna size
Transmission pattern	Pulse vs continuous-wave
	Frequency modulation
	Bandwidth
	Doppler capabilities
	Measurement update speed
	Peak and average powers
	Stability
Receiver dynamic range	Sensitivity (i.e., maximum range)
	Distortion (e.g., saturation)
Receiver bandwidth	Range resolution
	Maximum range and velocity
	Signal-to-noise ratio
	High-frequency receiver noise
	Low-frequency noise floor
	Interference
Antenna aperture	Sensitivity to reflections
	Size limitations
Antenna radiation	Efficiency and gain
	Beamwidth and directionality
	Radiation pattern and sidelobes

Duplexer

The *duplexer* allows for the use of a single antenna by keeping the powerful transmissions and weak incoming signals isolated. It protects the sensitive receiver from high-energy transmissions while relaying the low-energy echoes to the receiver.

Design parameters

In sum, there are several key radar design parameters: transmission waveform (e.g., frequency, bandwidth, and power, which affect the range accuracy and max-

imum), receiver properties (e.g., dynamic range and bandwidth, which affect the detection sensitivity as well as robustness to noise and distortion), and antenna characteristics (e.g., radiation pattern and effective aperture, which affect gain and angular resolution). These considerations are summarized in Table 2.2.

2.1.3 The radar equation

The signals received by radar can be analyzed to extract information about the radar's surroundings. This step is termed signal or data processing. Solving for an object's range relies on the *radar equation*, which defines the relationship between the radar, reflector, and environment using known physics [2].

Derivation

Consider a radar that radiates transmissions isotropically (i.e., in all directions equally) with power P_{trans}. Then, the power density of the transmission at range r away from the radar is given by:

$$\hat{D}_r^{\text{trans}} = \frac{P_{\text{trans}}}{4\pi r^2}.$$

Real antennas, however, are not isotropic and instead focus the transmission into a beam. The power density D_r^{trans} within that beam at range r is greater than the isotropic power density $\hat{D}_r^{\text{trans}}$ by a multiplication factor, called the gain G, such that: $D_r^{\text{trans}} = \hat{D}_r^{\text{trans}} G$.

If an object lies within that beam at range r, the amount of power it reflects $P_{\text{reflected}}$ is related to both the power density it intercepts and the reflective prop-

erties of the portion of the object exposed to the beam. The latter is termed the *radar cross section* (RCS) and represented by σ. The RCS is a function of various factors, including the area of the target within the transmission beam, reflecting surface's orientation relative to the beam, surface's texture, and target's material properties. Because radar operates under non-ideal conditions, losses occur due to internal radar characteristics and properties of the medium through which the signals travel, among other factors. These effects are incorporated using a loss term L. As a result, the power reflected by the target is:

$$P_{\text{reflected}} = \frac{\hat{D}_r^{\text{trans}} G\sigma}{\sqrt{L}} = \frac{P_{\text{trans}} G\sigma}{4\pi r^2 \sqrt{L}}, \tag{2.1}$$

where the loss factor is $\sqrt{L}$ for reasons that will be apparent shortly. Recall that the receiver gain of a monostatic radar is the same as that of the transmitter. In this context, the target becomes the transmitter and, by the same logic used above, the power radiated by the target and detected by the radar receiver is:

$$P_{\text{received}} = \frac{\hat{D}_r^{\text{reflected}} A_e}{\sqrt{L}} = \frac{P_{\text{reflected}} A_e}{4\pi r^2 \sqrt{L}}, \tag{2.2}$$

where A_e is the effective aperture of the antenna. Therefore, the total power reflected from the target back to the receiver can be found by combining equations (2.1) and (2.2) to produce the radar equation, as follows:

$$\boxed{P_{\text{target}} = \frac{P_{\text{trans}} G\sigma A_e}{16\pi^2 r^4 L},} \tag{2.3}$$

where P_{target} is the power received by the antenna due to a target at range r with

RCS function σ. If desired, the equation can be reformulated in terms of the transmission wavelength λ from the relationship $G = \frac{4\pi A_e}{\lambda^2}$, such that:

$$\boxed{P_{\text{target}} = \frac{P_{\text{trans}} G \sigma A_e^2}{4\pi \lambda^2 r^4 L}.} \tag{2.4}$$

These equations can be used to compute the range of reflectors in the environment and select the radar's design parameters according to the desired application.

Maximum range

From equation (2.3), it is possible to solve for the maximum range that the radar can observe based on the antenna's minimal detectable power P_{min}:

$$r_{\text{max}} = \left[\frac{P_{\text{trans}} G \sigma A_e}{16\pi^2 P_{\text{min}} L} \right]^{1/4}. \tag{2.5}$$

2.2 FMCW scanning radar

In this thesis, a *frequency-modulated continuous-wave* (FMCW) *scanning* radar is employed for all experiments. This type of sensor is defined by several features. As a scanning radar, it can survey 360 degrees around its platform. It mechanically rotates about its vertical axis such that each narrow, highly directional transmission beam inspects one angular slice at a time, as shown in Figure 2.4. The array of measurements obtained during a single transmission (represented by a green dotted line in Figure 2.4) are associated with the azimuthal angle around which the angular slice is centered. The radar collects information at every azimuth in the desired field-of-view to create a 2D array of measurements, each entry corre-

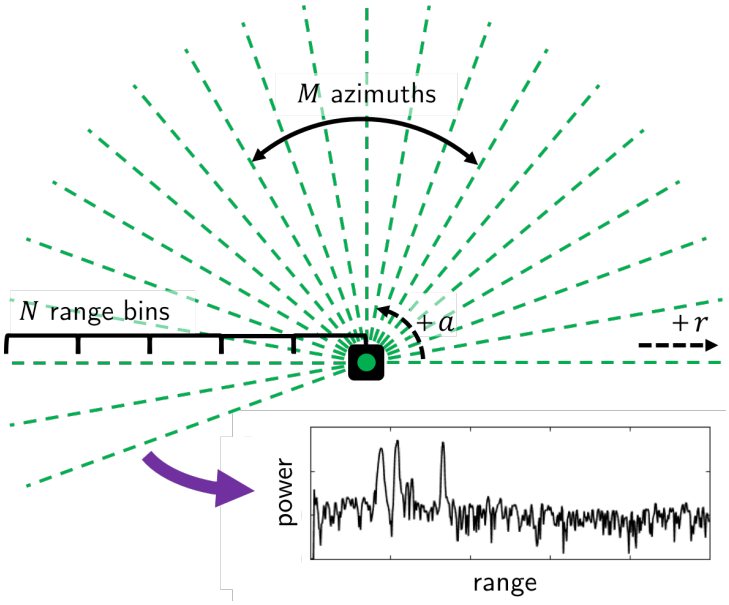


Figure 2.4: Visualization of the FMCW scanning radar viewed from above. The radar (green circle), centered on a vehicle (black box), sequentially gathers power-range spectra (dotted green rays) at each azimuth. A sample signal is plotted. Variables a and r denote azimuth and range, respectively.

sponding to a range and azimuth. As a result, although the radar's transmissions are 3D, as explained in Figure 2.3b, and the information is thus captured in 3D, radar outputs a 2D representation of its environment. In our case, the radar is mounted on a ground vehicle such that its axis of rotation is perpendicular to the ground, and it outputs a polar scan akin to a birds-eye view of the vehicle's surroundings, as further explained in Section 2.2.2.

FMCW radar's transmissions are continuous waves (rather than pulses) and, instead of operating at a single frequency, span a band of frequencies such that the waveform is frequency-modulated [7]. For each azimuthal slice, the output data encodes the ranges and, in some cases, velocities of the reflectors. This type of radar can obtain high range accuracy, measure at close ranges, and achieve good power efficiency, among other capabilities that are described in Section 2.2.3 though it suffers from several sources of error that are listed in Sectionx 2.2.4.

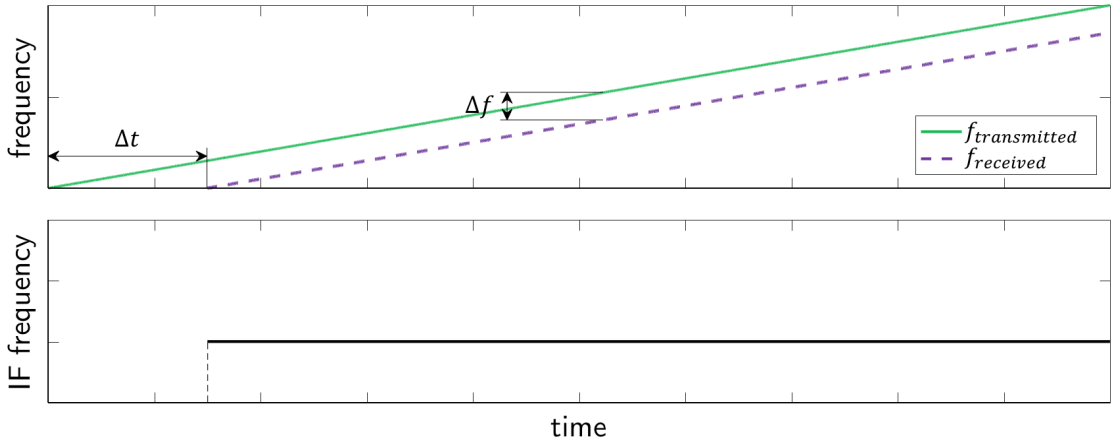


Figure 2.5: The transmitted signal (green) is a simple linearly modulated upchirp. If the signal reflects off of a single target that is static relative to the radar, the ideal echo (purple) retains the modulation pattern shape, displaced by some time-of-flight Δt, which relates to the frequency difference Δf by the slope of $f_{\text{transmitted}}$. The frequency of the IF signal is Δf and can be used to solve for the target's range.

2.2.1 Operating principle

Unlike a pulse radar, an FMCW radar deduces a target's range and, if desired, velocity from the frequency (rather than time) difference between the transmitted and received signals. As a result, the defining design parameter of an FMCW radar is its *modulation pattern*, which is the frequency f of the transmitted signal as a function of time t. Figures 2.5 and 2.6 show two common modulation patterns.

As described in Section 2.1.1, when a high-power transmitted signal encounters a target in the environment, it is reflected away, and a low-power echo returns to the receiver. In principle, the echo should retain the same shape as the transmitted modulation pattern (though many factors can cause distortion, such as those described in Section 2.2.4), except for a shift in time and/or frequency, as shown in Figure 2.5 for the simple upchirp modulation pattern. All information encoded by the FMCW radar can be extracted from this shift, as explained next.

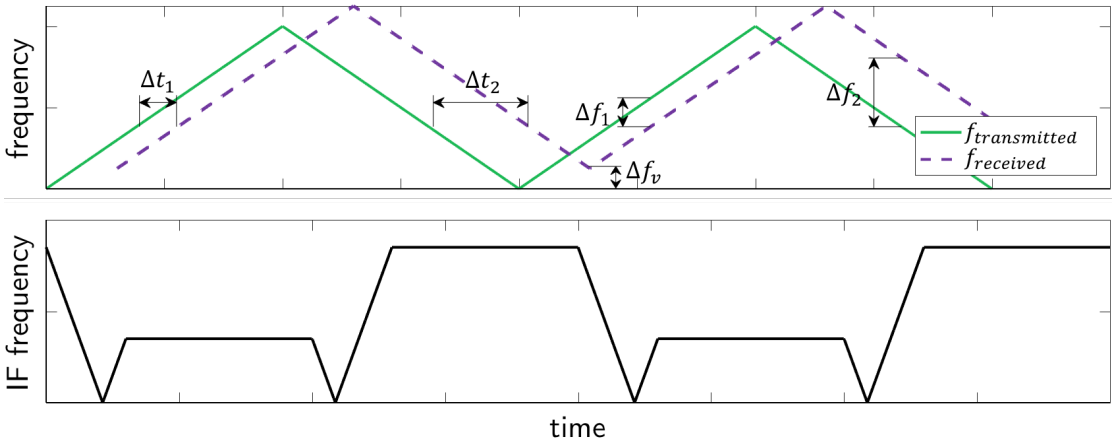


Figure 2.6: The transmitted signal (green) follows the triangular modulation pattern. If the signal reflects off of a single target with relative velocity v, the ideal echo (purple) retains the modulation pattern shape but is shifted in both time and frequency. As shown in the plot of the IF signal frequency, the frequency difference corresponding to the up- and down-chirp portions of the pattern are different. The use of both allows for the shifts due to the target's range and velocity to be resolved according to equations (2.7)-(2.8).

Computing range and velocity

Assuming that a target is static relative to the radar, its range r can be computed from the difference between the transmitted and received signals, which is known as the *intermediate frequency* (IF) signal, as follows:

$$r = \frac{c\Delta f}{2(df/dt)}, \tag{2.6}$$

where c is the speed of light, $\Delta f = f_{\text{transmitted}} - f_{\text{received}}$ is the frequency difference, and df/dt is the rate of change of frequency during the transmission chirp. This scenario is illustrated in Figure 2.5 for a linear modulation pattern.

Now, the assumption of the reflectors being static is removed. If a target is moving relative to the radar along the ray connecting the radar to the target, the reflected signal will undergo an additional frequency shift due to the *Doppler effect* [2]. As a result, the entire echo shifts not only laterally but also vertically

in the frequency-time plot, as shown in Figure 2.6 for the triangular modulation pattern. The frequency differences Δf_1 and Δf_2 in Figure 2.6 thus capture both range and velocity information. To decouple the frequency differences due to range Δf_r and velocity Δf_v, the following relationships can be used:

$$\Delta f_r = \frac{\Delta f_1 + \Delta f_2}{2} \tag{2.7}$$

$$\Delta f_v = \frac{\Delta f_2 - \Delta f_1}{2}. \tag{2.8}$$

Note that, according to the Doppler effect, a positive Δf_v means that the object's relative motion is towards the radar, and vice versa. The object's range r can be computed from Δf_r using equation (2.6), and its velocity v from Δf_v. The equation relating v to Δf_v and its derivation can be found in [2].

Not all modulation patterns allow for the differentiation between the frequency shifts due to range and velocity. Unless the pattern contains repeating segments, it is impossible to distinguish the Doppler from the time-of-flight shift. As a result, if there is relative motion between the radar and some object, the perceived range of the object is inaccurate, displaced by the amount caused by its motion. While repeating patterns allow the two frequency shifts to be resolved, they also introduce range ambiguity in that there is no longer a one-to-one correspondence between a given Δf_r and r. Moreover, adding repeating patterns either narrows the transmission bandwidth (which reduces the maximum range and range resolution) or increases the measurement update time (which slows the update speed). Whether to use Doppler capabilities depends on the desired application and the relative importance of velocity information to range ambiguity and resolution.

Implementation

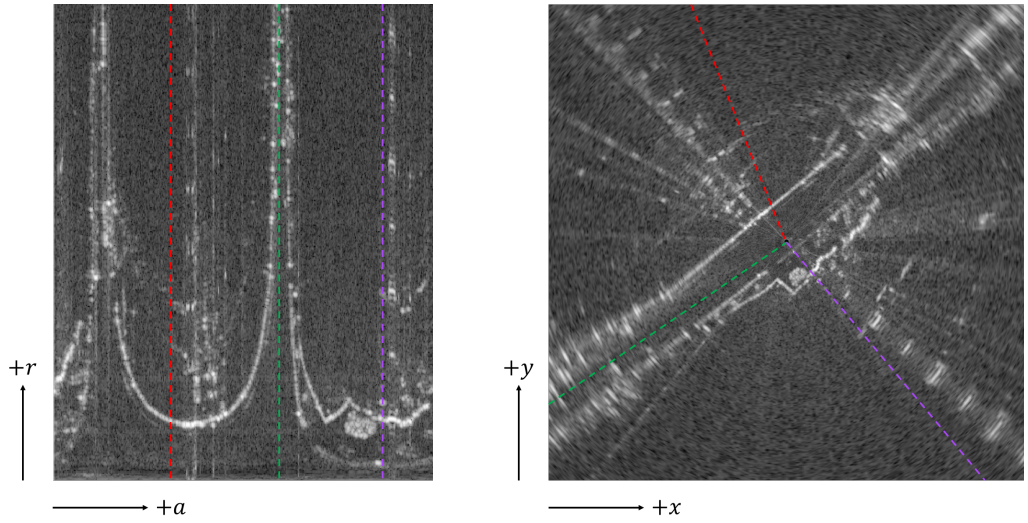
In practice, the transmitted and received signals are compared using a device called the mixer, which computes their difference. The result, which is known as the intermediate frequency (IF) signal, then undergoes a *fast Fourier transform* (FFT). The FFT outputs the frequency spectrum of the IF signal, and each entry of the frequency spectrum is the difference between the instantaneous frequencies of the transmitted and received signals. As such, each entry corresponds to a range, and the corresponding magnitude indicates the received power. This signal is termed the *power-range spectrum* in this thesis, but it is also known in relevant literature as the range-FFT. Although not reviewed here, the phase of the IF signal can also be used to determine a target's range and movements.

2.2.2 Output

An FMCW radar is capable of encoding several pieces of information in its output. As explained in Section 2.2.1, it obtains the ranges and velocities of surrounding objects via the frequency difference between the transmitted and received waves. It also captures each reflector's RCS characteristics, such as its material, size, and orientation relative to the radar, through the received power.

Format

The output of a scanning radar is a grid, in which each *cell* corresponds to a region in space. The size of each region is determined by the radar's angular and range resolutions. As the antenna rotates, it focuses each transmission into a narrow beam such that the radar inspects one azimuthal slice at a time. For



(a) (b)

Figure 2.7: (a) Example polar scan, which can be visualized in Cartesian space, as shown in (b). The radar is located at $r = 0$, or the bottom in (a), and at the origin, or center in (b). Pixel values scale with received powers such that brighter pixels indicate higher power, and vice versa. The axes a and r in (a) represent the measurement azimuth and range. The vertical lines in (a) correspond to the same-colored azimuthal lines in (b). The Cartesian image (b) is obtained by bilinear interpolation of (a). Note that (b) is less clear farther from the radar. In (b), the well-defined, continuous lines generally correspond to building walls. The circular object close to the vehicle and to the left of the purple line is a tree. The series of short lines to the right of the purple line and extending towards the bottom-right of the scan are false positives caused by receiver saturation. Another noticeable effect is the high level of speckle noise throughout the image.

each azimuth, the FMCW radar returns a power-range spectrum, which is a vector containing the received power readings for a series of range bins. Once the radar undergoes a full rotation across all azimuths, the spectra are combined into a *scan*, a two-dimensional matrix containing the power data such that each row corresponds to an azimuth and each column to a range. Some radars only gather measurements for a subset of azimuths and thus output partial scans.

Figure 2.7 visualizes an output scan as a grayscale image. It shows the raw output, which represents the world in polar space, as well as the radar image converted to Cartesian space. Note that the Cartesian spatial resolution decreases

with increasing range, straight lines in Cartesian space are not straight in polar space, and a rigid body changes shape in polar space depending on its location. The scan shown in Figure 2.7 is captured by mounting the radar on a vehicle such that its axis of rotation is perpendicular to the ground. As such, the scan in 2.7(b) is akin to a birds-eye view image centered at the vehicle's location.

Resolution

As range is computed directly from the frequency spectrum of the IF signal, range resolution is determined by the smallest measureable frequency difference Δf_{min}. In other words, two echoes can only be resolved in the frequency spectrum if their frequency difference is observable, such that:

$$\Delta f_{\text{obs}} > 1/T_{\text{chirp}}, \tag{2.9}$$

where T_{chirp} is the duration of the chirp (a.k.a. sweep) in the modulation pattern and Δf_{obs} is an observable frequency difference. From equation 2.6, note that:

$$\Delta f = \frac{2\Delta r(df/dt)}{c}.$$

Combined with equation 2.9, an observable difference in range Δr_{obs} must satisfy:

$$\Delta r_{\text{obs}} > \frac{c}{2T_{\text{chirp}}(df/dt)} = \frac{c}{2B}, \tag{2.10}$$

where $B = T_{\text{chirp}}(df/dt)$ is the transmission bandwidth. Therefore, the range resolution, $\Delta r_{\text{min}} = c/(2B)$, improves with greater transmission bandwidth.

As discussed in Section 2.1.2, the angular resolution is determined by the azimuthal beamwidth. A narrow beam provides greater angular resolution but observes a smaller region in space, meaning that the receiver or transmitter must compensate by being more sensitive or increasing power, respectively.

Maximum range

The maximum unambiguous range of an FMCW radar is determined by the bandwidth of the IF signal B_{IF} and the duration T_{chirp} of the transmission chirp:

$$r_{\max} = \frac{cB_{\text{IF}}}{2(df/dt)} = \frac{cT_{\text{chirp}}}{2}. \tag{2.11}$$

Recall that the maximum range is also restricted by the minimum power that the receiver can detect, as specified in equation (2.5).

2.2.3 Capabilities

The following list enumerates the advantages of the FMCW scanning radar:

- In comparison to pulse radars, FMCW scanning radar features:
  - Higher range resolution with simpler components due to processing the IF signal in frequency, rather than temporal, space.
  - The ability to detect targets at close range.
  - Lower peak power, which makes it energy efficient, low-cost, and safe to operate nearby humans and wildlife.
  - Quicker measurement updates due to continuous transmission.
- In comparison alternative non-radar sensors, the benefits of radar are the:

Table 2.3: Comparison of capabilities of popular sensors for mobile autonomy

Sensor	Type	Accuracy/ Resolution	Limits	Rate (s^{-1})	Power (W)
Navtech CTS350 [8]	FMCW scanning radar	Range: 0.25 m Azimuth: <1°	Min range: 1 m Max range: 500 m FOV: 360°	4	20
Bosch LRR4 [9]	Monostatic multimodal[1] FMCW radar	Range: 0.12 m Azimuth: 0.2° Velocity: 0.11 m/s	Min range: 0.36 m Max range: 250 m FOV: 12-40°	20	4.5
Delphi ESR [10]	Multimodal[1] electronically scanning radar	Range: 0.5 m Azimuth: 0.5° Velocity: 0.12 m/s	Min range: 1 m Max range: ∼150 m FOV: 20°-90°	20	12
Velodyne HDL-64E [11]	High-definition lidar scanner	Range: 0.02 m Azimuth: 0.2°	Min range: ∼0 m Max range: 120 m FOV: 360°	5-20	60
FLIR Bumblebee2 [12]	Stereo vision camera	1032 × 776 pixels	FOV: 97°	20	2.5
NovAtel SPAN-CPT [13]	GNSS + INS[2]	Position: ∼0.8 m Heading: 0.2-0.4°	Requires access to satellite signals	20	16

[1] Multimodal means that the radar returns Doppler information (i.e., speed of detected objects).
[2] GNSS denotes Global Navigation Satellite System and INS denotes Inertial Navigation System.

  - Long-range propagation capabilities of radio waves.
  - Resilience under variable lighting and temperature.
  - Robustness to diverse weather and atmospheric conditions, including fog, rain, snow, and dust due to its wavelength.

- The radar's IF signal simultaneously encodes the both range and velocity.
- The radar's information-rich output captures 3D data in a 1D signal and records multiple readings per transmission.
- Increasingly simplified processing requirements of FMCW scanning radars allow hardware components to be more compact and consistent compared to pulse radars and less costly than lidar.

Tables 2.3 and 2.4 compare the capabilities of various sensors.

Table 2.4: Qualitative comparison of FMCW radar capabilities to other individual sensor systems for mobile autonomy

Attribute	FMCW[1]	Pulse[1]	lidar	camera	GPS
Data per update	1D signal	1D signal	point(cloud)	image	location
Object velocity[2]	yes[3]	maybe	maybe	no	no
Fog[4]	yes	yes	no	no	yes
Rain[4]	yes	yes	no	no	yes
Poor lighting[4]	yes	yes	yes	no	yes
Indoors[4]	yes	yes	yes	yes	no
Noise levels	high	high	low	low	-
Ghost objects	yes	yes	no	no	-
Multiple returns	yes	yes	no	no	-

[1] Denoting the type of radar
[2] Whether the sensor can measure velocities of other objects
[3] Assuming the modulation pattern has repeating segments.
[4] Indicating the sensor system's ability to operate under this condition.

2.2.4 Sources of error

A sensor's drawbacks are as important to note as its capabilities. Compared to alternative systems, those of the FMCW scanning radar include the:

- Lower spatial resolution than lidar.
- Slower measurement update speeds than lidar.
- Higher levels of noise than lidar and vision systems.
- False detections (a.k.a. "ghost" objects).
- Disadvantages of being an active sensor (e.g., it is easier to detect and more susceptible to interference).
- Higher cost and lower range than traditional pulse radars.

The main sources of error, some of which contribute to the drawbacks listed above, are described below and enumerated in Tables 2.5 and 2.6:

- Hardware (e.g., receiver) and external (e.g., thermal) noise.

Table 2.5: The sources of error for FMCW scanning radar

Error	Source
Range error	Relative height of object
	Collapsing 3D information into 1D output
	Complex target (e.g., many surfaces, moving components)
	Doppler shift (if unable to resolve the frequency due to range and velocity)
Angular error	Location of object within spread beam
	Sidelobes
	Multipath reflections
	Beam redirection
False positive	Clutter
	Sidelobes
	Multipath reflections
	Receiver saturation
False negative	Atmospheric attenuation
	Clutter/object interference
	Highly reflective object at forefront
	Absorption of radio waves by target
	Target smaller than operating wavelength
	Rough surface causes diffuse scattering

- Speckle noise due to interference, which occurs when the IF signal is composed of multiple reflected waves.
- Clutter, which are echoes from unimportant reflectors in the environment and often cause interference.
- Target noise, which can be due to a target's complex shape, changing RCS, and/or moving parts.
- Sidelobes, which are radiation transmitted in unintended directions.
- Multipath reflections, which are waves that encounter multiple objects before returning to the receiver.
- Receiver saturation, which are due to echoes with powers that exceed the

Table 2.6: The types of noise for FMCW scanning radar

Noise	Source
Internal noise	Shot noise
	Thermal noise
	Phase noise
	Flicker noise
	Jitter
	Quantization noise
External noise	Atmospheric effects
	Speckle noise
	Interference
	Thermal & cosmic radiation

receiver's upper dynamic range.

- Signal attenuation and beam redirection due to properties of the medium of wave propagation (e.g., atmospheric attenuation, wave diffraction).
- Phase noise and jitter due to frequency instability in the transmitter.
- Flattening of the information obtained across a 3D spread beam into a 1D signal, in which target width and height manifest as range.

These internal and external sources of errors primarily affect the output in four ways: noise in the power readings, noise in the detected position, false positives, and false negatives, as summarized in Table 2.5 and 2.6.

These unwanted effects mean that radar scans are often difficult to interpret, which has made them less popular for mobile autonomy than lidar, cameras, and GPS. However, radar has some intrinsic characteristics that make it well suited for ensuring two crucial factors in mobile autonomy: reliability and safety. In particular, radar is long-range, lighting invariant, and robust to most precipitation. It also rarely returns false negatives, which is important for collision avoidance

Table 2.7: Attributes and consequences of FMCW scanning radar

Attribute	Compared to	Consequence(s)[1]
Radio waves	Vision	− No color or contrast data
	Lidar	− More susceptible to interference
		− Larger beamwidth
Longer wavelength	Lidar	+ Penetrates various materials[2]
		+ Undeterred by small particles (fog, dust, rain, snow, etc.)
		− Lower bandwidth and resolution
		− Slower update speed
Frequency-modulated	Unmodulated frequency	+ Encodes spatial & velocity data[3]
		+ Fourier, not time-of-flight, analysis (simpler components, cheaper, lighter)
Larger bandwidth	Pulse radar	+ Higher signal-to-noise ratio
		+ Higher range resolution
		+ Increased target detectability
Continuous-wave	Pulse radar and lidar[4]	+ Lower peak power
		+ Faster update speed
Scanning	Static or tracking radar	+ Full field-of-view with ability to locate targets
Active	Passive sensor	+ No dependence on external infrastructure or energy source
		− Easy to detect and interfere
Remote sensing	Tactile sensor	+ Contained and non-intrusive
Wider spread beam	Lidar	+ Encodes 3D information efficiently
		− Small position errors (unable to resolve location within spread beam)
		− Speckle noise (clutter interference)
		− Lower angular resolution
Outputs spectrum[5]	Lidar	+ Multiple returns per transmission
Lower peak power	Pulse radar and lidar	+ Safer for humans and environment
	Pulse radar	− Shorter range due to attenuation
Longer range	Lidar, vision	+ Detects faraway objects
		+ Better for monitoring object speed
		− Requires sensitive receiver (Susceptible to internal noise)
		− Positional errors due to beam bending
Less precise, slower updates	Lidar	+ Data efficient for post-processing
		+ Simpler, more durable components
		+ Lower cost
		− Lower positional accuracy
Monostatic	Bistatic sensor	+ Less complex, on-board system
		+ Provides spatial information
Directional	Omnidirectional	+ Encodes spatial information
		− Transmission sidelobes
On-board, contained	GPS	+ Functions indoors
		+ Easy to install

[1] Advantages are marked by + and disadvantages by −.
[2] FMCW radar can be placed behind a vehicle's bumper, ensuring that it does not protrude.
[3] Target velocities are derived from the Doppler shift.
[4] Most popular lidar sensors transmit pulses.
[5] FMCW radar outputs a 1D signal (power-range spectrum). Lidar outputs the first return (location of first object encountered).

In this thesis, we demonstrate that the lower spatial resolution, high noise, and other drawbacks of radar do not imply imprecise ego-motion estimation and localization because they are compensated for by the rich and long-range information encoded in radar scans. By acknowledging and accommodating for the negative and positive attributes of FMCW scanning radar, we develop a system that is highly robust to measurement errors. However, it is important to recognize the capabilities of complementary sensors. Table 2.7 summarizes the advantages and disadvantages of FMCW scanning radar and compares it to other sensor systems used for mobile autonomy.

2.2.5 The Navtech CTS350-X radar

In this work, the sensor used is the Navtech CTS350-X [14]. It is a MMW FMCW scanning radar with specifications given in Table 2.8. It scans a full 360-degree swathe every 0.25 seconds, dividing the angular space into 400 azimuths with a 2-degree beamwidth. The range resolution can be set to between 2 and 25 cm, and 2000 range bins are provided. Although it can be adapted to provide Doppler information, it is not currently equipped to do so. As a result, the range measurements may feature small errors due to Doppler shifts from relative motions of the platform or targets.

The Navtech CTS350-X has an operating frequency around 77 GHz with a 0.6 GHz bandwidth. It therefore falls under the class of MMW radars, which have gained popularity due to their ability to use large bandwidths that provide good range accuracy and compact antennas for practical size limitations. MMW radars are also more resilient to hostile countermeasures and do not interfere with



(a) (b)

Figure 2.8: The Navtech CTS350-X radar mounted on an extreme off-roading vehicle in Iceland. A close-up of the radar is shown in (a) while the full platform with the radar at the back is shown in (b).

commonly used low-frequency electronics. On the other hand, they are more close-range than low-frequency radars and experience non-negligible attenuation due to water vapor and oxygen. However, these effects do not have a large impact on our use case. Mobile robotics platforms typically only require knowledge of targets within several dozen or hundred meters, and the attenuation effects at these distances will generally be insignificant.

2.3 Summary

At its simplest, radar is an active sensor that transmits radio waves outwards and captures those that scatter off of surrounding objects in order to determine the objects' ranges and, if desired, velocities. Due to the propagation properties of radio waves, radar is a long-range sensor that functions well under diverse conditions that cause other sensor systems—like vision, lidar, and GPS—to fail.

Table 2.8: Specifications for the Navtech CTS350-X Radar [14]

Specification	Details
Classification	FMCW, millimeter-wave, scanning
Operating frequency	76 - 77 GHz
Update rate	4 Hz (rotations per second)
Sweep bandwidth	600 MHz
Range resolution	25 cm (V2), 17.52 cm (V4)
Number of range bins	2000 (V2), 3768 (V4)
Maximum range	500 m (V2), 660 m (V4)
Azimuth beamwidth	1.8 degrees
Measurement rate	400 azimuths per full rotation
Field of view	360 degrees
Elevation beamwidth	Spread beam
Power requirements	20 Watts at 24VDC
Weight	7 kg
Mounting orientation	Upright
Diameter	25 cm (V2)
Height	39 cm (V2)

The radar that we use, the Navtech CTS350-X, operates in the MMW range of the electromagnetic spectrum. As a scanning radar, it examines one angular slice of the environment at a time, sequentially changing its field-of-view. The power-range readings are collected into an information-rich spatial grid. It uses FMCW transmissions, which allows for the radar to be less expensive, more precise, and faster than pulse radars.

It is well suited for mobile autonomony due to its long-range capabilities, detection reliability, and robustness under varying conditions. Despite the many advantages it offers, radar is a challenging sensor due to its high levels of noise, frequent false detections, and mediocre resolution compared to alternative sensors used for this application domain. In sum, radar stands out as a practical, efficient, safe, and robust sensor, but any systems that rely on radar must accommodate for the unwanted artifacts in its output.

Chapter 3

Literature review

3.1 Uses of radar

Before discussing the related literature on radar for odometry and localization in mobile robotics, this section reviews the common early and modern uses of radar, from its military-focused origins to its incorporation onto intelligent vehicles.

3.1.1 Early applications

Although the core principles and equipment required for radar were available in the late 1800s [15], the first successful demonstration of radar took place in Britain in 1935. It was part of a larger initiative to invest in defense technologies due to the expectation of impending war, and these efforts only intensified during World War II and the Cold War. As a result, the early applications of radar were primarily military. From the ground, radar was used for detecting and tracking aircraft [5]. In the air, it was used for aircraft navigation, collision avoidance, and missile guidance. During the latter half of the twentieth century, radar began

being repurposed for other applications, including weather surveillance, terrain mapping, speed monitoring by police, and astronomy/planetary measurements.

In general, these uses take advantage of radar's ability to accurately estimate the range, direction, and velocity of a large reflective target from afar under conditions that pose difficulties for humans, like detecting objects at long ranges while traveling at high speeds in fog. The remainder of this section explores radar applications relevant to this thesis in more detail.

Target detection and tracking

Tracking radar seeks to detect the position of a large moving object, or target, and follow its trajectory based on its heading and velocity [2]. In order to center the narrow radar beam on the target and keep it within the radar's field-of-view, a tracking radar must also predict the target's likely future position.

These radars typically detect large, highly reflective, moving objects from a location with a clear line-of-sight (e.g., tracking aircraft from the ground). Therefore, if the target lies within the transmission beam, it is easily discernible in the radar output. Single-target tracking radars have narrow beamwidths, and detection is performed using the constant false alarm rate (CFAR) method [2].

If the line-of-sight is obstructed or the scene becomes more complex, then false alarms can occur and degrade tracking performance. In addition, because of the beam narrowness, initially acquiring the target is difficult. Surveying the space for targets can be accomplished by tracking while scanning or using a phased array radar, both of which allow for multi-target tracking. Methods for forming and extrapolating a target's trajectory are reviewed in [2].

Collision avoidance

The fundamental principles behind radar-based collision warning systems are similar to those of tracking radars. The main differences are that a collision-avoidance radar is placed on-board a moving vehicle with a fixed orientation, rather than being statically located with changing orientation. For instance, a vehicle may carry a collision avoidance system that issues warnings based on whether objects are too close to or travelling very quickly towards it. Each radar has a limited, fixed field-of-view and is placed at strategic points around the vehicle. Since collisions can only occur with the nearest approaching object, these radars are only required to detect first returns at close ranges. Radar, known for its reliability and lack of false negatives, has little difficulty accomplishing this task. In fact, although these collision-avoidance systems perform well, they fail to take full advantage of radar's long-range, multiple-return, information-rich capabilities.

Navigation and guidance

Early radar-based navigation was designed for aircrafts and missiles. The two most common navigation objectives were: (1) to travel directly to a destination, and (2) to follow a predefined route.

In the first scenario, the destination is known. Historically, missile guidance systems used powerful broadcasting antennas to illuminate the desired target with radio waves [16]. The receiving antenna on-board the missile could then easily identify the target, which would appear very brightly in the radar scan, and the missile would simply navigate towards it. These types of systems are capable of locating large targets and do not require high precision. They generally assume

that there is a clear line-of-sight to the target as well as free, unimpeded space through which the body (e.g., missile) can travel. Additionally, the use of broadcasting radars means that external infrastructure must be installed beforehand, which also makes it susceptible to hostile countermeasures.

In the second scenario, the pilot seeks to localize relative to a predefined route [17]. One way to accomplish this task is to place highly reflective beacons at known locations [18]. The aircraft can use these beacons to align itself as it travels past them. Another approach uses the radar scans to map the ground terrain during an initial traversal of the route, after which any incoming radar readings during subsequent traversals can be compared to and matched against this map [19–21].

Researchers have proposed several approaches for this matching operation. One method represents the radar data as a collection of edges and quantifies the difference between map and scan by the similarity of the resulting shapes using the Hausdorff distance [5]. It then searches through the possible translations, rotations, and scalings for the combination that minimizes the Hausdorff distance. Other map matching techniques seek to maximize some correlation criteria between the map and scan. They are often inspired by the vision community, substituting images with radar scans. Such correlation metrics, which are designed to accurately reflect the matching performance while remaining computationally efficient, include cross correlation, edge difference, and edge ratio [21]. Instead of matching based on the entire radar scan, it is also possible to match prominent keypoints or recognizable patterns in the radar scan using template matching [20, 22]. Similarly to map matching, template matching seeks to maximize some correlation score, but it only examines small patches (i.e., templates) in the scan. While

map matching is arguably more comprehensive, template matching can be less computationally intensive and more resilient to noise depending on which patches are selected. All in all, these match-based approaches have had limited success due to the high levels of noise and other unwanted artifacts in radar returns. Algorithmically, they require an exhaustive search of the transformation space for the optimal solution.

While these techniques successfully utilize radar to provide navigation and guidance for aircrafts and missiles, they are not suitable for all applications. In particular, many of the systems discussed in this section can tolerate large errors without sacrificing performance, assume no line-of-sight occlusions, require a priori knowledge or infrastructure, and contain computationally expensive operations. These characteristics would make them unfit for, as an example, high-precision, lightweight mobile robotics.

3.1.2 Radar for mobile robotics

The adaptation of the existing methods discussed in Section 3.1.1 to mobile robotics must address key differences between the two domains. Three important considerations for mobile robotic systems are:

1. Tasks that previously relied on human operators to provide fine-tuned adjustments, good priors, or context-dependent guidance should be automated without sacrificing performance.

2. The region observable by radar may change. Specifically, aircraft-tracking and terrain-mapping radar systems generally have an unobstructed line-of-

sight to the target. In contrast, robotic platforms, like ground vehicles, frequently experience occlusions, which cause the same scene to appear vastly different depending on the radar's location. Any mobile robotics system must account for this change in perspective and its effect on the field-of-view, keypoint permanence, etc.

3. The scale of problems in mobile robotics is vastly different. A typical robotic platform travels at a fraction of the speed of an aircraft, meaning that it requires a smaller visible range and its movements are more difficult to detect from consecutive measurements. Most importantly, because mobile robotic systems often interact with many other agents, including humans, in close proximity, they demand a higher level of accuracy than aircraft (e.g., positional errors of several meters are not tolerable).

These factors prevent the methods installed on common radar systems from being straightforwardly transferred to mobile robotics. A successful transition would continue to take advantage of radar's unique set of capabilities while addressing the aforementioned list of challenges. Current applications of radar to mobile autonomy and the relevance of this research to other sensors are discussed next.

Mobile autonomy with radar

Due to advancements that have made radars more compact and inexpensive over the last few decades, radar has become a common, often essential, part of the autonomous vehicle sensor suite. The most widespread use of radar is preventing collisions during lane changes, parking, and other driving scenarios by detecting the ranges and speeds of nearby objects [5]. As a sensor that performs well under

diverse conditions and very rarely returns false negatives, the inclusion of radar helps to achieve the most crucial requirement of autonomous systems: safety. These systems are straightforward applications of radar since they make direct use of the output (target position and velocity) and entail very little post-processing.

While these applications are uncomplicated and therefore reliable, they do not take utilize radar at its full potential. On these systems, the information provided by radar is frequently deemed less valuable than that provided by vision or lidar, obtained over a limited field-of-view, and mostly utilized for short-range warnings even though radar offers information-rich, long-range data with good resolution. As such, the application of radar to more complex tasks in mobile robotics remains an open and exciting research area.

Relevance to other sensing modalities

One of the main advantages of radar in comparison to alternative sensors is that it captures abundant information about its environment under varying conditions, which makes it versatile, reliable, and comprehensive. However, due to the amount of information recorded in the data-efficient representation, dissecting the salient and important features in the radar output from background and clutter can be difficult. As a result, radar is often regarded as less favorable than vision, lidar, or GPS because it is considered noisier.

Therefore, although radar provides benefits from practical and safety perspectives, other popular sensors are less noisy, higher resolution, and/or more straightforward. If a system is resilient to high noise, low resolution, and other unwanted effects that occur with radar, then it is highly robust and may be easily general-

izable to other sensors or applications. As explained in [5], "navigation work on radar is of direct relevance to robotic navigation and mapping concepts which *do not* employ radar". In other words, by studying a challenging sensor, our analysis offers insights into sensing for mobile autonomy that can be helpful for other sensor systems or applications with similar data.

3.2 Related work on odometry and localization

In the following sections, we review relevant work on ego-motion estimation and localization for radars. Sections 3.2.1 and 3.2.2 discuss works that adhere to the pattern of first detecting keypoints (or features) then performing data association. Section 3.2.3 examines alternative methods that forego the intermediate step of extracting keypoints and instead operate directly on the sensor outputs with minimal pre-processing. We restrict most of our analysis to monostatic scanning radars. Other configurations, such as the placement of two transmitters and one receiver in a known pattern to obtain surface shape information without post-processing [23], are out of the scope of this thesis.

3.2.1 Keypoint extraction

The first step of most radar ego-motion estimation approaches is the extraction of important keypoints from the raw radar returns. Methods that are considered successful identify meaningful keypoints that correspond to real objects in the environment while suppressing false detections. One straightforward strategy places fixed beacons at known locations [18]. Designed to be highly reflective, they are

clearly distinguishable from other objects in the radar scan. Because this method requires existing infrastructure, it fails in unknown settings. Most other works shed this requirement by detecting keypoints online.

Radar-based approaches

Approaches devised specifically for the radar acknowledge its irregular and high noise floor. Constant False Alarm Rate (CFAR), a common filtering algorithm, adapts to the variable noise floor along the received range-power signal by thresholding each power value according to the expected noise level, which is computed from neighboring cells, excluding the nearest cells whose values may be too similar to the one under consideration [24, 25]. Another technique infers the appearance of keypoints by estimating the radar's noise characteristics along the range signal and exploiting temporal-spatial continuity between azimuths [26, 27]. Specifically, Jose et al. use an averaging technique to characterize the radar's typical level of noise at each range, from which a threshold is acquired. Then, recognizing that each azimuth signal is highly correlated with the previous one, they predict the incoming signal and combine it with the thresholded version using a variable weight. Marck et al. avoid the filtering process altogether by recording only the range with the greatest power return per azimuth signal, but this simplification discards potentially relevant information [28]. Although these methods yield good results, they are highly susceptible to speckle noise and ghost objects, which we believe occurs because these approaches focus on computing an adaptive threshold along range and thus only address one attribute of radar—its high variable noise floor—rather than other important noise characteristics specific to radar.

Vision-based approaches

Some researchers draw from the literature on image processing, creating amplitude gridmaps that effectively transform the radar scans into grayscale images, where the image axes correspond to the range and azimuth, and the pixel values gives the radar power reading at that location. They then extract common image features, such as SIFT and FAST [29, 30]. Others use the gridmaps to find cohesive areas within the radar scan that are deemed interesting (i.e. represent continuous objects or surfaces). These regions are isolated using Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Maximally Stable Extremal Regions (MSER), and Connected Components, each yielding similar performance [31]. While intriguing in concept, these algorithms demonstrate inconsistent performance, partially due to the geometric operation of the scanning radar. By observing its environment radially, it often returns the surface of an object followed by a high-power, "fuzzy" tail in the radial direction, from which it is difficult to detect the edge of an object.

As expected, the keypoint extraction methods tailored to radar produce more meaningful and robust detections. Although commonly used features (e.g. SURF and FAST) work well with images, they are not effective or reliable with radar gridmaps because image feature detectors are designed to locate areas of high-gradient and high-intensity, otherwise known as corners. However, with radar, regions that satisfy these requirements occur often due speckle noise and multi-path reflections, among other effects. Thus, the application of image features to radar scans often results in false detections. Moreover, an ideal keypoint detector for radar marks all location at which it deduces that a real object exists in order to

provide a comprehensive yet sparse representation of the radar scan. Corner detectors give sparse pointclouds, but the keypoints lack meaningful interpretations. As a result, the radar-specific filters discussed above yield better results.

Deep learning-based approaches

Another method for identifying interesting and meaningful areas in the radar scan uses deep learning. One analog to this objective is image segmentation, for which the scan would be segmented into two classes: keypoints and background. A popular neural net for image segmentation is U-Net, which contains "skip" connections between non-adjacent convolutional layers and effectively returns image masks of the original input size, one for each of the classes [32]. A simpler fully convolutional network that contains fewer skip connections is presented by Long et al. [33]. Both outperform their contemporary methods, demonstrating the expansive capabilities of neural nets given sufficient training data and labels.

Alternatively, neural networks can be used to detect contours or edges. DeepContour is a convolutional neural net (CNN) for contour detection [34]. Its key innovation is the division of inputs into small patches, which are assigned to clusters based on the approximate shapes in the patches, in order to train on a finite set of possible contours. The neural net then classifies each patch by the contour it best represents or as background. The positive-sharing loss function that they adopt accommodates this strategy by penalizing the sum of any misclassifications between shape classes equivalently to the misclassifications of patches as background. DeepEdge also consists of a CNN, but it accepts an image at four different scales and candidate contour pixels that are selected by the Canny edge

detector as inputs, emulates feature extraction by applying center, average and max pooling at each layer, and trains against two evaluation criteria: a binary contour mask and the likelihood of classification as a contour by a human [35]. Holistically-nested edge detection (HED) handles the input at multiple scales, like DeepEdge, but it more closely monitors the network's hidden activity through deep supervision, which backpropagates error by comparing the labels against not only the final output but intermediate layer outputs as well [36]. These networks have their own advantages and disadvantages. DeepContour's success lies in extracting small patches and selecting a finite set of contours. While these choices reduce the complexity of the problem and increase the training data, they also limit the number of possible shapes. DeepEdge incorporates multiple sources of information (e.g. the Canny edge detector predictions and the human classification likelihood) at the risk of introducing complications that hinder its utility. HED harnesses the benefits of CNNs and deep supervision, but the net's prediction of thick edges at some edge locations and no edges at others demonstrates that there is ample fine-tuning by humans in order to select the appropriate image scales that may not generalize to all inputs.

3.2.2 Data association

The next step for ego-motion estimation is data association, which seeks to match keypoints that correspond to the same global location observed by the radar at different times. We note that all the following works compute ego-motion from the relative movement of the keypoints, thereby implicitly assuming that the majority of objects that are observed by radar are static [5].

Feature descriptors

Vision-inspired approaches to data association return pairs of keypoints whose feature descriptors are sufficiently similar [29, 30]. As previously discussed, feature descriptors work well for vision systems due to the complex high-density information in images, but they struggle to produce consistent results with radar, for which the readings are much noisier and less dense. Modifying this idea for radar occupancy grids, Schuster et al. identify and associate keypoints using the binary annular statistic descriptor (BASD) and Hamming distance, respectively [37, 38]. To compute the BASD of a point, the space around it is segmented into annuluses of a predetermined thickness; all neighbors that lie within an annulus are used to calculate its set of statistics, and this information defines the descriptor. Utilizing the BASD with radar is shown to yield superior results compared to its vision-inspired counterparts [38]. However, at the high level, because keypoint matching based on feature descriptors examines each keypoint independently, doing so neglects to consider the advantages of analyzing the radar scan as a whole. Moreover, as the radar returns for an object are highly dependent on the position of the vehicle as well as its orientation to the object, these descriptors are not robust across a variety of perspectives. We maintain that harnessing the long-range information (e.g. the relationships between distant keypoints) captured by the radar—potentially in combination with feature descriptors—improves functionality under diverse conditions.

Multi-sensor fusion

Many approaches address the difficulties of noise by integrating radar with other sensors in order to achieve odometry. These systems accept the ego-motion estimation of proprioceptive sensors, from which the incoming pointcloud is transformed into the global frame. The result is then compared to an existing map that is composed of previously acquired pointclouds. In one example, each incoming point is matched to its nearest neighbor in the map, from which a new rigid body transform is computed, allowing the motion estimate from the proprioceptive sensors to be corrected and the map to be updated [37]. Another applies probabilistic weights to each point match based on distance comparisons between the pointcloud and map, then uses Monte Carlo methods to select the final solution [23]. The data association between the radar pointcloud and map provides an additional ego-motion estimation that is then fused with the original odometry readings. While these methods are promising, none produce satisfactory results using only radar. They are aided by other sensors as their radar-based keypoint extraction and data association alone are insufficiently accurate. To compensate, these approaches demand the integration of model-reliant filters (e.g. Kalman and particle), the creation of maps to average out poorly extracted keypoints, and the regular calibration of sensors relative to one another, all of which introduce unnecessary complications and sources of error.

Feature-based scan matching

On the other hand, scan matching [39], which seeks to align keypoint sets in a manner that minimizes some overall cost function, accomplishes data associa-

tion by considering information captured across the entire radar scan and does so without other sensors. The most widespread scan matching approach, called Iterative Closest Point (ICP), iteratively performs point matching using naive methods, computes the optimal rigid body transformation that satisfies the data association result in the least-squares sense, updates the relevant pointcloud, and repeats until the alignment between the two pointclouds meets some termination condition [28, 40–42]. Researchers tend to modify the naive methods to their satisfaction, employing heuristics that rely on criteria such as the distance between points or direction of the surface's normal vector [43]. Although the iterative nature of ICP provides a level of control over the desired accuracy, it also implies that the algorithm is not highly efficient, especially if accuracy is prioritized.

Chandran and Newman adopt a different strategy, developing a function that quantifies the quality of a map that is created by superimposing multiple radar pointclouds [44]. The quality function indicates the clarity and crispness of the resulting map and depends on a set of unknown motion parameters. An exhaustive search is then performed to obtain the motion parameters that optimize this function [44]. Both works require that the initial scan and map are approximately aligned by assuming the movement between scans is small, which imposes an undesirable constraint on the algorithms and prevents them from being applied to arbitrary inputs. Vivet et al. design an innovative technique well suited for high velocities by utilizing the radar scan distortions, which occur solely due to the radar's motion [25, 45]. Most works typically regard the distortions as a drawback of mobile radar systems and implicitly classify them as acceptable sources of error or noise. However, Vivet et al. attempt to quantify the amount of distortion in

each scan in order to backsolve for velocity [25, 45]. Conceptually sound, its main weaknesses are in the data association step, which can mistakenly attribute real motion between scans to distortions, thus increasing the error at low speeds, and also imposes assumptions that do not accommodate high accelerations.

Related classical matching problems

There are several works that seek to perform data association using geometric constraints. One such approach is known as the constellation model [46]. The constellation model addresses the problem of object recognition in images by first learning the object model, which it represents as a collection of rigid features that are geometrically constrained, from a series of training images. The resulting constellation is modeled probabilistically and tested, according to a set of hypotheses, in order to identify whether the object appears in a new image. Several variations on the original constellation model have improved its efficiency and fully removed human supervision from the learning stage [47]. The constellation model exhibits the benefits of taking an approach that utilizes the geometric relationship between sparse features. It is not directly applicable to odometry as it requires a learning phase and handles object recognition, a tast that demands lower positional precision than scan matching. Furthermore, success depends on proposing a set of hypotheses that contains the correct one, and although efficient heuristics exist, the computation time for large constellations might still be prohibitively large.

A related area of research is the maximum weight clique problem (MWCP) [48, 49]. Given a graph, this problems seeks to find the fully-connected subgraph (i.e., clique) that is the largest (i.e., has the maximum total vertex and edge weights).

The MWCP is known to be NP-complete, but efficient algorithms exist for special types of graphs and heuristic-based approaches [51]. In [50], various strategies for vehicle relocation (i.e., the kidnapped robot problem) are tested, including the maximum clique approach, which determines whether or not the geometric constraints of the current environment match those of a previously traversed path. Effectively, it searches for the hypothesis that is consistent with the largest portion (i.e., clique) of a graph that captures geometric compatibility between feature sets.

A similar research problem is graph matching (also known as the graph isomorphism problem), which searches for a one-to-one mapping between two graphs. In the inexact version, a perfect one-to-one matching may not exist, and the objective is to minimize a distance function that rates the proposed graph match. This problem has been integral to pattern recognition and used extensively in computer vision [52]. Rather than match keypoints individually, both MWCP and graph matching utilize relationships between keypoints to generate robust matchings, and both can be adapted to find matchings that preserve geometric constraints (e.g., pairwise distances between points [53]). It has in fact been proven that the MWCP is equivalent to graph matching under certain conditions [54]. The main difficulties with these approaches are determining how to generate hypotheses and how to decide which ones to test, both of which affect computation time.

The data association algorithm in this thesis falls under the category of inexact graph matching. This approach allows us to utilize geometric information information captured across the entire radar scan—rather than local information around keypoints—while remaining robust to the false positives and other unwanted artifacts that appear in radar scans.

3.2.3 Direct methods

This class of methods does not require feature extraction and operates directly on the raw radar data. Unencumbered by heavy pre-processing, the concepts behind these direct methods are elegant. Yet, it is important to note that, in many cases, feature extraction is not only beneficial but necessary. It allows for the abundant information in radar scans to be efficiently represented by sparse features, which is crucial for efficient computation, particularly when repetitive operations are involved. It is also a prerequisite for other relevant tasks, like object tracking and surface detection, so forgoing this step reduces the overall system's versatility.

Direct scan matching

Checchin et al. work with radar images that are formed by applying a simple threshold to the raw radar scans before transforming them into amplitude gridmaps (i.e. images) [55, 56]. They then search for the rigid body transform that results in the greatest cross correlation between images by exploiting the property that translations and rotations manifest as phase differences and displacements in the Fourier domain. With the Fourier-Mellin transform, the authors are able to efficiently compute the vehicle's motion while utilizing the entire radar output [55, 56]. They supplement the radar with gyroscopes and odometers through an extended Kalman filter. Kellner et al. present another method that does not require keypoint detection [57]. Their Doppler radar returns the position and speed of the surrounding objects, from which they straightforwardly solve for the vehicle's relative motion. Their success is contingent on gathering a sufficient number of reliable detections and measurements. Cornick et al. opt for an entirely

different approach with a ground-penetrating radar that detects stable subsurface features for localization [58]. While robust, this technique requires a radar that is too large and heavy for most road vehicle applications. In addition, because the radar only observes the ground directly beneath it and not the surroundings, its capabilities are limited to routes almost identical to those previously driven.

End-to-end deep learning

In accordance with the recent surge in deep learning, there is potential to perform radar ego-motion estimation and localization with neural networks. The two types of nets most relevant to our research are as follows: (1) one accepts two consecutive inputs and computes the ego-motion estimation between them, and (2) the other accepts a single input and localizes it against a previously traversed path.

The first, termed DeepVO, has been explored for vision systems by Wang et al. and Mohanty et al [59,60]. Wang et al. combine a CNN—intended to simulate the feature extraction module of the traditional image processing pipeline—that feeds into a recurrent neural network—meant to capture the sequential motion dynamics—in order to acquire relative poses from a video sequence [59]. On the other hand, Mohanty et al. utilize a Siamese CNN that, in a known environment, takes only two images as inputs and, in an unknown environment, also accepts a mask highlighting the FAST features extracted directly from the images [60]. The latter approach does not incorporate motion dynamics, which often serves as a valuable piece of information, but this omission also makes it more flexible because it can predict relative poses for arbitrary pairs of images though it achieves qualitatively inferior results compared to Wang et al.

The second network design, termed PoseNet, infers the 6-DOF camera pose from a single image in a previously driven area and succeeds in obtaining median positional errors of less than 5 m (occasionally under 1 m) and median angular errors of less than 10 degrees (occasionally under 5 degrees) [61]. The structure is similar to GoogLeNet with the use of convolutional layers, inception modules, and intermediate classifiers with the appropriate modification of the final activation function from softmax to linear [62]. The authors later propose a strategy to estimate the uncertainty of the pose prediction by computing the variance of outputs when certain neurons are deactivated [63]. One of the main issues with inferring pose relates to the traditional purposes of the neural network, which made its name performing classification. For our purposes, in order to acquire the high accuracy necessary for pose estimation, the detailed information often discarded or blurred in classification becomes salient.

One neural net that attempts to remedy this issue for image segmentation is U-Net, which was explained in Section 3.2.1 and produces pixel-wise labels [32]. Another prominent issue with neural networks is the generation of training data and labels. Insufficient data means that the net is unable to learn the correct patterns and generalize to unseen inputs. Without a sizeable number of accurate labels, the network's performance deteriorates, but producing labels is often expensive and either requires manual tuning or alternative methods that frequently use other sensors, which arguably renders the neural net redundant. Although a generative adversarial network (GAN) is designed to address the problem of lacking enough training data, its performance still relies on having precise labels [64].

3.3 Summary

While all methods discussed are sufficient in certain settings, many do not generalize well and give unpredictable results due to the high levels of noise in radar scans. Adequate performance often requires fine tuning, *a priori* knowledge, restrictive assumptions, outlier detection, or expensive computation. Sharp gradients due to noise and difficulties uniquely identifying keypoints cause descriptor methods to perform inconsistently and require outlier detection. Several of the works rely heavily on other sensors for robustness, which compromises performance under conditions that cause these sensors to fail, or use simultaneous localization and mapping (SLAM), which is accompanied by large overhead costs and model-reliant motion filters [29, 37]. In this thesis, we present robust radar-only motion estimation with algorithms for keypoint extraction and data association. We utilize an indirect method that explicitly selects salient information from noisy artifacts and performs graph matching to fully utilize the information captured in scans. In contrast to the works above, our real-time system operates without other sensors, the creation of maps, model-reliant motion filters, or outlier detection.

Chapter 4

Radar-only odometry: robustness under variable conditions

As part of this integrated thesis, we now present our first paper on this work, which is published in the Proceedings of the 2018 IEEE International Conference on Robotics and Automation. The paper is followed by a brief discussion. Note that the paper adopts a different terminology than that in the rest of this thesis: all references to "landmark" should be treated as references to "keypoint".

2018 IEEE International Conference on Robotics and Automation (ICRA)
May 21-25, 2018, Brisbane, Australia

Precise Ego-Motion Estimation with Millimeter-Wave Radar under Diverse and Challenging Conditions

Sarah H. Cen and Paul Newman

***Abstract*— In contrast to cameras, lidars, GPS, and proprioceptive sensors, radars are affordable and efficient systems that operate well under variable weather and lighting conditions, require no external infrastructure, and detect long-range objects. In this paper, we present a reliable and accurate radar-only motion estimation algorithm for mobile autonomous systems. Using a frequency-modulated continuous-wave (FMCW) scanning radar, we first extract landmarks with an algorithm that accounts for unwanted effects in radar returns. To estimate relative motion, we then perform scan matching by greedily adding point correspondences based on unary descriptors and pairwise compatibility scores. Our radar odometry results are robust under a variety of conditions, including those under which visual odometry and GPS/INS fail.**

I. Introduction

In order to confidently travel through its environment, an autonomous vehicle must achieve robust localization and navigation despite changing conditions and moving objects. Currently, most platforms employ lidar, vision, GPS, internal sensors, or a combination of these systems to obtain information about their surroundings and perform motion estimation. While extremely fast and high-resolution, lidar is sensitive to weather conditions, especially rain and fog. Vision systems are versatile and cheap but easily impaired by scene changes, like poor lighting or the sudden presence of snow. Both optical sensors only yield dependable results for short-range measurements. A typical GPS on its own guarantees at best 3-m accuracy, experiences reception difficulties near obstructions, and relies on an external infrastructure. Additionally, proprioceptive sensors, like wheel encoders and IMUs, suffer from significant drift among other detrimental effects.

In contrast, radar is a long-range, on-board system that performs well under variable lighting and atmospheric conditions, and it is rapidly becoming more affordable and efficient. Due to its long wavelength and beam spread, radar can return multiple readings from the same transmission and generate a grid representation of its world. As a result, radar sensors detect stable, long-range features in the environment.

For these reasons, radar is a promising sensor for odometry, a task for which it is not typically utilized, and we seek to explore its capabilities via a radar-only system. In this paper, we demonstrate robust motion estimation using a frequency-modulated continuous-wave (FMCW) scanning radar alone. Our main contributions are two-fold: (1) a landmark extraction method that reliably identifies meaningful features while avoiding false detections; and (2) a robust radar-only scan matching algorithm that is effective under any rotations and large translations. While radar offers many benefits over the alternatives, working with it is non-trivial as it operates more slowly than lidar, generates noisy measurements, and often returns "ghost" objects. We address these issues and show that radar-only motion estimation succeeds across diverse settings and conditions. Moreover, we show that radar does well compared against vision and GPS even under conditions that are ideal for the latter two sensors.

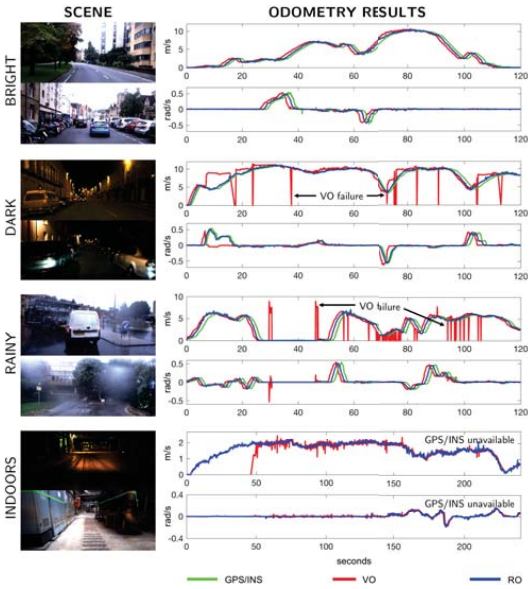


Fig. 1. A comparison of three odometry systems: radar (RO), vision (VO), and GPS/INS (note: the curves are shifted in time for better visual comparison). Only the radar is successful across all scenes, and its motion estimates closely match those of the other sensors when they are available. VO fails in poor lighting or rain, and GPS/INS is unavailable indoors.

Sections II and III-A discuss the relevant literature on radar and introduce the FMCW scanning radar, respectively. Section III details our algorithms on landmark extraction and data association. Section IV discusses our results, and Section V summarizes our work and suggests future directions.

II. Related Works

The first step of most radar motion estimation approaches is the extraction of important features from radar scans. Some researchers [1], [2] draw from the literature on vision, creating amplitude gridmaps that transform the radar scans

S. H. Cen and P. Newman are with the Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, UK (emails: sarah@ and pnewman@robots.ox.ac.uk).

into grayscale images, then extracting features, such as SIFT and FAST. Others [3] use the gridmaps to find continuous areas that are deemed interesting using DBSCAN, MSER, and Connected Components. In contrast, approaches devised specifically for the radar acknowledge its high and irregular noise floor. CFAR [4], [5], a common filtering algorithm, adapts to the variable noise floor along the received signal. Another technique [6], [7] infers the appearance of landmarks by estimating the radar's noise characteristics and exploiting temporal-spatial continuity. Marck et al. [8] avoid the filtering process altogether by recording only the range with the greatest power return per azimuth, but this simplification discards potentially relevant information. As expected, the landmark extractions tailored to radar produce the most meaningful and robust detections.

The next step for motion estimation is the data association of landmarks corresponding to the same object observed at different times. These approaches assume that the majority of the scene is static. Vision-inspired works [1], [2] pair radar landmarks that have sufficiently similar feature descriptors. Modifying this idea for radar occupancy grids, Schuster et al. [9] identify and associate landmarks using the binary annular statistic descriptor (BASD) [10] and Hamming distance. Although feature descriptors work well for images, which contain complex high-density information, they are unable to produce consistent results with radar, for which the readings are much noisier and less dense, and fail to consider the advantages of analyzing the radar scan as a whole.

An alternative to feature-based radar odometry uses multi-sensor fusion. These systems use the other sensors' odometry to transform the incoming radar landmark pointcloud and compare it against an existing map of landmarks. Schuster et al. [9] match each point to its nearest neighbor in the map. Diessler et al. [11] use Monte Carlo methods to select a solution from probabilistically assigned weights. The data association between the radar pointcloud and map provides a motion estimate that is then fused with the original odometry readings. While these multi-sensor methods are promising, none produce satisfactory results using only the radar, and they thus rely on the availability of the other sensors. In addition, they often require the integration of model-reliant filters (e.g. Kalman and particle) and the creation of maps, both of which introduce unnecessary complications.

On the other hand, scan matching, which aligns landmark sets in order to minimize some cost function, accomplishes data association by considering information captured across the entire radar scan and does so without other sensors. One widespread approach, called Iterative Closest Point (ICP), iteratively matches and aligns the pointclouds until the desired termination condition is met [8], [12], [13]. Chandran and Newman [14] adopt a different strategy, developing a function that quantifies the quality of a map created by superimposing radar pointclouds according to the unknown motion parameters; they then perform an exhaustive search to optimize over the motion parameters. Both works assume that the movement between scans is small, which imposes an undesirable constraint on the algorithms and prevents them from being applied to arbitrary inputs. Vivet et al. [5] design an innovative technique well suited for high velocities, using the radar scan distortions, which are typically regarded as a drawback of mobile radar systems, to backsolve for velocity with the help of an extended Kalman filter. Rapp et al. combine spatial scan matching with Doppler information for a joint ego-motion estimate [15].

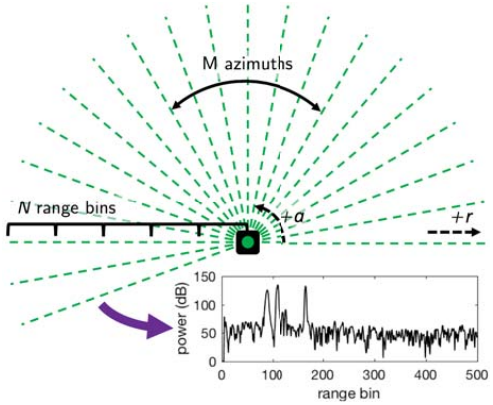


Fig. 2. Visualization of the FMCW scanning radar viewed from above. The radar (green circle), centered on the vehicle (black box), sequentially gathers power-range spectra (dotted green rays) at each azimuth. A sample signal is plotted. Variables a and r denote azimuth and range, respectively.

Other scan matching algorithms do not extract features and operate directly on the radar outputs instead. Checchin et al. [16] apply the Fourier-Mellin transform to efficiently compute the vehicle's rotation and translation from the entire radar output. The Doppler radar used by Kellner et al. [17] returns the position and speed of the surrounding objects, from which the vehicle's motion is easily computed with enough detections. Both concepts are unencumbered by heavy preprocessing. Yet feature extraction is often necessary for other tasks, like object tracking, so forgoing this step reduces the system's overall versatility.

In this paper, we present robust radar-only motion estimation using our own algorithms for landmark extraction and scan matching. We adopt this approach in order to fully utilize the information captured by the radar while providing a method that identifies meaningful radar features that are useful for other tasks. In contrast to the works above, we accomplish these goals without other sensors, the creation of maps, model-reliant filters, or outlier detection.

III. Our Approach

A. FMCW Scanning Radar

We employ the FMCW scanning radar, which is visualized in Figure 2. This radar rotates about its vertical axis while continuously transmitting and receiving frequency-modulated radio waves. The received power corresponding to a position in the environment indicates the reflectivity, size, and orientation of an object at that location. The radar inspects one azimuth at a time. For each, it produces a

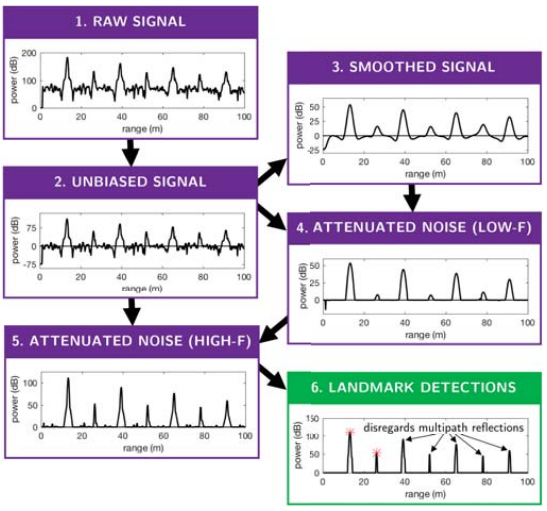


Fig. 3. Procedure for landmark extraction from a power-range spectrum. The input (raw signal) is processed from the top-left to produce the output on the bottom-right, in which the landmarks are denoted with red asterisks. Box 6 in this example highlights the ability of our approach to remove detections due to ghost objects and noise. Boxes 3 and 5 show the importance of incorporating the high-frequency signals, as using the smooth ones in boxes 2 and 4 alone would discard the high range resolution of the FMCW radar.

1D signal, called the power-range spectrum, that contains the power readings for a series of range bins, and one full rotation across all azimuths constitutes a scan. Let N be the number of range bins in a power-range spectrum and M the number of azimuths in one rotation.

The FMCW radar's advantages are its range-measurement accuracy, ability to gather readings at close range, and low peak power. Its disadvantages include sidelobes—radiation sent in unintended directions—and multipath reflections—which result when a wave encounters additional surfaces before returning to the receiver. These two effects cause the scan to contain "ghost" objects at locations where in fact none exist. Errors in range can also occur due to relative motion (via the Doppler effect) and the compression of 3D information into the 1D spectrum. Other issues include phase jitter, saturation, and atmospheric attenuation.

B. Landmark Extraction

Our first objective is to accurately detect objects in the radar's environment with minimal false positives. Specifically, our detector should find all landmarks perceived by the radar while minimizing the number of redundant returns per landmark and avoiding the detection of nonexistent ones, such as those due to noise, multipath reflections, harmonics, and sidelobes. In this section, we present our method for extracting landmarks while adhering to the aims above.

Our method accepts power-range spectra (i.e. 1D signals), as inputs and returns a set of landmarks, each specified by its range and azimuth. The core idea is to estimate the signal's noise statistics then scale the power value at each range

Algorithm 1: Landmark Extraction Method

Input: Power-range spectra $\boldsymbol{s} \in \mathbb{R}^{N\times 1}$ for azimuth a
Output: Set of landmark detections $L(\mathbf{s})$
Parameters: Median filter width w_{median}; binomial filter width w_{binom}; z-value $z_{\mathbf{q}}$ to threshold noise; boolean F and threshold d_{thresh} (optional) for multipath reflection removal.

1 $\mathbf{q} \leftarrow \mathbf{s} - \text{medianFilter}(\mathbf{s}, w_{\text{median}})$
2 $\mathbf{p} \leftarrow \text{binomialFilter}(\mathbf{s}, w_{\text{binom}})$
3 $Q \leftarrow \{q_i : q_i \le 0\}$
4 $\mathcal{N}(\mu_{\mathbf{q}}, \sigma^2_{\mathbf{q}}) \leftarrow \text{normalDistribution}(Q \cup -Q)$
5 Initialize $N \times 1$ vector $\hat{\mathbf{y}}$ to zeros.
6 **for** $i \leftarrow 1$ **to** N **do**
7 **if** $q_i > 0$ **then**
8 $\hat{y}_i \leftarrow p_i \cdot \left(1 - \frac{f(p_i|0,\sigma^2_{\mathbf{q}})}{f(0|0,\sigma^2_{\mathbf{q}})}\right)$
9 $\hat{y}_i \leftarrow \hat{y}_i + (q_i - p_i) \cdot \left(1 - \frac{f(q_i - p_i|0,\sigma^2_{\mathbf{q}})}{f(0|0,\sigma^2_{\mathbf{q}})}\right)$
10 Threshold $\hat{y}_i$ values below $z_{\mathbf{q}}\sigma_{\mathbf{q}}$.
11 $L(\mathbf{s}) \leftarrow \{(a, r(i)) : \hat{y}_i > 0 \cap \hat{y}_{i+1} > 0 \cap \hat{y}_{i-1} = 0\}$
12 If F, remove multipath reflections in $L(\mathbf{s})$ using d_{thresh}.

by the probability that it corresponds to a real detection. Continuous peaks in this reshaped signal are treated as objects; per peak, only the range at the center of the peak is added to the landmark set.

Let the vector $\mathbf{s}(t) \in \mathbb{R}^{N\times 1}$ be the power-range spectrum at time t such that the element s_i is the power return at the i-th range bin, and $a(t)$ is the associated azimuth. Let $r(i) = \beta(i - 0.5)$ give the range of bin $i \in \{1, 2, \ldots, N\}$, where β is the range resolution. Suppose that $\mathbf{y}(t) \in \mathbb{R}^{N\times 1}$ is the ideal signal if the environment was recorded perfectly. Then, $\mathbf{s}(t) = \mathbf{y}(t) + \mathbf{v}(\mathbf{y}(t))$, where $\mathbf{v}$ represents unwanted effects, like noise. Therefore, inferring $\mathbf{y}(t)$ from $\mathbf{s}(t)$ in order to accurately isolate the landmarks requires an approximation of $\mathbf{v}(\mathbf{y}(t))$ such that $\hat{\mathbf{y}}(t) = \mathbf{s}(t) - \hat{\mathbf{v}}(\mathbf{s}(t))$. Removing $\hat{\mathbf{v}}$ from $\mathbf{s}$ is the aim of our method. The landmark detections extracted from $\hat{\mathbf{y}}(t)$ are stored in the set $L(\mathbf{s}(t))$.

The landmark extraction method, as described next, references Figure 3 and Algorithm 1. To begin, an unbiased signal $\mathbf{q}$ that preserves high-frequency information (box 2) is acquired by subtracting the noise floor of $\mathbf{v}(\mathbf{s})$ from $\mathbf{s}$ (line 1). The result is then smoothed to obtain the underlying low-frequency signal $\mathbf{p}$ (box 3), which better exposes obvious landmark peaks (line 2). At this point, $\mathbf{q}$ is not discarded for two reasons: radar landmarks often manifest as high frequency peaks, so smoothing would dampen their presence; and smoothing muddles the peaks of landmarks that are in close proximity, making it difficult to distinguish between them. Thus, we integrate the information of both $\mathbf{q}$ and $\mathbf{p}$.

To estimate the noise characteristics, we treat the values of $\mathbf{q}$ that fall below zero as Gaussian noise with mean $\mu_{\mathbf{q}} = 0$ and standard deviation $\sigma_{\mathbf{q}}$ (line 4). Let $f(x|\mu, \sigma^2)$ be the probability density at x for the normal distribution $\mathcal{N}(\mu, \sigma^2)$.

Then, for every range bin, the power values are scaled by the probability that they do not correspond to noise in two steps. First, each value of the smoothed signal p_i is scaled by $f(p_i|0,\sigma_{\mathbf{q}}^2)$ (box 4 and line 8). This process is repeated for the high-frequency signal q_i relative to the smoothed signal p_i such that the scaling factor is $f(q_i|p_i,\sigma_{\mathbf{q}}^2)$ (box 5 and line 9). The sum of both values is stored in $\hat{y}_i$. These steps integrate high- and low-frequency information to preserve range accuracy while suppressing signal corruptions due to noise. Finally, the $\hat{y}_i$ values that are below the upper $z_{\mathbf{q}}$-value confidence bound of $\mathcal{N}(0,\sigma_{\mathbf{q}}^2)$ and therefore less likely to represent real landmarks are set to zero (box 6 and line 10).

The method extracts landmarks from $\hat{y}_i$ (the black signal in box 6) as follows. All values of $\hat{\mathbf{y}}$ are now either zero or belong to a peak. For each peak's center located at range bin i, the tuple $(a, r(i))$ is added to the landmark set $L(\mathbf{s})$ (line 11). These landmarks are then tested, and those identified as multipath reflections (MR) are removed (box 6 and line 12). Since MRs cause peaks with similar wavelet transform (WT) signatures to appear in the power-range spectrum at different ranges with amplitudes that decrease with distance, this step compares the continuous WTs $\mathbf{w}_i, \mathbf{w}_j \in \mathbb{R}^{H\times 1}$ for each set of peaks P_i and P_j where $j > i$. If $d_{ij}/H < d_{\text{thresh}}$ and the maximum power of P_i is greater than that for P_j, where $d_{ij} = |\frac{\mathbf{w}_i}{\max(\mathbf{w}_i)} - \frac{\mathbf{w}_j}{\max(\mathbf{w}_j)}|$ is a measure of dissimilarity, then P_j is considered a MR, and $(a, r(j))$ is removed from $L(\mathbf{s})$. MR removal produces good results but requires significant computation time, making it optional.

Our method requires three free parameters with an optional fourth. In general, w_{median} should represent a distance large enough to span multiple landmarks, and w_{binom} should be around the width of an average peak ($\sim \frac{w_{\text{median}}}{2}$). A greater $z_{\mathbf{q}}$ value raises the standard for peaks to be chosen as landmarks over noise, and d_{thresh} is the minimum difference between WTs for detections to be considered independent. For the following analyses, let $\mathbb{L} = \bigcup_{t\le\tau<t'} L(\mathbf{s}(\tau))$ be the set of all landmarks in one full scan from time t to t'.

C. Data Association

In this section, we present a scan matching algorithm that achieves robust point correspondences using high-level information in the radar scan. Intuitively, it seeks to find the largest subsets of two pointclouds that share a similar shape. Unlike ICP, this method functions without a priori knowledge of the scans' orientations or displacements relative to one another. Thus, our algorithm is not constrained to have a good initial estimate of the relative pose and can compare pointclouds captured at arbitrary times without a map. The only requirements are that the areas observed lie in the same plane and contain sufficient overlap.

One of the key attributes of our approach is to perform data association using not only individual landmark (i.e. unary) descriptors, but also the relationships between landmarks. For instance, imagine three landmarks that form the vertices of a scalene triangle. Then, the set of distances from each point to its neighbors is unique to that point regardless of the overall pointcloud's placement, allowing the landmark to be straightforwardly matched to its counterpart in any other pointcloud acquired by applying a rigid body transformation to the original triangle. The greater the number of points, the less likely it is for an individual point to have the same set of pairwise distances to its neighbors as another. Moreover, the exact position and orientation of the pointcloud does not influence the pairwise relationships within it, so great disparities between the placements and orientations of the pointclouds are inconsequential. We harness these observations to obtain reliable matches for our large landmark sets. With real data, the main challenges are that the landmark locations and detections are noisy, meaning that points do not always survive the rigid body transformation and the locations of those that do are affected by noise.

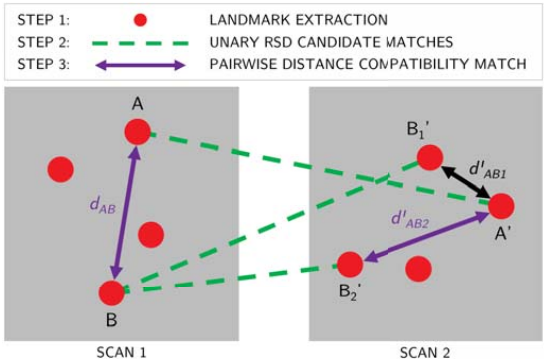


Fig. 4. The core idea behind our data association algorithm that seeks to find similar shapes within the two landmark pointclouds (in red) extracted from radar scans. The unary candidate matches (dotted green lines) are generated by comparing the points' angular characteristics. The selected matches (A, A') and (B, B_2') minimize the difference between pairwise distances ($|d_{AB} - d'_{AB2}| < |d_{AB} - d'_{AB1}|$). In this way, we approximate a shape matching by sequentially comparing angles and side lengths.

A simple example illustrating the concept behind our data association algorithm is shown in Figure 4. The method is given in Algorithm 2, which we reference in the following explanation, and its helper functions appear in the Appendix. As inputs, it accepts two pointclouds $\mathbb{L}^O$ and $\mathbb{L}^I$ for each of the two radar scans. The first $\mathbb{L}^O$ is the original set of landmarks in Cartesian coordinates. Because landmarks are detected in polar space, the resulting pointcloud will be dense at low ranges and sparse at high ones. The second $\mathbb{L}^I$ compensates for this by generating a binary Cartesian grid of resolution β that is interpolated from the binary polar grid of landmarks. The latter pointcloud is less exact and only used to sidestep the range-density bias when processing the layout of the environment while data association is performed on the former (i.e. the algorithm returns a set of matches $\mathcal{M}$ that contains tuples (i, j) such that the landmark $\mathbb{L}_1^O\{i\}$ corresponds to $\mathbb{L}_2^O\{j\}$). This distinction is a key insight. It preserves accuracy by operating on the landmarks detected in polar space while correcting for a main difficulty of scanning radars by interpreting the environment in Cartesian space.

The data association is then performed in four steps. First, for every point in $\mathbb{L}_1'$, the unaryMatches function suggests a

Algorithm 2: Data Association Method

Input: Landmark sets for two scans $\mathbb{L}_1^O$, $\mathbb{L}_1^I$, $\mathbb{L}_2^O$, and $\mathbb{L}_2^I$.
Output: Set of landmark matches $\mathcal{M}(\mathbb{L}_1^O, \mathbb{L}_2^O)$.
Parameters: Percentage α of landmarks that, if matched, cause the algorithm to terminate.

1 $B \leftarrow$ unaryMatches$(\mathbb{L}_1^O, \mathbb{L}_1^I, \mathbb{L}_2^O, \mathbb{L}_2^I)$ and $W \leftarrow |B|$
2 $C_{W\times W} \leftarrow$ pairwiseCompatibilities$(B, \mathbb{L}_1^O, \mathbb{L}_2^O)$
3 $\mathbf{u}^* \leftarrow$ normalizedMaxEigenvector(C)
4 Initialize the empty set $\mathcal{M}$.
5 Initialize the $W \times 1$ vector *unsearched* to all *True*.
6 **while** *(any True in unsearched)* $\cap$ $(|\mathcal{M}| < \alpha W)$ **do**
7 $\quad$ $(max_match, max_reward) \leftarrow$ $(i, u_i^{*2}) : u_i^{*2} \geq u_j^{*2}\ \forall\, i, j \in unsearched)$
8 $\quad$ Terminate function if $(max_reward \cdot W < 1)$.
9 $\quad$ Add the match $B\{max_ind\}\}$ to $\mathcal{M}$.
10 $\quad$ $searched \leftarrow \{i : B\{max_match, 1\} = B\{i, 1\} \cup B\{max_match, 2\} = B\{i, 2\}\}$
11 $\quad$ $unsearched_{i \in searched} \leftarrow False$

potential point match in $\mathbb{L}_2'$ based on some unary comparison method (line 1). We discuss the *unaryMatches* function in Section III-D. Next, the non-negative compatibility score for each pair of proposed matches $g = (i, i')$ and $h = (j, j')$ is computed and assigned to the elements (g, h) and (h, g) of the $W \times W$ matrix C such that it is symmetric and diagonally dominant (line 2). If the landmark matches g and h are correct, then the relationship between i and j in the first radar scan is similar to that between i' and j' in the second; the compatibility score reflects this pairwise similarity. In our method, the value is computed from the distances between corresponding pairs of points in the two scans. It reflects the understanding that real, correctly identified landmarks are the same distance apart in any two radar scans.

The optimal set of matches $\mathcal{M}$ maximizes the overall compatibility, or reward. Suppose that $\mathbf{m} \in \{0, 1\}^W$ such that (1) $m_i = 1$ if the unary match $B\{i\}$ is deemed plausible and $m_i = 0$ otherwise; and (2) the selected matches do not conflict (i.e. a point in one pointcloud cannot correspond to two in the other). Then, the optimal solution $\mathbf{m}^*$ satisfies

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

Due to the discretization of $\mathbf{m}$, this maximization is computationally difficult, so we relax the aforementioned constraint to seek the continuously-valued $\mathbf{u}^*$ such that

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

Under these conditions, $\mathbf{u}^*$ is the normalized eigenvector of the maximum eigenvalue of the positive semi-definite matrix C. The optimal solution $\mathbf{m}^*$ is then be approximated from $\mathbf{u}^*$ using the greedy approach shown in lines 3-11.

In short, the greedy method iteratively adds satisfactory matches to the set $\mathcal{M}$. On each iteration, the remaining valid matches are evaluated (line 7), that which returns the maximum reward is accepted (line 9), and those that conflict with it are removed from further consideration (lines 10 and 11). When the most recently selected match yields a reward less than the that if all matches were valued equally (i.e. is a weak match) or more than α percent of the landmarks in either set are matched, the algorithm terminates (lines 6 and 8). Note that α is the only free parameter in this method, and no outlier removal is required.

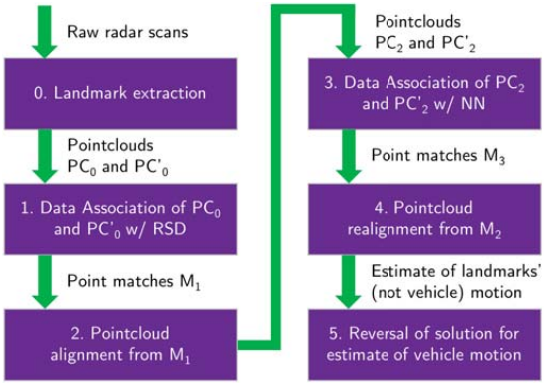


Fig. 5. The relative motion estimation pipeline.

D. Relative Motion Estimation

Given two sets of corresponding points, Challis [18] presents a method for finding the rigid body motion that optimally aligns them in the least-squares sense using singular value decomposition (SVD). We apply this technique to $\mathcal{M}$ to estimate the relative motion between two radar scans.

Our motion estimation algorithm, summarized in Figure 5, first performs a data association (step 1) that is capable of aligning pointclouds separated by any distance and rotation. The relative motion is then estimated with Challis' method (step 2). The second round of data association (step 3), which serves only to refine the motion estimate, makes use of the fact that the new pointclouds are now approximately aligned when generating unary matches. The final motion calculation is performed on the new set of dense point matches (step 4). Two examples of this process are shown in Figure 7.

The only difference between steps 1 and 3 is the choice of unary matching function. In step 1, each point is associated with its closest match in the other pointcloud according to what we term the *Radial Statistics Descriptor* (RSD). This descriptor specifies each landmark $\boldsymbol{\ell}$ by the radial statistics of its neighboring points along every angular slice centered at $\boldsymbol{\ell}$ (Algorithm 3 in the Appendix). In step 3, the unary match of a point is its nearest neighbor in the other pointcloud.

Recall that the overarching intuition for our scan algorithm is to find and match the largest subsets of points in two pointclouds that share a similar shape. Generally, a shape is uniquely defined by two characteristics: its side lengths and angles. While shape matching is computationally demanding,

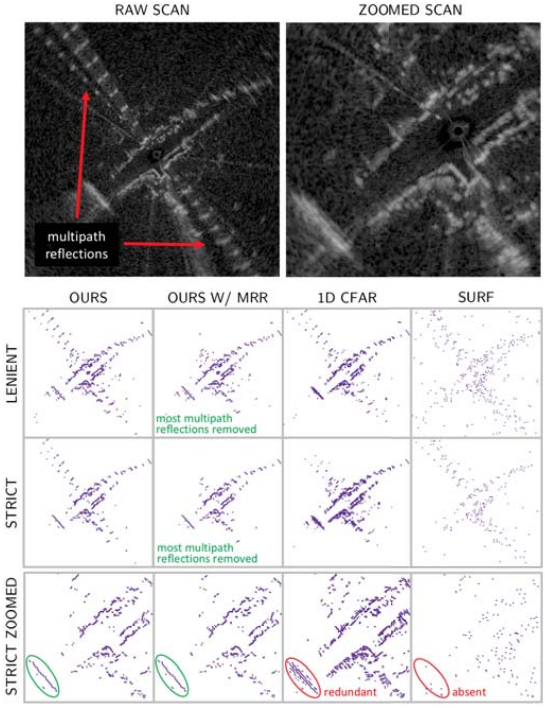


Fig. 6. A comparison of our landmark extraction methods to 1D CFAR [4], [5] and SURF [1], [2]. The raw radar images are at the top, and each table column features a method. The first and second rows give the pointclouds for lenient (i.e. more detections with the potential for more noise) and strict parameters, respectively. The bottom row zooms in on the latter. Only our method with multipath reflection removal (MRR) successfully disregards MRs. 1D CFAR contains redundant returns (e.g. for the wall at the bottom left), and SURF does not consistently find real objects.

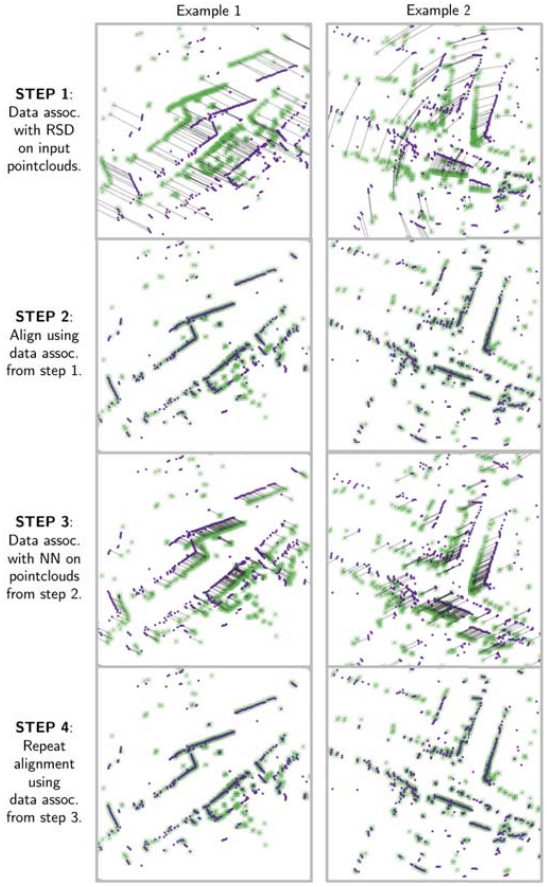


Fig. 7. Illustration of the data associations and pointcloud alignments corresponding to steps 1 through 4 in Figure 5. The displacement between the two pointclouds (purple and green) in step 3 is inserted to better visualize the point matches (black lines). Note that step 1 operates on pointclouds separated by any distance and rotation since it only uses the pointclouds' shapes. Steps 2 and 4 show the alignments produced by the matches in steps 1 and 3, respectively. Step 3 refines the alignment of step 2, thus improving the motion estimate, as evidenced by the denser matches.

our algorithm efficiently approximates this operation by capturing angular information in the RSD unary matches and side lengths in the pairwise compatibility calculation, and this combination ensures robust scan matching.

IV. Results

We utilize the Navtech CTS350-X, a FMCW scanning radar without Doppler information. For this radar, $M = 399$, $N = 2000$, and $\beta = 0.25$ m. The beam spread is 2 degrees in azimuth and 25 degrees in elevation. The radar operates at 4 Hz, and our algorithm (not fully optimized) operates at approximately 3 Hz. The radar is placed on the roof of a ground vehicle with an axis of rotation perpendicular to the driving plane. We adopt the usual odometry assumptions that the environment is mostly static and non-deformable. We also assume that the instantaneous motion of the vehicle is planar. We utilize the following parameters, chosen empirically: $w_{\text{median}} = 200$, $w_{\text{binom}} = 50$, $z_{\mathbf{q}} = 2.5$, $d_{\text{thresh}} = 0.1$, and $\alpha = 0.5$ with MR removal. When driving, the vehicle typically travels between 5 and 10 m/s; when turning, up to 0.6 rad/s (see Figures 1 and 8). The vehicle is driven through various parts of downtown Oxford, UK.

Figure 6 compares our landmark extraction algorithm to other common approaches. It examines the use of image processing features for radar localization and mapping [1], [2] via SURF. Designed to detect sharp gradients, SURF is highly susceptible to unwanted radar artifacts. As shown, it returns numerous false positives, and many landmarks do not correspond to meaningful objects in the scene, demonstrating the need for radar-specific landmark detection in order to perform robust motion estimation. In contrast, 1D CFAR yields a better depiction of the surroundings, from which structures (e.g. walls and buildings) are easily discernible. However, we maintain that the landmark set produced by our algorithm is preferable for the following reasons. Our pointclouds are qualitiatively clean and crisp with few false detections, especially with multipath reflection removal (MRR). The

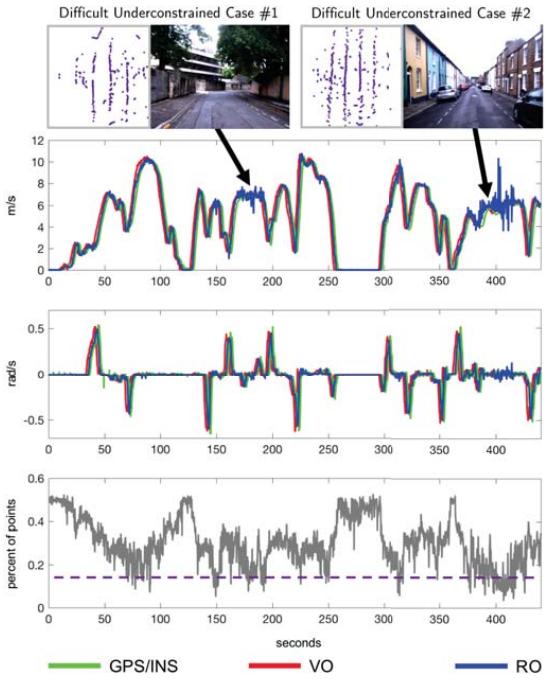


Fig. 8. A comparison of odometry using the radar (RO), vision (VO), and GPS/INS under favorable conditions for vision and GPS/INS (note: the curves are slightly shifted in time for better visual comparison). This route illustrates the success of RO (it closely matches the estimates of VO and GPS/INS) while including portions of its worst performance. The images and landmark sets when RO is noisiest show that this occurs in featureless narrow environments where the pointclouds are underconstrained between scans. The top and middle plots show the estimated translational and angular velocities, respectively. The bottom plot tracks the portion of the landmark pointcloud that is matched during the data association step using RSD. This value appears to provide an uncertainty measure, as it is approximately inversely correlated with RO performance, an added benefit of our approach.

landmarks accurately reflect the coherent structure of the surroundings, and our approach avoids returning redundant points corresponding to the same object. Finally, adjusting $z_{\mathbf{q}}$ in our algorithms tunes the leniency in an intuitive manner (i.e. raising $z_{\mathbf{q}}$ reduces the number of landmarks, beginning with those that appear to be noise). Yet adjusting the parameters for 1D CFAR deteriorates the pointcloud quality in some places and improves it in others. This property makes the CFAR filter parameters difficult to tune.

Figure 7 displays the outcomes of our data association algorithm through two scan matching examples. The steps 1 and 3 show point matches that are dense, accurate, and consistent with one another without any outlier removal. These correspondences provide the highly precise and clean scan matches in the steps 2 and 4. Because step 1 can be applied to pointclouds separated by any displacement as long as they contain sufficient scene overlap, the matches in step 3, which exploit the approximate alignment from step 2, are more dense, thus producing the refined alignment in step 4.

Figures 1 and 8 illustrate the odometry results across three

TABLE I

ERROR STATISTICS FOR RO COMPARED AGAINST VO

Translational Velocity Error (m/s)			
Quantile	Diverse setting	Busy city center	Narrow backstreets
0.25	0.0293	0.0420	0.0495
0.50	0.1057	0.1134	0.1218
0.75	0.2116	0.2094	0.2321
Rotational Velocity Error (deg/s)			
Quantile	Diverse setting	Busy city center	Narrow backstreets
0.25	0.0824	0.0801	0.2108
0.50	0.3210	0.3188	0.5165
0.75	0.7118	0.7017	0.9589

sensor systems. RO is robust across a variety of settings and conditions. Importantly, even when visual odometry (VO) [19] and GPS/INS are available, they are closely matched by RO. In the dark and rain, VO periodically fails while RO gives a clean and smooth result. The data for the bottom example of Figure 1 was captured indoors in a dark, crowded, and enclosed area. As a result, GPS/INS is unavailable and VO is sporadic. Due to the presence of large nearby metallic objects, which cause the scans to be noisy, the RO curves are less smooth than the outdoor ones, but the radar still produces the most reliable motion estimation. Figure 8 shows a route chosen because it contains areas in which our RO performs suboptimally. Our RO system experiences difficulties when the landmark pointclouds contain insufficient information to conclusively deduce the vehicle's motion (i.e. are underconstrained). Specifically, narrow corridors appear as parallel lines in consecutive radar scans; due to the lack of variability in the scene, our algorithm generates visually satisfactory alignments that do not reflect the actual movement between the scans. We can evaluate the uncertainty of our motion estimate by examining the portion of the landmark pointcloud that is matched using RSD during data association. Since our approach seeks the largest number of matches that are mutually consistent, a low number indicates either that the two pointclouds are dissimilar or multiple solutions exist.

Table I confirms that the error for RO is greater when driving through narrow backstreets than in the busy city center. Over a 10 km route through Oxford, UK (the diverse setting), the median RO error is about 0.106 m/s in translation and 0.321 deg/s in rotation. We compare against VO because it provides accurate and fine-grained odometry while that from GPS/INS odometry is smoothed.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced a robust radar-only motion estimation system that rivals the performance of VO (even under conditions optimal for vision) and demonstrates the importance of radars for autonomous vehicles. As an onboard sensor that operates under diverse conditions, the successful implementation of RO improves the reliability and versatility of mobile systems. Our method stands out because it is not only dependable and accurate, but also straightforward and intuitive with few free parameters, no outlier-detection methods, and no model-reliant filters. Although multi-sensor fusion is beneficial, we show that RO can stand

alone if the other sensors drop out. Our landmark extraction algorithm produces sparse yet meaningful detections with minimal false positives. It feeds into our data association algorithm, which performs scan matching using a greedy approach that, intuitively, seeks to find the largest subsets of the two pointclouds that share similar shapes. The resulting radar-only motion estimation is accurate and robust under conditions for which other common sensor systems fail.

In the future, we intend to address the scenarios in which RO encounters difficulties and to develop a technique that accurately quantifies the uncertainty of our component estimates. We also aim to include comprehensive comparative studies against other RO methods.

REFERENCES

[1] J. Callmer, D. Törnqvist, F. Gustafsson, H. Svensson, and P. Carlbom, "Radar slam using visual features," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 71, 2011.

[2] F. Schuster, C. G. Keller, M. Rapp, M. Haueis, and C. Curio, "Landmark based radar slam using graph optimization," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 2559–2564.

[3] K. Werber, J. Klappstein, J. Dickmann, and C. Waldschmidt, "Interesting areas in radar gridmaps for vehicle self-localization," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on*. IEEE, 2016, pp. 1–4.

[4] H. Rohling, "Ordered statistic cfar technique-an overview," in *Radar Symposium (IRS), 2011 Proceedings International*. IEEE, 2011, pp. 631–638.

[5] D. Vivet, P. Checchin, and R. Chapuis, "Localization and mapping using only a rotating fmcw radar sensor," *Sensors*, vol. 13, no. 4, pp. 4527–4552, 2013.

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

[7] ——, "Relative radar cross section based feature identification with millimeter wave radar for outdoor slam," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 1. IEEE, 2004, pp. 425–430.

[8] 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 *Radar Conference (EuRAD), 2013 European*. IEEE, 2013, pp. 471–474.

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

[10] M. Rapp, K. Dietmayer, M. Hahn, F. Schuster, J. Lombacher, and J. Dickmann, "Fscd and basd: Robust landmark detection and description on radar-based grids," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on*. IEEE, 2016, pp. 1–4.

[11] T. Deissler and J. Thielecke, "Uwb slam with rao-blackwellized monte carlo data association," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–5.

[12] P. J. Besl, N. D. McKay *et al.*, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

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

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

[15] M. Rapp, M. Barjenbruch, M. Hahn, J. Dickmann, and K. Dietmayer, "Probabilistic ego-motion estimation using multiple automotive radar sensors," *Robotics and Autonomous Systems*, vol. 89, pp. 136–146, 2017.

[16] 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.

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

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

[19] W. Churchill, "Experience based navigation: Theory, practice and implementation," Ph.D. dissertation, University of Oxford, Oxford, United Kingdom, 2012.

APPENDIX

Algorithm 3: Compute RSD for Landmark

Input: Landmark sets $\mathbb{L}^O$ and $\mathbb{L}^I$; index i of landmark.
Output: The $3 \times M$ matrix RSD. Element (i, j) is the i-th statistic of the points in j-th azimuth slice, which are ordered CCW, starting with the slice that contains the most points.

Initialize $3 \times M$ matrix RSD to zeros.
Separate space around $\mathbb{L}_i^O$ into M equal azimuth slices.
for azimuth slice $j \leftarrow 1$ **to** M **do**
    Find set S of all points in $\mathbb{L}^I$ that lie in slice j.
    $RSD[:, j] \leftarrow [|S|, \text{arithmMean}(S), \text{harmonMean}(S)]$
Normalize all rows in *RSD*.
Shift *RSD* columns s.t. highest-density slice is the first.

Algorithm 4: Generate Unary Matches based on RSD

Input: Landmark sets $\mathbb{L}_1^O$, $\mathbb{L}_1^I$, $\mathbb{L}_2^O$, and $\mathbb{L}_2^I$.
Output: Set of unary match indices B.

Initialize the empty set B.
RSD1 $\leftarrow$ getAllPointsRSD($\mathbb{L}_1^O, \mathbb{L}_1^I$)
RSD2 $\leftarrow$ getAllPointsRSD($\mathbb{L}_2^O, \mathbb{L}_2^I$)
for point $i \leftarrow 1$ **to** $|\mathbb{L}_1^O|$ **do**
    Find point j in $\mathbb{L}_2^O$ to minimize $||RSD2_j - RSD1_i||$.
    Add unary match (i, j) to set B.

Algorithm 5: Generate Unary Matches based on NN

Input: Landmark sets $\mathbb{L}_1^O$ and $\mathbb{L}_2^O$.
Output: Set of unary match indices B.

Initialize the empty set B.
for point $i \leftarrow 1$ **to** $|\mathbb{L}_1^O|$ **do**
    Find point j in $\mathbb{L}_2^O$ to minimize $||\mathbb{L}_2^O\{j\} - \mathbb{L}_1^O\{i\}||$.
    Add unary match (i, j) to set B.

Algorithm 6: Compute Pairwise Compatibility

Input: Set of unary matches B; landmark sets $\mathbb{L}_1^O$ and $\mathbb{L}_2^O$; indices of unary matches i and j.
Output: Pairwise compatibility score C_{ij} for (i, j) pair.

if i equals j **then**
    $C_{ij} \leftarrow |B|$ and return.
$\boldsymbol{\ell}_{1,i}, \boldsymbol{\ell}_{1,j} \leftarrow \mathbb{L}_1^O\{B\{i, 1\}\}, \mathbb{L}_1^O\{B\{j, 1\}\}$
$\boldsymbol{\ell}_{2,i}, \boldsymbol{\ell}_{2,j} \leftarrow \mathbb{L}_2^O\{B\{i, 2\}\}, \mathbb{L}_2^O\{B\{j, 2\}\}$
$d_1 \leftarrow ||\boldsymbol{\ell}_{1,i} - \boldsymbol{\ell}_{1,j}||^2$ and $d_2 \leftarrow ||\boldsymbol{\ell}_{2,i} - \boldsymbol{\ell}_{2,j}||^2$
$C_{ij} \leftarrow (1 + |d_1 - d_2|)^{-1}$

4.2 Discussion

4.2.1 Summary of contributions

The aim of this work is to achieve ego-motion estimation with performance comparable to state-of-the-art methods when using a difficult sensor: radar. This paper demonstrates that a radar-only system with 25-cm range resolution can provide centimeter-level odometry accuracy. It then extends the results to establish that radar maintains good performance even in conditions under which other popular sensors (e.g., lidar, vision, and GPS) fail. By demonstrating ego-motion estimation, a classical task for mobile autonomy, this work shows that radar can be a valuable sensor for autonomous vehicles.

This work presents two algorithms—one for keypoint extraction and another for data association—that address the difficulties of working with radar. The odometry pipeline is designed to be straightforward and easy to use, lessening the reliance on human expertise. The keypoint extraction method prioritizes keypoint sparsity and location accuracy in order to supply the data association module with a concise representation of the environment that will be consistent from scan to scan. These priorities differentiate it from other works that perform smoothing or segmentation on radar scans. The keypoint sets are then fed into the data association module, which matches keypoints based on global geometric relationships in an efficient approximation of shape matching. The data association module naturally removes outliers, making it robust to unwanted artifacts in radar scans.

The advantages and drawbacks of the keypoint extraction and data association algorithms are analyzed in-depth in Chapter 6.

Further explanation of this work and results can be found in our video [65].

4.2.2 Design choices

This section explains the design choices of our approach.

Why develop a radar-only system?

This work does not suggest that mobile autonomous systems should utilize only radar for ego-motion estimation or localization. To the contrary, radar has many characteristics that would complement other sensors in a well-rounded sensor suite. However, this work focuses on radar-only odometry in order to demonstrate the capabilities of radar, which is generally underutilized in mobile autonomy. It aims to show that radar is an information-rich, resilient sensor while also emphasizing the benefits of deliberate algorithm design.

Why use an FMCW scanning radar?

As described in Chapter 2, FMCW scanning radar offers several benefits. Compared to other radars, FMCW radars have high range resolution, low peak power, low minimum range, and fast measurement updates, and they are generally more compact and less expensive. These characteristics make it well suited for vehicles, for which accuracy, safety, payload, and cost are important factors. Although some applications use radar for obstacle avoidance, most tasks require a holistic view of the surroundings. For these tasks, scanning radars can provide a 360-degree field-of-view around the vehicle.

Why work on ego-motion estimation?

Ego-motion estimation is a classical, foundational problem in mobile autonomy. Furthermore, radar is well suited for ego-motion estimation because it can detect stable, long-range features in the environment.

Why use an indirect method?

Separating ego-motion estimation into two steps—keypoint extraction then data association—allows for a modular pipeline that can be adapted for different purposes. For instance, (1) the keypoint extraction module can be modified for a different sensor or data type, (2) the data association module can be modified for dynamic object detection, or (3) extracted keypoints can be re-purposed for other tasks, like object tracking.

How was the keypoint extraction method formulated?

There are several methods commonly used in the literature for keypoint extraction or segmentation in radar scans. As further explained in Chapter 6, we deliberately design a new keypoint extraction method for several reasons, as follows:

- **Sparsity**. The computational resources required by data association scales with the number of keypoints it needs to match, making it desirable to have as few keypoints as necessary to adequately describe the scene. To achieve this objective, the our keypoint extraction algorithm first identifies peaks in the power-range spectrum that are believed to correspond to objects. Then it returns one keypoint per peak (i.e., object) in the power-range spectrum. In comparison, methods like CFAR, MSER, and DBSCAN are region-based,

meaning that they produce good occupancy grids but are less suitable for our system needs.

- **Object correspondence**. There are a few reasons why radar keypoints should correspond to meaningful features, like objects. In particular, radar scans are noisy, so if keypoints are based on features in one scan that do not correspond to objects, it is likely that the same feature does not appear in a subsequent scan, making it difficult to perform feature association. Because objects (e.g., walls or trees) are more permanent, they are more consistent keypoints to use in data association. In addition, keypoints that correspond to objects can be used for other tasks, like object detection and tracking. In comparison, common image features, like SURF and FAST, search for corners using intensity gradients. When applied to radar, these features perform poorly because high gradients are often caused by noise and other unwanted artifacts rather than persistent objects, as shown in [66].

- **Spatial accuracy**. Unlike object classification or occupancy grid prediction, ego-motion estimation requires high spatial accuracy. As radar already has lower resolution than lidar and vision sensors, it is important that the spatial information that is provided by radar is properly utilized and not discarded. Many radar-based works pre-process the scans by smoothing in order to mitigate the effects of speckle noise. However, doing so also smooths away high-frequency information (e.g., the location of a wall can be shifted by a range bin due to an asymmetric peak). In order to also maintain sparsity (as described above), our method first finds peaks corresponding to objects, then

returns the location of a single keypoint that is most likely to correspond to the reflector.

- **Noise resistant**. The identification and location of keypoints should be as minimally affected by noise as possible. As mentioned directly above, features borrowed from image processing do not respond well to speckle noise. The performance of other methods, like CFAR, depend on the environment and require (often unintuitive) tuning of parameters, as explained in our paper [66]. In comparison, our approach seeks to reduce the effects of noise by first identifying objects, then returning a single keypoint rather than returning keypoints corresponding to each peak in the power-range spectrum. Like CFAR, it also relies on a few parameters; the parameters are intended to be intuitive and thus easy to tune.

- **Efficiency**. Keypoint extraction should be efficient. In our case, there is a trade-off between computational efficiency and perfect identification of keypoints (i.e., all keypoints correspond to real objects in the environment). For ego-motion estimation, keypoint sparsity, reliability, and spatial accuracy are the main priorities. As such, we allocate the task of outlier detection to the data association module.

How was the data association method formulated?

Our data association approach is based on the following considerations:

- **Geometric or shape-based methods**. Radar is best suited for identifying objects in the environment. In comparison to images, radar scans lack the

rich, multi-channel information that allows individual objects to be easily recognized between scans. As such, we take a geometric approach that makes use of the relationships *between* keypoints. In this sense, it approximates shape matching by seeking to preserve the geometric constraints between keypoints. To do this, data association is formulated as an optimization problem that falls under the category of graph matching [53].

- **Motion priors**. Many data association methods (e.g., ICP) rely on the availability of a good prior on the relative displacement between scans, which can be restrictive. This consideration also motivates the use of a geometric approach because shape matching is coordinate-independent and can be performed without a prior on the displacement between scans.

- **Long-range capabilities**. One of the main benefits of radar is its ability to detect long-range objects. Akin to how torque scales with the length of the lever arm, correctly matching faraway objects can significantly improve ego-motion estimation accuracy. However, the spatial resolution of radar decreases with range (because radar information is gathered in polar coordinates). Because the density of keypoints also decreases with range, faraway keypoints are naturally weighted less than close ones. Due to this observation, our data association algorithm does not apply keypoints weights based on reliability. In various tests, we empirically verified that performance is better when the reliability of a keypoint is not weighted by its range.

- **Feature descriptors**. Our method uses a feature descriptor to propose unary matches. The matches provided by feature descriptors are only treated

as proposals because they are often incorrect due to the many artifacts in radar scans. In designing the descriptor, it became clear that local gradient-based feature descriptors, such as those used in vision, do not work well for radar (see the discussion in the previous section). Instead, we use a descriptor that is based on information across the entire radar scan similar to BASD [38]. An important step in the computation of the descriptor is the weighting of keypoints by their distance from the radar. Recall that radar has a sampling bias: it produces many more close-range samples than faraway ones because it scans its environment in a polar fashion. Because the descriptor for a keypoint should depend on the object's characteristics and *not* on the radar's position, it is essential that the sampling bias be corrected in the descriptor. In various experiments, we verified that this step significantly affects performance.

- **Robustness to false positives and negatives**. As many of the keypoints correspond to false positives or negatives in the radar scan, the data association method must be robust to changes in appearance. To achieve this goal, our approach formulates data association as a global optimization problem, thus searching for a solution that is most favorable across the entire set of keypoints.

- **Outlier detection**. Identifying and removing outliers is essential to data association. This step usually requires tuning a threshold. In recognition of this, our algorithm seeks to threshold in a principled manner by terminating when the next match scores lower than if all matches were weighted equally.

- **Efficiency**. Computational complexity is a key consideration of data association. Various methods for data association are formulated around hypothesis testing or exhaustively search for motion parameters that best align the scans. Other methods, like ICP, repeatedly transform the keypoint sets until they are aligned. We sought to avoid both strategies. Our approach approximates the optimal solution in one step, then greedily adds matches that are mutually consistent without any hypothesis testing.

What methods of ground truth were used to evaluate performance?

In this paper, RO was tested against two well known sensor systems: VO and GPS. The VO system used is described in [67]. For the Kitti Vision Benchmark Suite, this VO system generally achieves between 2 and 3 percent translational error for path lengths between 0 and 400 m, and rotational errors of approximately 0.01 deg/m for path lengths between 0 and 400 m. The GPS system used is the NovAtel SPAN-CPT [13,68], which combines a GNSS and IMU for higher accuracy. In the experiments of [66], the SPAN-CPT offers a positional accuracy (RMS) of between 0.6 and 1.2 m, velocity accuracy (RMS) of around 0.03 m/s, and heading accuracy (RMS) of 0.08 deg.

4.2.3 Next steps

This paper demonstrates the viability of radar-only odometry to match the performance of state-of-the-art systems. Although the proposed pipeline performs well under a variety of conditions, it struggles in less structured environments. All experiments shown in this paper take place in Oxford, UK, a relatively ur-

ban setting with many stable and diverse features. Even so, there are still cases in Oxford for which the version of radar odometry presented in this chapter is unsatisfactory, as shown in Figure 8 in [66]. When it is tested in unstructured environments, like a forest or in the countryside, performance deteriorates further. Upon closer investigation, we observe that the main problems are that the keypoints in this paper are not well suited for unstructured environments. Due to the poor keypoints and chosen feature descriptor, the number of good unary candidate matches that are proposed during data association is so low that the algorithm has trouble distinguishing between true matches and outliers.

These observations motivated the need for designing a revised system that works well across various settings. Our goal for the subsequent work was to develop a radar-only odometry system for diverse environments that improves reliability, increases accuracy, and requires minimal human intervention compared to the paper given in this chapter.

Chapter 5

Radar-only odometry: versatility and reliability across settings

We now present our second paper on this work, which will be published in the Proceedings of the 2019 IEEE International Conference on Robotics and Automation. It improves upon [66], which is included in Chapter 4, and demonstrates accurate ego-motion estimation across diverse settings.

Radar-only ego-motion estimation in difficult settings via graph matching

Sarah H. Cen and Paul Newman

***Abstract*— Radar detects stable, long-range objects under variable weather and lighting conditions, making it a reliable and versatile sensor well suited for ego-motion estimation. In this work, we propose a radar-only odometry pipeline that is highly robust to radar artifacts (e.g., speckle noise and false positives) and requires only one input parameter. We demonstrate its ability to adapt across diverse settings, from urban UK to off-road Iceland, achieving a scan matching accuracy of approximately 5.20 cm and 0.0929 deg when using GPS as ground truth (compared to visual odometry's 5.77 cm and 0.1032 deg). We present algorithms for keypoint extraction and data association, framing the latter as a graph matching optimization problem, and provide an in-depth system analysis.**

I. Introduction

Despite the rapid expansion and advancement of research into sensor systems for mobile autonomy, those used for intelligent transportation typically feature the same suite: vision, lidar, GPS, and proprioceptive sensors. Recent interest in radar for mobile autonomy [1], [2] has uncovered its potential as a highly robust, multipurpose sensor. Unlike vision and lidar systems, it is resilient to variable lighting and weather, and, unlike GPS, it functions both indoors and outdoors. Radar is reliable at short and long ranges, and it is becoming increasingly more compact and affordable. In addition to its practical benefits, radar returns are simultaneously data-efficient and information-rich: radar can observe multiple objects per transmission and detect their locations, velocities, and cross-section characteristics.

As a result, radar is a highly versatile sensor that can withstand adverse conditions. Its ability to detect stable, long-range features makes it particularly suitable for odometry and localization. In this paper, we build on our previous work [2], in which we demonstrate precise ego-motion estimation using only a millimeter-wave (MMW) frequency-modulated continuous-wave (FMCW) scanning radar. As shown in [2], our system performs comparably to state-of-the-art visual odometry (VO) and GPS. Even under conditions that cause other sensors to fail, radar odometry (RO) is robust.

Common approaches to odometry for rangefinders have several key drawbacks, such as sensitivity to noise, reliance on other sensors, requiring *a priori* knowledge, and utilizing restrictive models. In contrast, our radar-only system achieves precise odometry without model-reliant motion filters, outlier detection, or map creation. At its core is a geometrically-based, non-iterative scan matching algorithm that is displacement-independent and approximately optimizes a global objective function. Consequently, our pipeline works without a prior on the temporal or spatial relationship between keypoint sets and can handle significant measurement noise and missing detections, as shown in Fig. 1.

S. H. Cen is with the Department of Electrical Engineering and Computer Science, Massachussetts Institute of Technology, Cambridge 02139, USA (shcen@mit.edu). P. Newman is with the Oxford Robotics Institute, University of Oxford, Oxford OX1 3PJ, UK (pnewman@robots.ox.ac.uk).

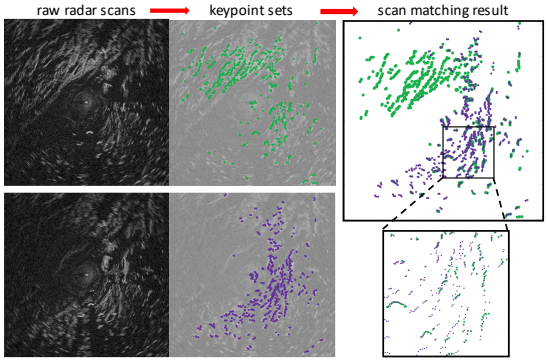


Fig. 1: Demonstration of our one-parameter, radar-only odometry pipeline performing scan matching in a challenging scenario with no motion prior. On the left are two consecutive radar scans, which are 250×250 m and taken 0.25 seconds apart while driving over uneven open terrain in Iceland. The keypoints extracted by our algorithm are shown in green and purple, and the alignment proposed by our data association algorithm is visualized on the right. Precise scan matching is achieved despite radar's high levels of measurement noise as well as significant appearance changes.

In this paper, we present a revised system that features the same advantages enumerated above with several major improvements to [2]. Specifically, our main contributions are:

- A new gradient-based, one-parameter keypoint extraction algorithm that functions well in diverse settings.
- More efficient and robust data association.
- Successful real-time odometry in varied and challenging settings (e.g., off-road path in Iceland boulder field).
- In-depth theoretical analysis on system complexity and estimation uncertainty.

Following a literature review in Section II, we present and analyze our pipeline in Sections III and IV. We provide experimental results and conclude in Sections V and VI. Both our works are presented and analyzed in detail in [3].

II. Related works

While visual [4], [5], lidar [6], [7], and wheel [8] odometry are well studied, radar odometry remains challenging. Due to its wide spreadbeam and long range, radar has lower resolution than lidar and is highly susceptible to interference from clutter, which generates speckle noise. Radar scans also contain false positives from multipath reflections and receiver saturation. As a result, radar odometry must be robust to measurement noise and false detections, and it must demonstrate high precision despite low-resolution data and slow update speeds. Odometry methods for radar can

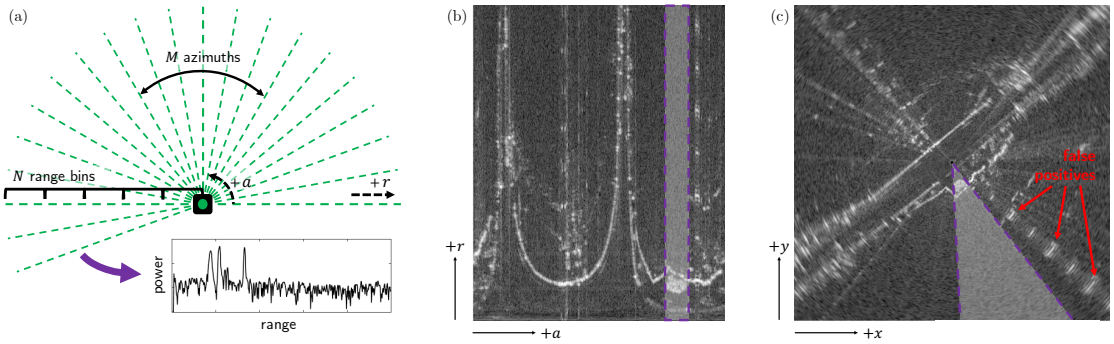


Fig. 2: (a) Visualization of scanning radar operation from a bird's-eye view. The radar (green dot) sequentially observes M azimuths (dotted lines), collecting readings for N range bins along r. For each azimuth a, it outputs a power-range spectrum, an example of which is given on the bottom right. (b) The polar radar scan formed by combining the spectra of a full rotation of azimuths, as observed in Oxford city center, UK. (c) The corresponding Cartesian scan with equivalent regions in (b) and (c) highlighted. Note the high levels of speckle noise and false positives due to saturation.

be categorized as indirect or direct. Indirect methods first extract salient keypoints, then associate those that correspond to the same location. Direct methods [9]–[11], which forego keypoint extraction and operate on minimally pre-processed sensor outputs, are discussed in [2], [3] and not in this paper. All methods assume the majority of observed objects are static.

The first step of indirect methods is keypoint extraction, for which the most popular approach is constant false-alarm rate (CFAR) detection [12], which distinguishes peaks from noise using sliding-window thresholding. CFAR and its variants generally require at least three tunable parameters, which are based on assumed noise characteristics and do not behave consistently across datasets (see [2] for comparison). Some works leverage the knowledge that coherent structures make good keypoints by clustering or detecting the edges of bright regions in scans [13]. Others elect to represent the surroundings using predetermined geometric primitives [14] or models, like the normal distribution transform (NDT) [15]. Vision-inspired works treat radar scans as images and extract features, like SIFT and FAST [16].

These keypoints must then undergo data association, also known as scan matching [1], [14] in robotics. The most common technique is iterative closest point (ICP) [17], which iteratively matches points using naive methods until the alignment between keypoint sets is sufficiently close [14], [18]. ICP relies on a good estimate of the relative displacement (i.e., motion prior) between scans. Other data association techniques search for motion parameters that optimize some objective function, like maximizing similarity (e.g., overlap of Gaussian distributions for NDT [19]) or minimizing distance (e.g., cluster edge difference [9]) between keypoint sets. Two further examples of objective functions characterize map quality [20] and radar scan distortion [21] in terms of motion. Feature-based approaches associate keypoints using descriptors, like BASD [22] and SURF [16].

Many of the methods discussed do not generalize well due to the high levels of noise in radar scans. Adequate performance often requires fine tuning, *a priori* knowledge, restrictive assumptions, or outlier detection. Several of the works rely heavily on other sensors for robustness, which compromises performance under conditions that cause these sensors to fail, or use simultaneous localization and mapping (SLAM), which is accompanied by overhead costs and model-reliant motion filters [16], [23].

We utilize an indirect method to explicitly select salient information from noisy artifacts. We present two algorithms that improve on our previous work [2]. The new pipeline is more efficient and easily adaptable across diverse settings. Accordingly, our keypoint extraction algorithm returns interpretable features with minimal redundancy, only one input parameter, and no assumptions about the scene structure or noise. Our data association algorithm does not need a motion prior or parameter tuning, and it is robust to large amounts of noise and false detections. An in-depth literature review and description of our contributions can be found in [3].

III. Radar-only ego-motion estimation

A. FMCW scanning radar

FMCW radar, which is becoming more compact, safe, and affordable than alternative radars, collects long-range measurements with high accuracy and remains resilient under variable lighting and weather [3]. Scanning radar, as shown in Fig. 2, sequentially observes narrow angular regions as it rotates, allowing it to locate an object that falls inside the transmitted beam by its range and azimuth. Received power depends on object reflectivity, surface area, orientation, and material. A wide spreadbeam in elevation and long wavelength allow multiple objects to be detected per transmission.

For each azimuth, radar outputs a one-dimensional signal, termed the power-range spectrum, which encodes the power reflected by the scatterers within the beam at each range. After a full rotation, radar returns a two-dimensional scan, as shown in Fig. 2. Radar scans are both information-rich and data-efficient, but they contain several unwanted artifacts visible in Fig. 2, including noise and false detections [3]. Another consideration of radar is its lower resolution and slower measurement update speeds compared to lidar.

Algorithm 1: Keypoint Extraction

Input: Radar scan $S \in \mathbb{R}^{m \times n}$
Output: Set of keypoints $L(S) \in \mathbb{R}^{p \times 2}$
Parameters: Maximum number of keypoints $\ell_{\max}$

1 $G \leftarrow$ computeNormalizedGradientMagnitude(S)
2 $S' \leftarrow S - \text{mean}(S)$
3 $H \leftarrow (1 - G) \times S'$
4 $I \leftarrow$ getIndicesOfElementsInDescendingOrder(H)
5 Initialize marked matrix $R \leftarrow [\textit{false}]^{m \times n}$, counter $\leftarrow \ell$
6 **while** $(\ell < \ell_{\max})$ and (any *false* in R) **do**
7 $(a, r) \leftarrow$ getNextIndices(I)
8 **if** not $R(a, r)$ **then**
9 $(r_{\text{low}}, r_{\text{high}}) \leftarrow$ findRangeBoundaries$(S'[a, :])$
10 **if** none in $R[a, r_{\text{low}} : r_{\text{high}}]$ **then**
11 Increment ℓ
12 $R[a, r_{\text{low}} : r_{\text{high}}] \leftarrow \textit{true}$
13 **for** a from 1 to m **do**
14 **for** each marked region Q in $R[a, :]$ **do**
15 **if** any in $R[a-1, Q] \cup R[a+1, Q]$ **then**
16 $L \leftarrow L \cup (a, r)$ for $H[a, r] = \max(H[a, Q])$

B. Keypoint extraction

In this section, we present a keypoint extraction algorithm that is designed to:

- Return interpretable and meaningful keypoints.
- Be straightforward to use and require only one input parameter specifying the maximum number of keypoints.
- Adapt to diverse settings without *a priori* knowledge, like the expected noise characteristics.
- Avoid redundant returns to improve data efficiency.
- Ignore bright swathes due to rolling, pitching, or clutter.

The procedure is given in Algorithm 1. It takes as an input the raw radar scan, which contains the power readings for m azimuths and n range bins, and the desired maximum number of keypoints $\ell_{\max}$. The radar scan is re-scaled to favor areas of high intensity and low gradients (using the Prewitt operator), storing the result in H (lines 1-3). The algorithm then iterates through points (a, r) in order of highest to lowest $H(a, r)$ and marking off adjacent regions in R that should be ignored in future iterations. These regions are identified by searching S' along a for the closest range indices $(r_{\text{low}}, r_{\text{high}})$ below and above r with values less than 0 (lines 4-12). After $\ell_{\max}$ regions are marked, iterate through a's. For each continuous marked region in a, the range bin r with the highest value is added to keypoint set L. Isolated detections, which are more likely to be speckle noise, are eliminated when a marked region has no adjacent neighbors in azimuth (lines 13-16).

The algorithm returns meaningful keypoints by seeking peaks in the scan (areas of high intensity and low gradient). It minimizes redundancy by returning only one keypoint per continuously marked sequence of range bins. This strategy also prevents large bright regions due to rolling, pitching, or clutter (RPC) from producing many keypoints that can distract from important keypoints corresponding to stable objects. Furthermore, by minimizing the amount of prior knowledge required and seeking coherent structures irrespective of noise, the algorithm adapts to diverse settings with varying levels of noise. The only input parameter is an upper bound on the number of desired keypoints.

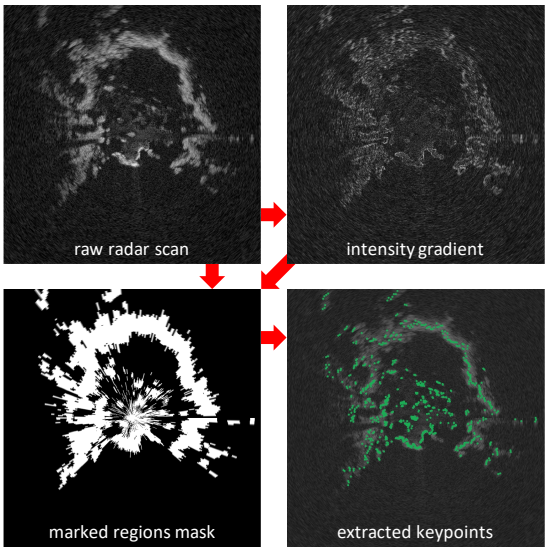


Fig. 3: Visualization of our keypoint extraction approach operating in a forest in Iceland. From the 250×250 m radar scan (top left) and its intensity gradient (top right), important regions in the scan are identified. Those marked (bottom left), the number of which is capped by the only tunable parameter $\ell_{\max}$, are used to produce a keypoint set (bottom right) that captures the scene structure while minimizing redundant keypoints.

In [3], we discuss how CFAR is not ideal for our objectives. It requires multiple tunable parameters, provides redundant returns, and often struggles with bright patches caused by RPC. In comparison to our previous work [2], our new algorithm has fewer parameters and operates well even in unstructured environments, like forests.

C. Data association

In this section, we present an improved data association algorithm, which is revised from [2] and designed to:

- Have zero tunable parameters.
- Work without prior knowledge on the relative displacement (translational or rotational) between scans.
- Handle high levels of measurement noise, false (or ghost) objects, and other appearance changes.
- Utilize global information across the scans to generate correspondences that are mutually consistent.

Our data association algorithm is an example of graph matching [24]. It seeks point correspondences based on global geometric constraints. Unlike ICP, which relies on the approximate alignment of scans, our method does not need prior knowledge of the relative displacement between scans. With no parameter tuning, it is easy to apply across datasets. Moreover, by formulating data association as a global opti-

Algorithm 2: Data Association

Input: Keypoint sets L_1 and L_2, where $|L_1| \leq |L_2|$
Output: Selected matches $M(L_1, L_2)$
Parameters: Azimuth and range resolutions α and ρ, respectively, of the radar being used

1 $U^{u\times 2} \leftarrow \text{unaryMatchesFromDescriptors}(L_1, L_2, \alpha, \rho)$
2 $C^{u\times u} \leftarrow \text{pairwiseCompatibilityScores}(U, L_1, L_2)$
3 $\mathbf{v}^* \leftarrow \text{principalEigenvector}(C)$
4 Initialize empty set $\mathcal{M}$ and $\hat{\mathbf{m}} \leftarrow [0]^{u\times 1}$
5 Initialize *unsearched* $\leftarrow \{1, 2, \ldots, u\}$ and *score* $\leftarrow 0$
6 **while** *unsearched* is not empty **do**
7 $\quad \hat{m}_g \leftarrow 1$ given g s.t. $v_g^{*2} \geq v_h^{*2}\ \forall\, h \in$ *unsearched*
8 $\quad$ Terminate if $\hat{\mathbf{m}}^\top C \hat{\mathbf{m}} <$ *score*
9 $\quad$ Add the match $U[g, :]$ to $\mathcal{M}$
10 $\quad$ Remove values h from *unsearched* if $(U[g, 1] = U[h, 1] \cup U[g, 2] = U[h, 2])$

mization problem, our approach utilizes information across the entire scan and naturally removes outliers. Compared to [2], our current method is more efficient and robust due to changes to the unary keypoint descriptor and termination condition. As given in Algorithm 2, our approach takes in two keypoint sets L_1 and L_2 and proceeds in three stages.

Stage 1: Without loss of generality, assume $u = |L_1| \leq |L_2|$. A unary candidate proposal method (see [2]) matches each keypoint $i_1 \in L_1$ to another $i_2 \in L_2$, and the pair is added to U (line 1). In this paper, we present a new rotation-invariant descriptor $d^i \in [0, 1]^{\alpha+\rho}$ for keypoint i. To compute $d[1{:}\alpha]$, we create a histogram of the number of elements in each angular slice around keypoint i, take its fast Fourier transform, then normalize its phase. To compute $d[\alpha+1{:}\alpha+\rho]$, we create a histogram of the number of elements in each annulus around keypoint i, then normalize it. The angular slices and annuli have the same spatial resolution as the radar scan. Note that, as done in [2], each element contributing to the descriptor is weighted by its range relative to the radar to correct for radar's range-density bias.

Stage 2: To refine the match proposals in U, each pairwise combination of keypoints $g = (i_1, j_1) \in L_1$ is compared to its associated (according to U) match pair $h = (i_2, j_2) \in L_2$. If i_1 and i_2 do indeed represent identical keypoints in the environment and the same holds true for j_1 and j_2, then the pairwise relationships g and h must be similar, which is quantified by the compatibility score $C_{gh} \in [0, 1]$ of the non-negative symmetric compatibility matrix C (line 2). The formulation of this score is given in [2].

Stage 3: Maximizing the global compatibility score is done by finding the optimal match vector $\mathbf{m}^*$ such that:

$$\mathbf{m}^* = \underset{\mathbf{m}\in\{0,1\}^{u\times 1}}{\arg\max} \left(\frac{\mathbf{m}^\top C \mathbf{m}}{\mathbf{m}^\top \mathbf{m}} \right). \tag{1}$$

Logically, $\mathbf{m}$ must satisfy the (i) integrality constraint: all elements are either 0 or 1 (i.e., a match is true or false) and (ii) uniqueness constraint (i.e., each keypoint in L_1 cannot maps to more than one keypoint in L_2, and vice versa). We

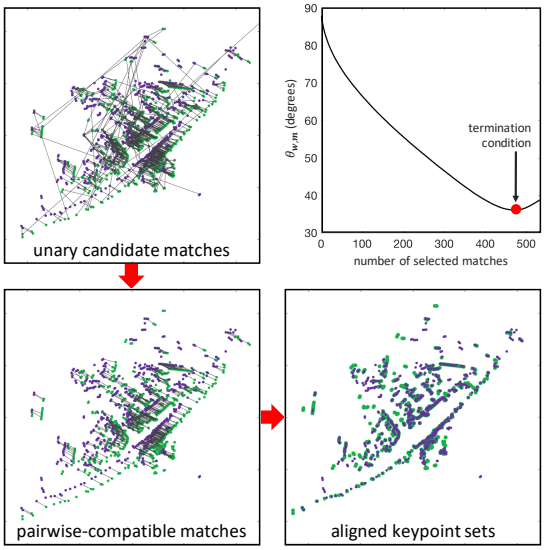


Fig. 4: Visualization of our data association approach operating in Warwick, UK. It first proposes unary matches using feature descriptors (top left), then uses pairwise relationships to identify mutually consistent matches (bottom left). Our method approximates the solution to an NP-hard problem using a greedy algorithm, for which the termination condition is visualized on the top right. The final alignment (bottom right) shows the robustness to noise.

define $\mathbf{m}$ such that $m_k = 1$ if $U_k = (i_1, i_2) \in M$ and $m_k = 0$ otherwise, and let $\mathcal{M} : L_1 \rightarrow L_2$ be the mapping between keypoint sets according to M. This formulation is known as graph matching [24], and it is NP-hard, so we relax the constraints and solve instead for $\mathbf{v}^*$ such that:

$$\mathbf{v}^* = \underset{\mathbf{v}\in\mathbb{R}^{u\times 1}}{\arg\max} \left(\frac{\mathbf{v}^\top C \mathbf{v}}{\mathbf{v}^\top \mathbf{v}} \right). \tag{2}$$

By the Raleigh's ratio and Perron-Frobenius theorems [24], the solution $\mathbf{v}^* \in [0, 1]^{u\times 1}$ is the principal eigenvector of C. We use $\mathbf{v}^*$ to approximate the optimal constrained solution to (1) via the greedy approach in lines 4-10. We present a new termination condition that stops adding matches when the new one is incompatible with those already in M, as explained in Section IV-B. Compared to our previous work [2], the new condition is more robust and consistent because it does not depend on the magnitude of elements in $\mathbf{v}^*$.

The rigid-body motion (RBM) is found using a well-known technique [25]. As radar scans are in 2D, ego-motion is restricted to 3-DOF: planar translation and rotation.

IV. System Analysis

A. Algorithmic complexity

The complexity of our system is summarized in Table I. For most radars, $A < R$. To estimate the average complexity, it is assumed that $A < \ell_{\max} < R$. The second row gives the complexity for data association with no prior on the relative motion between scans. When running the system in real time, because the relative motion between consecutive time steps is limited, a mild prior can be used to improve computational

efficiency by reducing the number of unary descriptor comparisons. For this analysis, the intensity gradient, ordering, and principal eigenvector are computed using the Prewitt operator, merge sort, and power iteration, respectively.

Overall, the system computation is generally dominated by the unary candidate matching step in *Stage 1* of data association, such that the average system complexity is given by $\Theta(\ell_{\text{max}}^2(A+R)) \sim \Theta(\ell_{\text{max}}^3)$. Because the bottleneck depends on ℓ_{max}, we intentionally allow ℓ_{max} to be set by the user as the *only* tunable input parameter of our system.

B. Theoretical analysis and performance measures

The fundamental principle behind our system is the optimization problem (1). Intuitively, it seeks the set of matches that maximizes the global compatibility score $\mathcal{G}(\mathbf{x}, C) = \frac{\mathbf{x}^\top C \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$ in (1). Because this problem is NP-hard, our algorithm approximates the optimal solution. This section investigates how well solution M approximates M^*. Let $C = C^* + E$, where $C^*_{ij} = C_{ij}$ if and only if $U_i, U_j \in M^*$, and $C^*_{ij} = 0$ otherwise. C^* is the corrected C in which $k \notin M^*$ does not inflate the compatibility scores of $i \in M^*$, and E can be considered a perturbation from C^* due to incorrect matches.

Proposition 1: $\mathcal{G}(\mathbf{m}, C)$ always overestimates the true global compatibility score, except when all proposed matches are correct (i.e., $U \in M^*$), in which case they are equal.

Proof: By definition of C and C^*, E is non-negative: $\mathbf{x}^\top E \mathbf{x} \geq 0$ for any $\mathbf{x} \geq 0 \Rightarrow \mathbf{m}^\top C \mathbf{m} = \mathbf{m}^\top C^* \mathbf{m} + \mathbf{m}^\top E \mathbf{m} \geq \mathbf{m}^\top C^* \mathbf{m}$, or $\mathcal{G}(\mathbf{m}, C) \geq \mathcal{G}(\mathbf{m}, C^*)$. ■

Proposition 1 implies that match selection is overly optimistic, and it is more likely to add too many matches, some of which are bad, than to not add good ones. Therefore, when using $\mathbf{v}_1$, we seek to: (1) resolve conflicting proposed matches in order to satisfy the constraints on (1) by selecting the best of them, and (2) avoid adding poor matches. Our algorithm achieves the first goal by adding matches in order of largest corresponding value in $\mathbf{v}_1$. It achieves the second with the termination condition, explained next. Let $\overline{C}$ be the compatibility matrix assuming perfect measurements.

Proposition 2: Assuming no perfect symmetry exists in the keypoint sets, $\mathbf{m}^* = \arg\max_{\mathbf{x} \in [0,1]^u} \mathcal{G}(\mathbf{x}, \overline{C}^*)$.

Proof: By the definition of compatibility scores [2], the maximum score $\overline{C}^*_{ij} = 1$ if $i, j \in M^*$, and $0 \leq \overline{C}^*_{ij} < 1$ otherwise. Since we assume perfect measurements and no symmetries, to meet the uniqueness constraint in (1), $m^*_i \neq 1 \Rightarrow \overline{C}^*_{ij} \neq 1 \forall j \in \{1, \ldots, u\}$. By the integral constraint, $m^*_i \neq 1 \Rightarrow m^*_i = 0$, and $m^*_i = 1$ if $i \in M^*$. ■

By Proposition 2, given the true compatibility matrix $\overline{C}^*$ and optimal match set M^*, the solution is, in fact, the binary vector $\mathbf{m}^*$ even though no integral constraints are placed on $\mathbf{x}$. Since all accepted matches in $\mathbf{m}^*$ are equally weighted, they are equally valued and agree with one another. We use this fact to seek a solution M that, like M^*, contains matches that are mutually consistent. By Proposition 1, the global compatibility score is an optimistic measure (i.e., $\mathcal{G}(\mathbf{m^1}, C) > \mathcal{G}(\mathbf{m^2}, C)$ does not imply that matches in M^1 are better than those in M^2). We use Proposition 2 to construct a more reliable measure, as follows.

TABLE I: Algorithmic time complexity

	Average	Worst
KE[1]	$\Theta(AR \log(AR) + \ell_{\text{max}})$	$O(AR \log(AR) + \ell_{\text{max}} R)$
DA	$\Theta(\ell^2_{\text{max}}(A+R))$	$O(\ell^2_{\text{max}}(A + R + \log(\ell_{\text{max}})))$
DA[2]	$\Theta(\ell^2_{\text{max}} \log(\ell_{\text{max}}))$	$O(\ell^2_{\text{max}} \log(\ell_{\text{max}}))$
ME[3]	$\Theta(\ell^2_{\text{max}})$	$O(\ell^3_{\text{max}})$

[1] Keypoint extraction
[1] Data association with mild prior on relative motion
[2] Rigid-body motion estimation

TABLE II: Datasets and their characteristics

Dataset	Setting[1]	Avg. T. (m)[2]	Avg. R. (deg)[2]	Dist. (km)	RO Prior[3]
City	Oxford, UK	1.268	1.489	8.36	None
Backstreets	Oxford, UK	1.252	1.804	2.76	None
Highway	M40, UK	5.164	1.076	19.6	Max. Acc.
Countryside	Rural UK	2.297	1.264	16.8	Max. Acc.
Forest	Heidmörk, IS	3.168	0.769	9.37	None
Rocks	Thórsmörk, IS	1.267	1.036	2.35	Max. Acc.

[1] UK and IS stand for the United Kingdom and Iceland, respectively.
[2] These quantities are the average movement (in translation and rotation according to RO) between consecutive scans when the vehicle is not stationary. Scans are obtained at ∼4 Hz (i.e., speed is ∼4 times displacement).
[3] "None" means that no motion prior is used during scan matching. "Max. Acc." employs a very mild prior that limits the maximum distance the vehicle could have traveled based a maximum acceleration of 8 m/s^2.

Measure 1: The index $\mathcal{I}_M = \cos(\theta_{\mathbf{w},\mathbf{m}}) \in [0, 1]$ measures the mutual compatibility of selected matches M, where $\theta_{\mathbf{w},\mathbf{m}}$ is the angle between vectors $\mathbf{w} = C(\mathbf{m} \odot \mathbf{v}_1)$ and $\mathbf{m}$.

Using this property, we terminate the greedy algorithm when the next match would cause $\mathcal{I}_\mathcal{M}$ to decreases. Intuitively, the new match would cause $\mathbf{w}$ to move farther from its corresponding binary solution, where this separation is inversely related to the consistency between matches. Although similar to $\cos(\theta_{\mathbf{m},C\mathbf{m}}) = \mathcal{G}_{\mathbf{m},C}$, the continuous values $v_{1,i}$ in $\mathcal{I}_\mathcal{M}$ scale match i's contribution by its compatibility.

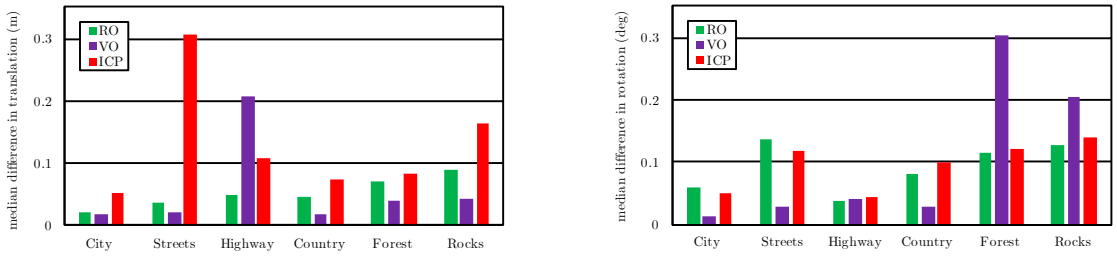
Measure 2: Assume the chosen solution is the optimal one: $M = M^*$. Let λ^*_i represent the i-th largest eigenvalue of C^*. Then, the normalized eigengap $\mathcal{I}_{\text{eg}} = (\lambda^*_1 - \lambda^*_2)/u \in [0, 1]$ measures the solution's robustness to perturbations.

While $\mathcal{I}_M$ is introspective (quantifies how well the selected matches agree), the second measure $\mathcal{I}_{\text{eg}}$ considers all proposed matches U and quantifies the system's confidence it its solution M^* over the next-highest scoring solution.

V. Results

In [2], we demonstrate RO that performs comparably to VO and GPS (under ideal conditions for the latter two) and maintains precise odometry even under conditions for which VO and GPS fail. In this section, we build on this result and show that our new RO pipeline, which improves upon [2], achieves high accuracy in real time across diverse settings, from UK city streets to off-road landscapes in Iceland.

For keypoint extraction, we use $\ell_{\text{max}} = 1000$. The six experimental datasets examined in this paper are described in Table II. As explained in the footnotes, the final column indicates whether a motion prior is used for RO. For three datasets, no motion prior is used at all. The other three employ a very mild motion prior that uses the system's



(a) (b)

Fig. 5: Evaluation of odometry performance across six diverse settings, measured by the median difference between each method's estimated displacements relative to those of GPS in translation (a) and rotation (b). Displacement is the relative motion between consecutive scans obtained at ∼4 Hz. For each setting, RO (our pipeline), VO, and ICP are compared to GPS. Both RO and ICP use the keypoint sets returned by our KE algorithm.

TABLE III: Comparative summary of odometry methods

	Translation (m)[1]		Rotation (deg)[1]	
Comparison	Median	Std. Dev.	Median	Std. Dev.
RO to GPS	0.0520	0.0660	0.0929	0.1632
RO to VO	0.0724	0.0777	0.1414	0.2105
VO to GPS[2]	0.0577	0.0546	0.1032	0.1396

[1] Statistics for difference between estimated displacement (in translation and rotation) between consecutive radar scans, which are obtained at ∼4 Hz.
[2] Contextualizes RO performance by comparing two well accepted methods.

TABLE IV: Summary of RO scan matching performance relative to GPS

	Translation (m)		Rotation (deg)	
Setting	Median	Std. Dev.	Median	Std. Dev.
City	0.0208	0.0318	0.0597	0.1442
Backstreets	0.0362	0.0398	0.1375	0.2238
Highway	0.0480	0.0598	0.0384	0.0673
Countryside	0.0462	0.0698	0.0811	0.1380
Forest	0.0711	0.0899	0.1141	0.1796
Rocks	0.0897	0.1050	0.1267	0.2267

maximum possible acceleration of 8 m/s^2 to remove impossible matches from consideration. Empirically, our system very rarely requires this prior (e.g., 2 scan matches out of 10,000). Incorrect alignments occur when the scene structure contains high levels of symmetry, resulting in many possible alignment solutions. The lenient prior serves to reduce the number of possible solutions. Due to space constraints, odometry plots for this work can be found in our video [26].

Fig. 5 evaluates RO performance against GPS and compares it to state-of-the-art VO [27] and popular scan matching method ICP. For ICP, we use a convergence tolerance of 1e-5. We improve ICP performance by providing it with a good motion prior and limiting the possible matches to nearest neighbors within 2 m. Even so, RO outperforms ICP in every setting, as shown in Fig. 5(a). The occasional lower rotational error of ICP in Fig. 5(b) is misleading because it only occurs when ICP's estimated translation is highly inaccurate, in which case the rotation error is meaningless. Relative to VO, RO generally achieves comparable though slightly lower accuracy, which can be explained by the fact that RO only captures 3-DOF motion compared to VO's 6-DOF. Thus, if the platform experiences rolling or pitching, RO registers this motion into its 3-DOF estimate. Similarly, if the environment is less structured, variability in elevation is discarded in radar's 2D scan, also affecting the resulting motion estimates. This reasoning is consistent with RO's slightly higher error in countryside, forest, and boulder field (i.e., rocks) datasets. RO greatly outperforms VO on the highway, on which the average travel speed is highest, making radar odometry appealing from a safety perspective.

A summary of RO performance relative to GPS is given in Table IV, and a comparison of the three odometry methods using results averaged across all datasets is given in Table III. Table IV shows that our RO pipeline achieves high accuracy, differing from GPS by only 5.2 cm and 0.09 deg in translation and rotation, respectively, on average. Notably, the third row of Table III shows that the overall median difference between VO and GPS exceeds that between RO and GPS. On the other hand, the standard deviation of differences is larger for RO, meaning that RO provides highly precise yet slightly noisier motion estimates compared to VO. Importantly, as detailed in Table IV, RO is remarkably accurate in common settings (e.g., city center). Although the error seemingly increases for more complex environments (e.g., in Iceland), these discrepancies may be due to factors discussed in the previous paragraph.

This work achieves lower error in Oxford city compared to our previous work [2]. Though not shown, it consistently outperforms [2], which cannot handle unstructured settings.

VI. Conclusion and Future Work

In this paper, we motivate the use of radar as an information-rich, long-range, and reliable sensor that is robust under challenging conditions. We propose keypoint extraction and scan matching algorithms that are designed to require minimal *a priori* knowledge and human intervention. We demonstrate our pipeline's high accuracy across diverse settings, performing comparably and at times better than VO. We discuss our approach's theoretical underpinnings, propose two performance measures, and analyze system complexity. Three areas of promising future work are: streamlining unary candidate proposal, addressing scene symmetry (e.g., using tertiary graph matching or quantifying motion observability), and deducing 3D information from 2D scans.

REFERENCES

[1] M. Adams and E. Jose, *Robotic navigation and mapping with radar*. Artech House, 2012.

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

[3] S. H. Cen, "Ego-motion estimation and localization with millimeter-wave scanning radar," Ph.D. dissertation, University of Oxford, Oxford, United Kingdom, 2019.

[4] M. O. Aqel, M. H. Marhaban, M. I. Saripan, and N. B. Ismail, "Review of visual odometry: types, approaches, challenges, and applications," *SpringerPlus*, vol. 5, no. 1, p. 1897, 2016.

[5] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE robotics & automation magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[6] J. Zhang and S. Singh, "Low-drift and real-time lidar odometry and mapping," *Autonomous Robots*, vol. 41, no. 2, pp. 401–416, 2017.

[7] R. W. Wolcott and R. M. Eustice, "Robust lidar localization using multiresolution gaussian mixture maps for autonomous driving," *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 292–319, 2017.

[8] J. Borenstein, H. Everett, L. Feng *et al.*, "Where am i? sensors and methods for mobile robot positioning," *University of Michigan*, vol. 119, no. 120, p. 27, 1996.

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

[10] 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.

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

[12] M. Skolnik, *Radar Handbook, Third Edition*, ser. Electronics electrical engineering. McGraw-Hill Education, 2008. [Online]. Available: https://books.google.co.uk/books?id=76uF2Xebm-gC

[13] K. Werber, J. Klappstein, J. Dickmann, and C. Waldschmidt, "Interesting areas in radar gridmaps for vehicle self-localization," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on*. IEEE, 2016, pp. 1–4.

[14] 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.

[15] M. Rapp, M. Barjenbruch, M. Hahn, J. Dickmann, and K. Dietmayer, "Clustering improved grid map registration using the normal distribution transform," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*. IEEE, 2015, pp. 249–254.

[16] J. Callmer, D. Törnqvist, F. Gustafsson, H. Svensson, and P. Carlbom, "Radar slam using visual features," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 71, 2011.

[17] P. J. Besl, N. D. McKay *et al.*, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[18] M. Yang, B. Dong, H. Wang, and B. Zhang, "Laser radar based real-time ego-motion estimation for intelligent vehicles," in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 1. IEEE, 2002, pp. 44–51.

[19] M. Rapp, M. Barjenbruch, M. Hahn, J. Dickmann, and K. Dietmayer, "Probabilistic ego-motion estimation using multiple automotive radar sensors," *Robotics and Autonomous Systems*, vol. 89, pp. 136–146, 2017.

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

[21] D. Vivet, P. Checchin, and R. Chapuis, "Localization and mapping using only a rotating fmcw radar sensor," *Sensors*, vol. 13, no. 4, pp. 4527–4552, 2013.

[22] F. Schuster, C. G. Keller, M. Rapp, M. Haueis, and C. Curio, "Landmark based radar slam using graph optimization," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 2559–2564.

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

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

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

[26] S. H. Cen and P. Newman, "Radar-only ego-motion estimation in difficult settings via graph matching," YouTube Spotlight Video for ICRA, 2019.

[27] W. Churchill, "Experience based navigation: Theory, practice and implementation," Ph.D. dissertation, University of Oxford, Oxford, United Kingdom, 2012.

5.2 Discussion

5.2.1 Summary of contributions

Similarly to [66], the aim of this work is to illustrate the capabilities of radar for mobile autonomy by testing a radar-only ego-motion estimation system in diverse settings. This paper shows that RO achieves performance comparable to other state-of-the-art systems despite difficult or adverse conditions, including driving at high speeds, through unstructured settings, and over uneven terrain.

This work has three main contributions. It introduces a new keypoint extraction algorithm that has only one input parameter (the maximum number of desired keypoints) and adapts well across various environments. It modifies the data association method proposed in [66] to make the algorithm more principled while also improving performance. Finally, the paper offers a discussion on the system's algorithmic complexity and a theoretical analysis of the data association module that includes performance measures.

As explained in Section 4.2, the odometry pipeline is designed to be straightforward and easy to use, thus minimizing reliance on human expertise. The keypoint extraction method seeks to supply the data association module with a concise representation of the environment that will be consistent from scan to scan. The data association module matches keypoints based on global geometric relationships in an efficient approximation of shape matching. The data association module naturally removes outliers, making it robust to unwanted artifacts in radar scans. Performance indices measuring the mutually compatibility of the chosen matches and the solution's robustness to perturbations affect the algorithm's termination

condition and indicate confidence in the solution, respectively.

The advantages and drawbacks of the keypoint extraction and data association algorithms are analyzed in-depth in Chapter 6.

Further explanation of this work and results can be found in our video [69].

5.2.2 Design choices

To avoid redundancy, explanations for most our design choices are not included below, as they are already discussed in Section 4.2.2. The following section describes the design choices that are relevant to our second work [70].

How was the keypoint extraction method formulated?

There are several methods commonly used in the literature for keypoint extraction or segmentation in radar scans. As further explained in Chapter 6, we deliberately design a new keypoint extraction method for several reasons, as follows:

- **Sparsity vs. Expressiveness**. The computational resources required by data association scales with the number of keypoints it needs to match, making it desirable to have as few keypoints as necessary to adequately describe the scene. However, the keypoint set must also be expressive enough to capture the most important information in the scan. To balance this trade-off, our keypoint extraction method first identifies salient areas in the scan. The central step comes next: the algorithm then expands the regions by adding bins either (1) because it believes that they correspond to the same object as a previously identified area or (2) to rule them out as uninteresting areas (i.e., mark them off). It finishes by iterating through each power-range

spectrum, returning one keypoint per area/region in each spectrum. In contrast to other region-based methods like CFAR, MSER, and DBSCAN, the marked regions are not necessarily indications of occupancy because they can also denote areas that are uninteresting. This approach allows the algorithm to efficiently search the scan by exploring the most salient areas first while simultaneously ruling out other areas as uninteresting.

- **Computational bottleneck**. As explained in [70], the system's bottleneck is the keypoint set size. With this knowledge, we design the only tunable parameter of the keypoint extraction method to be the maximum number of keypoints returned. Imposing a maximum means that the previous method in [66] no longer suffices. The new method adapts to maximum set size by balancing sparsity and expressiveness, as explained in the bullet point directly above.

- **Object correspondence**. There are a few reasons why radar keypoints should correspond to meaningful features, like objects. In particular, radar scans are noisy, so if keypoints are based on features in one scan that do not correspond to objects, it is likely that the same feature does not appear in a subsequent scan, making it difficult to perform feature association. Because objects (e.g., walls or trees) are more permanent, they are more consistent keypoints to use in data association. In addition, keypoints that correspond to objects can be used for other tasks, like object detection and tracking. In comparison, common image features, like SURF and FAST, search for corners, which are areas with intensity gradients. When applied to radar,

these features perform poorly because high gradients often results from noise and other unwanted artifacts rather than persistent objects, as shown in [66].

- **Spatial accuracy**. Unlike object classification or occupancy grid prediction, ego-motion estimation requires high spatial accuracy. As radar already has lower resolution than lidar and vision sensors, it is important that the spatial information that is provided by radar is properly utilized and not discarded. Many radar-based works pre-process the scans by smoothing in order to mitigate the effects of speckle noise. However, doing so also smooths away high-frequency information (e.g., the location of a wall can be shifted by a range bin due to an asymmetric peak). In contrast, the keypoint extraction in [70] does not smooth the original scan.

- **Noise resistant**. The identification and location of keypoints should be as minimally affected by noise as possible. As mentioned directly above, features borrowed from image processing do not respond well to speckle noise. The performance of other methods, like CFAR, depend on the environment and require (often unintuitive) tuning of parameters, as explained in our paper [66]. In comparison, our approach seeks to reduce the effects of noise and avoids the need to characterize or threshold the noise (which are drawbacks of [66] and CFAR).

- **Efficiency**. Keypoint extraction should be efficient. In our case, there is a trade-off between computational efficiency and perfect identification of keypoints (i.e., all keypoints correspond to real objects in the environment). For ego-motion estimation, keypoint sparsity, reliability, and spatial accuracy are

the main priorities. As such, we allocate the task of outlier detection to the data association module.

Why alter the feature descriptor?

As described in Chapter 4, one of the main issues with the pipeline proposed in [66] when tested in difficult settings is that the unary match being proposed can be bad. This step relies on the feature descriptor. The descriptor in the previous work [66] is not rotation-invariant and computes three statistical measures for each azimuthal slice. It is not clear that these statistical measures are particularly important or helpful. The descriptor introduced in this chapter [70] is both rotation-invariant and more straightforward, using simple histogram counts.

Why alter the data association termination condition?

The previous termination condition in [66] works well and is justified by its interpretation (i.e., the algorithm terminates when the value of the next match is lower than if all matches have equal value). However, the termination condition should depend on the keypoint sets being matched. For instance, if keypoint set A can be aligned with set B in two different ways (alignments 1 and 2) that both appear valid, then the scores for matches belonging to alignment 1 can be similar to those for alignment 2. In this case, if the scores for both alignments are high enough, the algorithm may begin adding matches from both alignments even though they are conflicting. Using this logic, the new termination condition encourages the algorithm to select the most promising of the two alignments and stop adding matches when subsequent matches conflict with the most promising alignment.

The mathematical reasoning behind the termination condition is given in [70].

How does performance compare to the previous work?

In comparison to [66], the odometry performance of [70] is better across the board. For instance, Figure 8 of [66] provides an example of a setting that causes the first pipeline difficulties. The pipeline presented in [70] has no difficulties in the described dataset (Jericho backstreets); the radar odometry curve is completely smooth. Comparing Table I in [66] to Table IV in [70] verifies these results. In addition, although the first pipeline [66] is not thoroughly tested in the non-urban environments that are examined in [70], preliminary experiments showed that it gives inconsistent performance outside of urban areas. Note, however, that the first pipeline is a radar-only system and could be fused with other sensors for better robustness.

What methods of ground truth were used to evaluate performance?

In this paper, RO was tested against two well known sensor systems: VO and GPS. The VO system used is described in [67]. For the Kitti Vision Benchmark Suite, this VO system generally achieves between 2 and 3 percent translational error for path lengths between 0 and 400 m, and rotational errors of approximately 0.01 deg/m for path lengths between 0 and 400 m. The GPS/INS system used in the Iceland datasets is the LORD MicroStrain 3DM-GX4-45 [71]. The system used in all other datasets is the NovAtel SPAN-CPT [13, 68], which combines a GNSS and IMU for higher accuracy. In the experiments of [70], the MicroStrain 3DM-GX4-45 offers a positional accuracy (RMS) of 2.0-2.5 m, velocity accuracy (RMS)

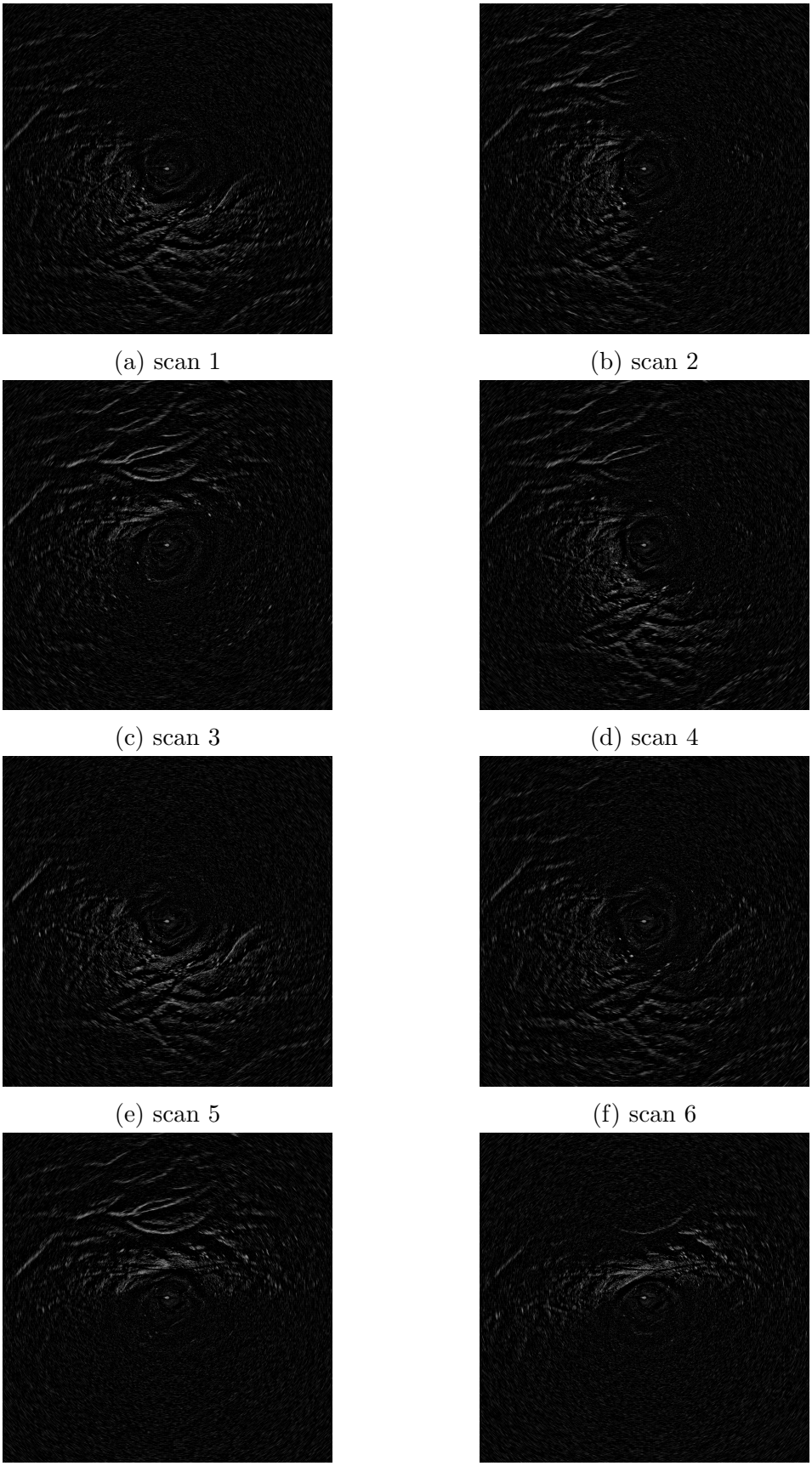
of around 0.1 m/s, and heading accuracy (RMS) of 0.8-2.0 deg. The SPAN-CPT offers a positional accuracy (RMS) of between 0.6 and 1.2 m, velocity accuracy (RMS) of around 0.03 m/s, and heading accuracy (RMS) of 0.08 deg.

5.3 Illustrative results

Due to space constraints, the paper presented in this chapter does not include many figures or plots illustrating our pipeline at work. For more visualizations of our results, we encourage the reader to watch our spotlight videos [65] and [69] for ICRA 2018 and 2019, respectively.

This section provides additional results to help visualize our pipeline and to provide evidence that [70] outperforms [66].

Figures 5.1 through 5.4 provide the outputs of intermediate steps in our pipeline for the most difficult setting tested in [70]: the Thórsmörk Boulder Field in Iceland. A sequence of eight consecutive radar scans are examined from the raw radar scans (transformed into Cartesian space) in Figure 5.1 to the extracted keypoints in Figure 5.2 to the selected matches from data association in Figure 5.3 and finally the product obtained when overlaying the keypoint sets according to the estimated transformation between the radar scans. Note that Figure 5.3 artificially displaces the keypoint sets in order for the reader to better see the matches (given by the black lines). Even so, because our method is displacement-independent (i.e., does not depend on the actual coordinates of the keypoint sets relative to one another), any such artificial displacement can be handled by our pipeline.



(a) scan 1

(b) scan 2

(c) scan 3

(d) scan 4

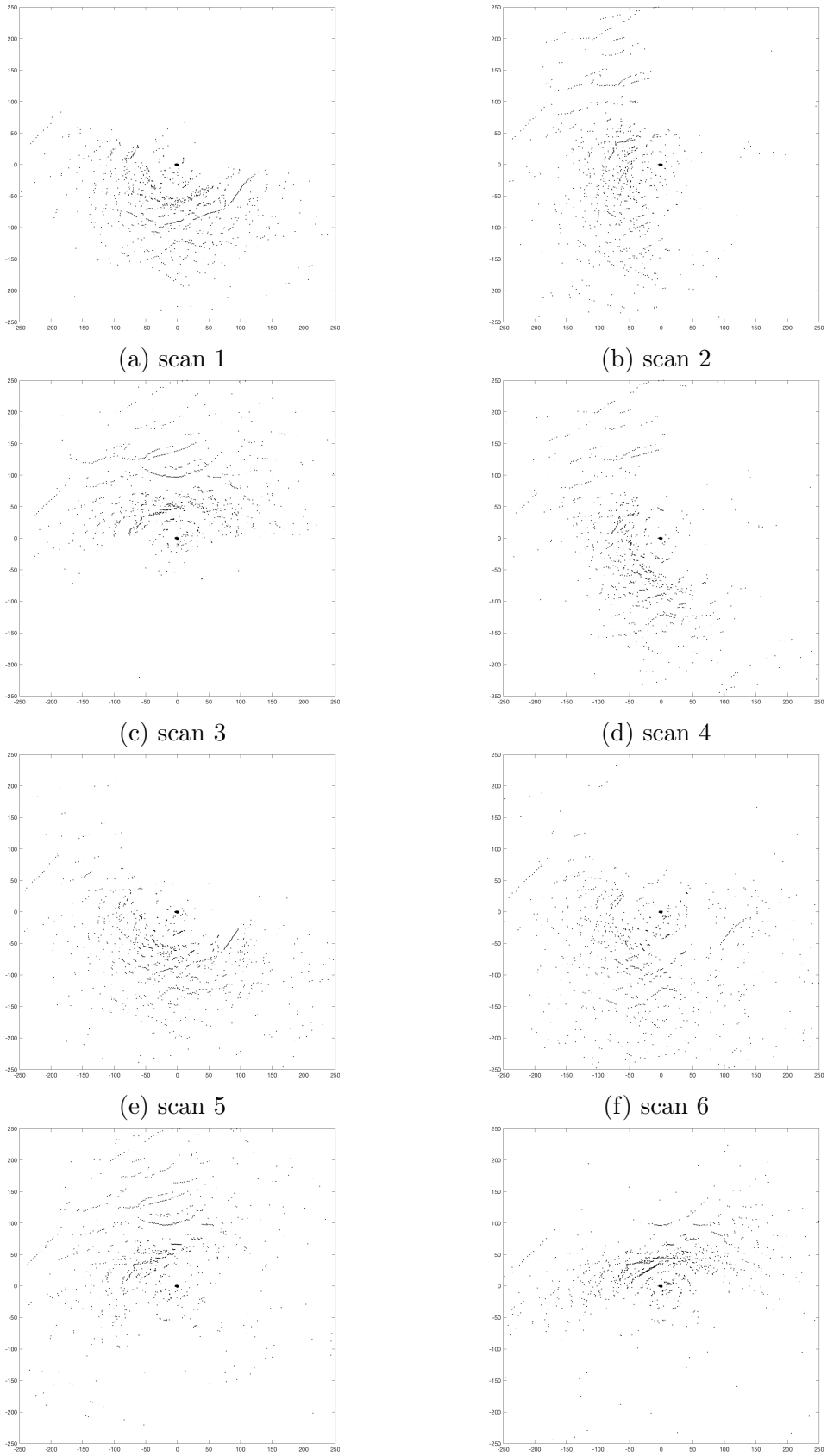
(e) scan 5

(f) scan 6

(g) scan 7

(h) scan 8

Figure 5.1: Cartesian radar scans for route in Thórsmörk Boulder Field, Iceland.



(a) scan 1

(b) scan 2

(c) scan 3

(d) scan 4

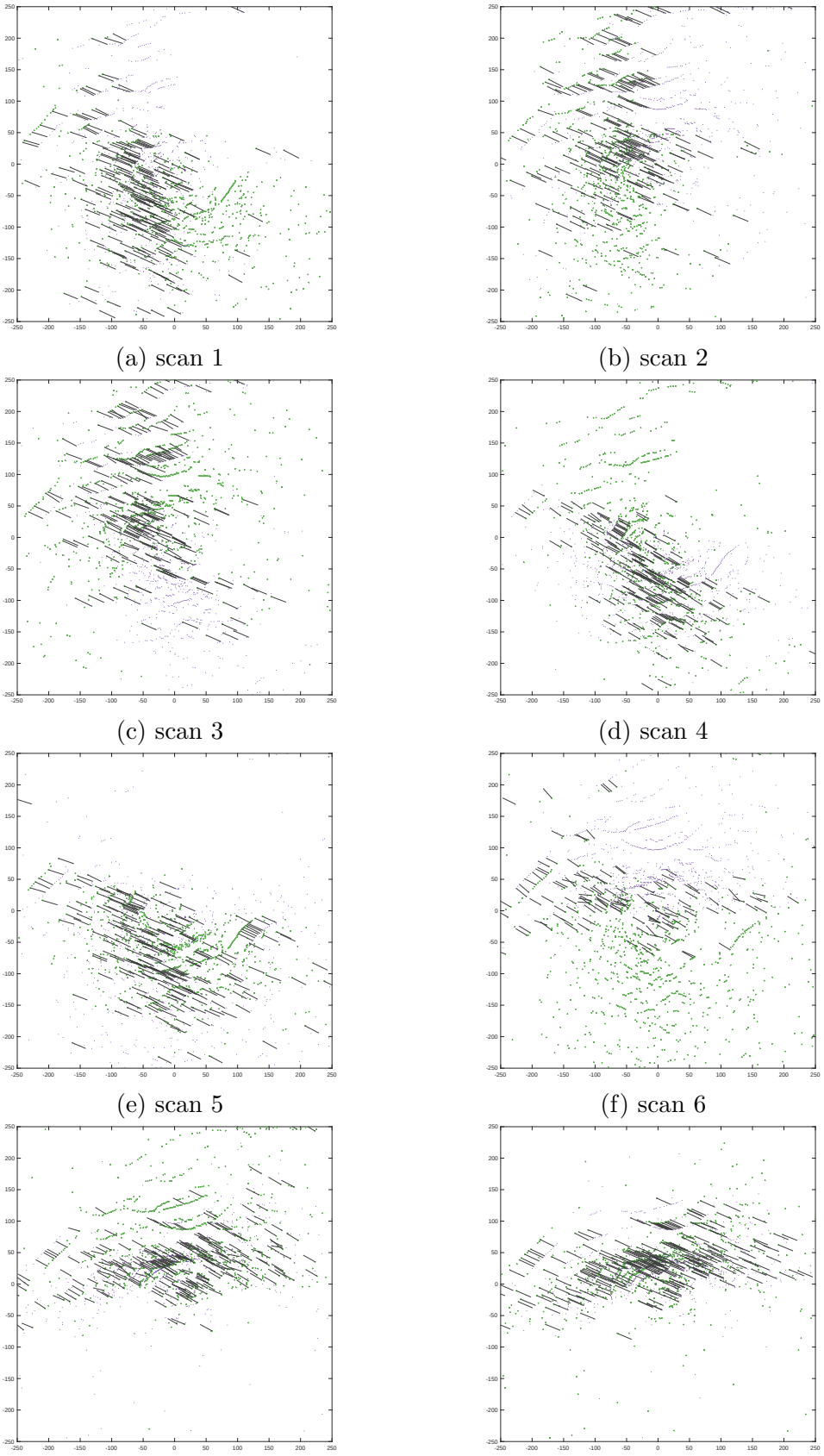
(e) scan 5

(f) scan 6

(g) scan 7

(h) scan 8

Figure 5.2: Keypoint sets for scans shown Figure 5.1.



(a) scan 1

(b) scan 2

(c) scan 3

(d) scan 4

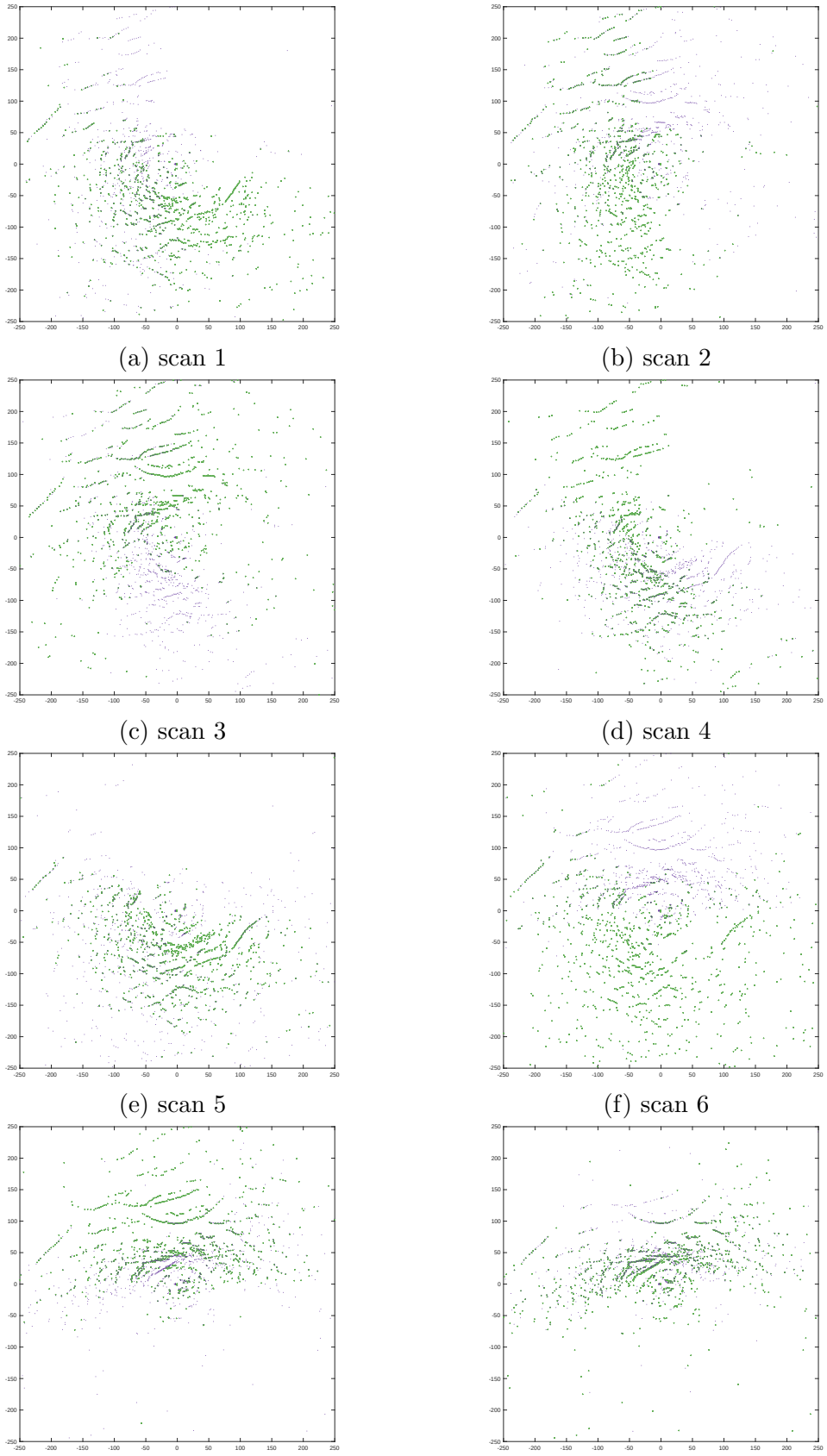
(e) scan 5

(f) scan 6

(g) scan 7

(h) scan 8

Figure 5.3: Matches between keypoint sets shown Figure 5.2.



(a) scan 1

(b) scan 2

(c) scan 3

(d) scan 4

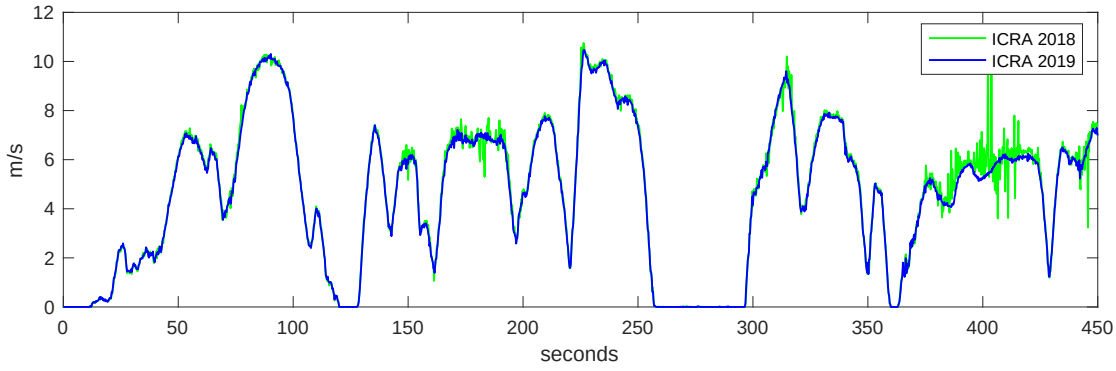
(e) scan 5

(f) scan 6

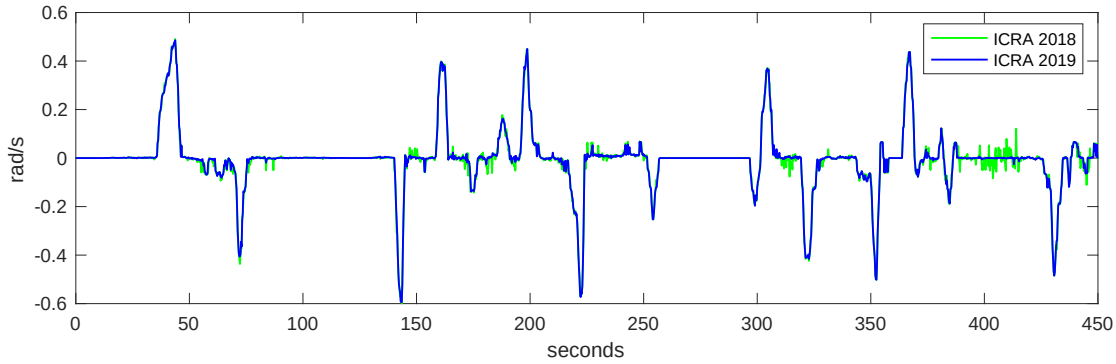
(g) scan 7

(h) scan 8

Figure 5.4: Keypoint sets overlaid according to matches in Figure 5.3.



(a) translational velocity



(b) rotational velocity

Figure 5.5: Comparison of odometry results for pipelines proposed in ICRA 2018 paper [66] and ICRA 2019 paper [70]. The estimated translational velocities are shown in (a) and the estimated rotational velocities in (b). The data is for a route through the Jericho backstreets of Oxford, UK, which poses the problem of organized conspiracy explained in Section 6.2.2. A similar plot is included in [66] to visualize the weaknesses of our first pipeline. As shown, our second pipeline [70] does significantly better.

We limit our comparison of the two papers [66] and [70] to the dataset given in Figure 5.5. We do so because Oxford city center is the easiest dataset for our pipeline to handle, meaning that all other datasets examined in [70] are more difficult than that shown in Figure 5.5. As such, it makes sense to demonstrate that our second pipeline outperforms our first through the Jericho dataset. Although our first pipeline occasionally works in unstructured environments, it is unable to handle difficult portions of non-traditional datasets, such as the forest, countryside, and boulder field ones.

Chapter 6

Discussion

6.1 Keypoint extraction

6.1.1 Problem statement

In order to interpret raw sensor outputs, a common first pre-processing step is to create a reduced representation that captures the most important information conveyed by the sensor while also describing the scene in a holistic manner. For radar, this task can be accomplished via keypoint (or feature) extraction, which seeks to identify areas in a radar scan that most likely correspond to stable or prominent objects in the environment. Such keypoints can include walls, buildings, trees, rocks, and more. However, they can also be dynamic objects, like humans, cars, and cyclists. As a result, keypoint extraction provides concise representations that can be used for many purposes, such as ego-motion estimation, object detection and tracking, path planning, and more.

6.1.2 Considerations and challenges

The main considerations of keypoint extraction for radar are:

- **Type of output**. For instance, whether to produce a grid, set of points, probability distributions, etc.
- **Spatial accuracy**. In the context of ego-motion estimation and localization, keypoints should be located in a precise manner and appear in consistent locations as the radar moves.
- **Compatibility and reliability across environments**. Performance should be consistent under changes in scene structure.
- **Ease of use**. In particular, the number of tunable parameters and the amount of human intervention required for good results should be low.
- **Efficiency in time and space**. This step must be computationally efficient, but it must also produce a representation that is both concise and comprehensive. Conciseness is important because it allows subsequent modules to be executed efficiently. Comprehensiveness ensures that crucial information (e.g., a human near the vehicle) does not go unnoticed.
- **Interpretability of keypoints**. Keypoints that correspond to objects are stable and easy to work with, and they can be used for other tasks, such as object tracking.

The main challenges of keypoint extraction for radar are:

- **False positives**. Radar effects, such as receiver saturation and multipath reflections, cause "ghost" objects to appear. Identifying and removing them is difficult.
- **False negatives**. Occlusions cause objects to be hidden behind those closer to radar. Additionally, radar radiation patterns illuminate objects unevenly such that some are more easily discernible as keypoints than others.
- **Low spatial resolution**. In comparison to lidar, radar's identification of an object's location is less precise.
- **Speckle noise**. Because this type of noise is so pervasive, common gradient-based methods used to identify corners in image processing are not applicable to radar scans.
- **Radar as a 2.5D sensor**. Although the output of radar is a 2D scan, readings are gathered over 3D radiation beams. This discrepancy means that the height and width of radar cross sections are encoded as range, which can lead to measurement errors and blurred object boundaries.

Several of these issues are discussed and explained in Chapter 2.2.4.

6.1.3 Analyzing our approach

We present two algorithms for keypoint extraction. We term the first attenuated noise keypoint extraction (ANKE) and the second the marked region keypoint extraction (MRKE).

Attenuated Noise Keypoint Extraction

ANKE is described, visualized, and tested in [66], which provides the full algorithm. Therefore, the details of our approach are not included here. We discuss the attributes and behaviors of our system. The principal advantages are:

- **Three input parameters**, for which we provide suggestions based on empirical observations. All three are easy to interpret and tune.
- **Meaningful keypoints**. It outputs keypoint sets that are sparse yet interpretable, in the sense explained in Section 6.1.2.
- **Retains spatial accuracy**. It locates keypoints using both low- and high-frequency components of the power-range spectrum. The low-frequency components (i.e., peaks obtained after smoothing the signal) help to distinguish objects from noise. Importantly, the *high-frequency* components preserve high-resolution spatial information encoded in the original signal. In contrast, many other works identify objects in radar scans by performing smoothing or segmenting the scan into regions. These approaches discard key spatial information provided by the radar, an already low-resolution sensor, and, in doing so, reduce positional accuracy.
- **Designed for noise**. It removes radar's high noise floor using the median filter, addresses speckle noise by returning keypoints only after identifying objects, and limits the number of keypoints per object to one.

We note that our algorithm has several flaws, including:

- **False positives**. It does not address the existence of "ghost" objects. Although a perfect keypoint extraction method would remove "ghost" objects, we do not do so for two reasons. First, given limited computation time, we prioritize spatial accuracy in the keypoint extraction step, which makes sense for the objective of accurate ego-motion estimation. Second, we reason that the data association step, which already considers outliers, should be robust to false positives, so we allocate the task of distinguishing between real and "ghost" objects to our scan matching method.

- **Power-based**. It relies primarily on the magnitude of received power to indicate an object's existence, which is an unreliable standard. This characteristic means the algorithm performs well when power correctly indicates occupancy, which is true in structured settings with well-defined buildings, walls, etc.

- **Prior knowledge**. It has three tunable parameters that require some prior knowledge of the environment and the noise characteristics.

Marked Region Keypoint Extraction

MRKE is described, visualized, and tested in [70], which provides the full algorithm. Therefore, the details of our approach are not included here. We discuss the attributes and behaviors of our system. The principal advantages are:

- **One input parameter**. It has only one parameter that specifies the maximum number of desired keypoints.

- **Adapts well across diverse settings**, reducing the reliance on human intervention and expertise.

- **Meaningful keypoints**. It returns keypoints sets that are sparse yet interpretable, in the sense explained in Section 6.1.2. The approach first looks for regions of interest in the scan, which allows it to then return a small number of keypoints to describe the scene. This conciseness reduces computation time in the data association step.

- **Intensity gradients**. In comparison to ANKE, the new approach uses both power magnitude and intensity gradients to select keypoints. As a result, MRKE works well in various different environments (e.g., forest) unlike ANKE, which works best in structured settings (e.g., city).

- **Noise resistant**. Similarly to ANKE, MRKE is designed to function well under high levels of noise. In contrast to ANKE, it extracts keypoints without needing to model the noise characteristics, which further removes reliance on human expertise.

- **Clutter resistant**. It handles scans containing large bright swathes in the scan caused by rolling/pitching or clutter, which are unimportant features in the scan that often distract other keypoint extraction methods.

We note that our algorithm has several flaws, including:

- **False positives**. It does not address the existence of "ghost" objects. Although a perfect keypoint extraction method would remove "ghost objects", we do not do so for two reasons, which are the same as those given for ANKE

in the previous section.

- **Degeneracy**. It can produce noisy keypoints if the input parameter is set very high. This behavior occurs because the algorithm sequentially selects keypoints based on an ordering protocol and, if the number of desired keypoints is too high, the system begins returning those near the bottom of the ordering. It would be helpful to explore how to adaptively predict or learn the number of ideal keypoints in future work.

- **Range errors**. It can have small range errors that result when the marked regions corresponding to the same object in two different scans are of different sizes. This behavior occurs because marked regions are added in a specified order. If a new region overlaps with an already identified one, it grows. Therefore, the size of a region can depend on its rank in the ordering, which may be fluctuate if other parts of scan change dramatically. Addressing the problem listed directly above may mitigate this type of error.

6.2 Data association

6.2.1 Problem statement

Also known as scan matching, this procedure finds correspondences across radar scans. Given the keypoint sets of two separate radar scans, it seeks to match those that represent the same object or location in the environment. As discussed in 3.2.1, data association can be performed by optimizing locally (e.g., using feature descriptors like SURF) or globally (e.g., using scan correlation measures). For

odometry, it is typically assumed that the majority of observed keypoints are static. Under this condition, it is possible to compute the platform's relative motion. We take advantage of this rigid-body property to select matches that approximately optimize a geometrically-based global objective function, which provides better robustness than optimizing locally due to the high levels of noise and other unwanted artifacts in radar scans.

6.2.2 Considerations and challenges

The main considerations of data association of radar keypoint sets are:

- **Algorithm efficiency**. In particular, how the computation time scales with the number of keypoints being matched.
- **Ease of use**. One way to gauge this characteristic is by the number of tunable input parameters.
- **Amount of prior knowledge** required to guarantee robustness. Many existing approaches rely on a good motion prior to approximately align the radar scans and reduce the search space, which is undesirable because it limits the system's possible use cases and makes restrictive assumptions about the motion between scans.
- **Outlier detection**. Removing outliers generally involves human intervention. In addition, outlier thresholds usually require adjustments when operating in different settings. Unless these steps are taken in a principled manner, parameter tuning can be arbitrary and damage performance.

- **Iterations**. Some approaches exhaustively search for the optimal motion parameters using a heuristic. This process involves many repetitive iterations. This can be inefficient and does not guarantee good results if the search space is too small or the algorithm finds an incorrect local minimum.

The main challenges of data association for radar keypoint sets are:

- **High levels of noise**, which can affect both amplitude and location (i.e., range). Significant noise makes certain tasks difficult. For instance, local feature descriptors do not work well because they rely on fine-grained information close to the feature, which can change significantly from scan to scan due to speckle noise.

- **False positives and negatives** due to many factors, including saturation, multipath reflections, and occlusions. These artifacts can significantly alter scene appearance and are the main reason why radar-based systems are challenging.

- **Symmetries in the scan**. Radar scans are noisy and lack multi-channel information, meaning different settings can look similar to radar. The scan alignment strategy that we utilize matches the overall geometry of keypoint sets. Given sufficient geometric diversity, this approach works well for scan matching. Conversely, settings that are perfectly symmetric do not have unique alignment solutions, which can result in scan matching ambiguity. This problem, also known as organized conspiracy [72], typically occurs when the radar is unable to observe a sufficient number of interesting keypoints due to large, nearby occlusions.

- **Range-dependent spatial resolution**. Information obtained at long ranges can be crucial to achieving good alignment, but the detections of faraway objects are less precise than those obtained at short ranges. As a result, it might seem intuitive to assign long-range detections lower weights during data association, but we assert that this is not necessary; radar samples faraway objects fewer times than close objects, which provides a natural weighting for detections. While data association should account for the reliability of a detection using the natural weighting just described, there are certain cases when the sampling bias (i.e., an object's relevance scales inversely with its range) should be accounted for. In particular, our data association algorithm uses a feature descriptor to propose unary matches. Because the feature descriptors should be perspective-invariant (i.e., independent of the radar's location), failing to account for the sampling bias would result in a poor feature descriptor. In our case, objects are weighted by their distance from the vehicle when computing the histograms of the feature descriptor (i.e., a faraway keypoint receives higher weighting). In general, methods that use polar radar scans must consider this bias based on the intended use, as we have done here.

6.2.3 Analyzing our approach

We present a geometrically-based data association method that greedily searches for an optimal alignment using both unary and pairwise keypoint information. Two related but different algorithms are presented in [66] and [70]. Details for both can be found in these papers. The discussion below focuses on the approach

in [70], which implements three improvements, described in the paper, relative to [66]. The principal advantages of our system are:

Advantages

- **Built-in outlier removal**. Outliers are removed naturally without requiring a user-defined threshold.
- **No motion prior**. Our method does not require prior knowledge on the relative displacement between scans because it uses an approach that approximates shape matching. Keypoint sets are matched using their geometry rather than their coordinates. As a result, the system does not need a good prior on the relative motion between scans, which is a common requirement of other ego-motion estimation systems.
- **No tunable parameters** are required, making the system easy to adapt across datasets. As noted in [66], providing a very mild prior (e.g., disregarding impossible motions based on the vehicle's maximum acceleration) can be used if scene symmetry poses an issue (explained in Section 6.3).
- **Global consistency**. Matchings are obtained by approximating the optimal solution to a global objective function. This technique makes use of information across the entire scan and selects matches that are mutually consistent, which improves robustness in comparison to local, feature-based data association methods.
- **Robustness**. Our method handles high levels of measurement noise, occlusions, false positives, and false negatives.

We note that our algorithm has several weaknesses, including:

- **Outlier tipping point**. If the signal-to-noise ratio is very low, there may be too many outliers for the data association algorithm to correctly determine which detections are real. Our system is generally robust to noise, occlusions, and false detections because it (approximately) optimizes a global function defined over pairwise relationships between keypoints. However, this function relies on unary match candidates proposed using feature descriptors. Although our approach can handle bad proposals and rejects them as outliers, if the number of incorrect proposals is overwhelming or they jointly suggest an incorrect solution, then the algorithm will not arrive at the correct solution. In other words, because the optimization can only select matches that are proposed as unary candidates, the system struggles if all the proposals are very poor.

- **Observability sensitivity**. The data association method can return feasible but incorrect solutions when the environment is highly symmetric or certain motion parameters are unobservable. Since our data association approach is geometrically-based and designed to require as little *a priori* knowledge as possible, symmetry means there are many possible feasible solutions. Similarly, if motion in a certain direction is underconstrained by the scene geometry, the algorithm does not know which unique solution to return. This problem can be easily remedied by providing a mild motion prior, as described in [70].

- **Computational bottleneck**. The algorithmic complexity scales quasi-

quadratically with the number of keypoints. We analyze complexity in [70] and identify this bottleneck. Since the bottleneck scales with number of keypoints, we design the only tunable parameter of the keypoint extraction method to be the maximum number of keypoints to select per scan, which the user can easily adjust based on the available computational resources.

6.3 System robustness

The keypoint extraction and data association algorithms are combined to form a full radar-only odometry system that demonstrates good robustness under challenges conditions and in diverse settings. However, the system contains limitations that arise due to sensor characteristics as well as elements of the algorithm, which are discussed next. As described in [70], the following challenges arise when working with radar:

1. False positives (from multipath reflections, side lobes, receiver saturation) and false negatives (from wave absorption, diffuse scattering, occlusions).

2. Measurement errors in keypoint locations caused by wave propagation imperfections, spatial resolution, and other effects.

3. Representational bias of radar, which provides significantly more sensor readings at short ranges than long ranges.

4. Range "errors" due to the projection of the 3D scene onto a 2D scan (e.g., objects below the main beam register as farther away).

In our pipeline, items 1-2 are registered as perturbations during data association.

They manifest as incorrect match proposals and errors in the compatibility matrix values. Our system is robust to false positives and negatives because the framework finds mutually consistent matches while discarding poor ones. It handles measurement and range errors by averaging over many, long-range matches. Accurately representing the scene is important in the feature descriptor, which addresses item 3. If the radar is mounted very high above the ground, item 4 means that radar scans can no longer be interpreted as top-down views of the scene and must be corrected to perform 2D scan matching, which is out of the scope of our paper.

Our pipeline's main sources of error due to environmental factors, computing constraints, and use of a 2D sensor are:

1. Incorrect matches in settings with a lot of symmetry due our geometrically-based approach, which can be avoided by using a prior on inter-scan motion.

2. The relaxation of an NP-hard optimization problem and use of greedy algorithm in data association, which results in an approximated solution.

3. The assumption of 3-DOF motion in the rigid-body motion computation, which means that travelling over a uneven surfaces causes small errors in the ego-motion estimation.

There are several possible failure conditions for our system. It would be constructive to examine how performance is affect by the following factors:

1. The ratio of false positives or false negatives to true detections

2. The ratio of dynamic to static objects

3. Non-planar movements (e.g., rolling and pitching)

4. Perspective changes

5. The availability of long-range keypoints

6. The availability of keypoints and diversity of structures

Empirically, our pipeline can tolerate all these factors within reason (e.g., in the various settings in [70]), but the tolerations limits have not be tested.

6.4 Understanding radar odometry

Many of the topics in this section are discussed in detail in Sections 4.2 and 5.2. This section provides a brief analysis of radar odometry, its results, and design choices in order for the reader to gain an intuition of the main considerations when working with radar.

As explained in Chapter 2 and summarized in Section 6.3, there are many unwanted artifacts in radar scans that make extracting the desired information from them difficult. However, in spite of these challenges, radar offers key benefits that would complement popular sensor suites for mobile autonomy. In particular, its long-range capabilities, lighting invariance, resilience under variable weather conditions, and low false negative rate for nearby objects mean that radar can improve sensor suites in two very important dimensions: *reliability* and *safety*. Therefore, we make the following three observations:

1. Radar is largely underutilized on current mobile autonomous platforms. Rather than incorporate all radar's strengths, most systems employ radar

for specific purposes, like estimating an object's velocity using the Doppler effect or detecting obstacles at close range.

2. The main deterrents of radar are high levels of noise and false positives. One of the challenges of radar is in designing a pipeline that can handle the unwanted artifacts in its scans. Achieving robustness to these effects would allow researchers to take full advantage of the benefits that radar offers.

3. For ego-motion estimation and localization, a second challenge is generating results that are accurate enough to justify the use of radar, which has lower spatial or temporal resolution than other sensors, including lidar, vision, and GPS/IMU-based systems.

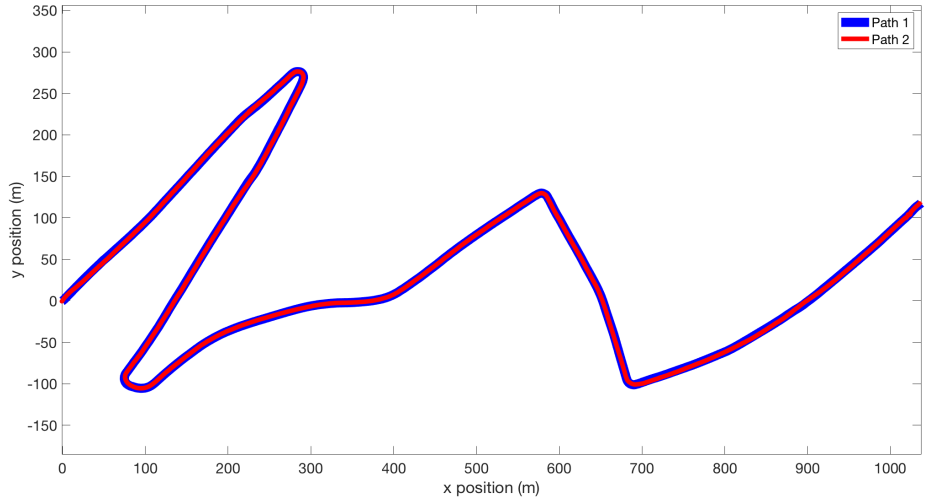
This thesis is centered around these three observations. Although radar should be integrated into a multi-sensor suite, we design a radar-only pipeline in order to illustrate the capabilities of radar on its own. Because the levels of noise and unwanted artifacts in radar scans are much higher than for other sensors, achieving a robust radar-only system could not be done using many of the methods from the vision or lidar literature (e.g., image features and ICP both perform poorly in our tests), Finally, it may seem daunting to produce results at a level of precision higher than the resolution of radar, but this thesis demonstrates centimeter-level scan matching accuracy using a radar with 25-cm resolution and does so without prior knowledge on the relative displacement between scans. We believe that this is possible because of the lever arm effect of long-range detections' as well as our emphasis on sparse, spatially accurate keypoints (see Sections 4.2 and 5.2).

6.5 Localization

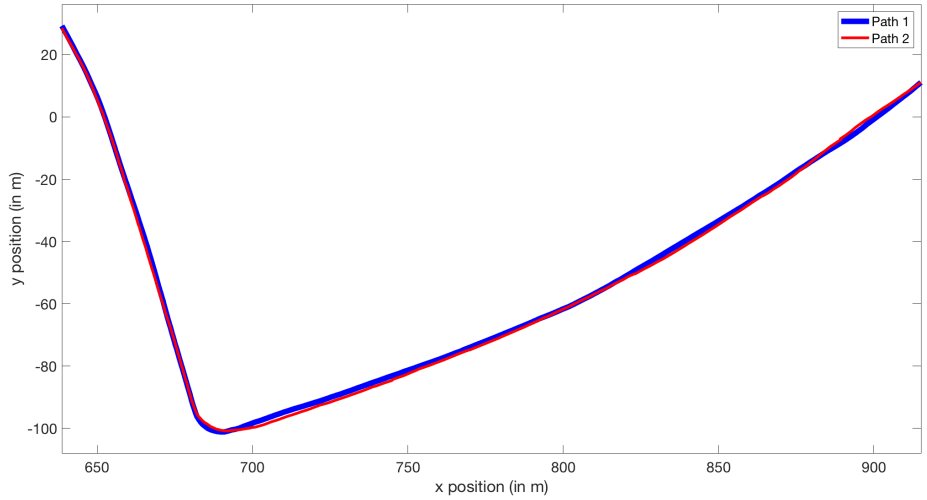
In the two publications included in Chapters 4 and 5, we present systems for ego-motion estimation. However, the pipeline is intentionally modular. The keypoint extraction algorithm aims to return meaningful and stable keypoints. The data association algorithm matches scans without a motion prior. In combination, this means that the pipeline can be easily adapted for localization.

As long as two radar scans are obtained in the same environment and contain sufficient scene overlap, our system can be transformed into a localization pipeline. For example, under a teach-and-repeat framework, given a previously driven path, localizing the vehicle when traveling along the same path becomes a problem of data association. The main difference is that, unlike in ego-motion estimation, the scans being compared are obtained at very different times without a good prior on the relative displacement between the two. Importantly, our data association method is designed to work under these conditions. Therefore, localizing against a known path simply requires matching current scans against former ones. One characteristic of our system is that data association performance is independent of the radar's orientation (i.e., it will align two scans in the exact same manner for any relative rotation between them). Furthermore, as shown empirically in [70], localization is successful under large translations as well.

Under this framework, localization can be interpreted as scan matching across two different datasets. Figure 6.1 verifies that our pipeline can perform teach-and-repeat for paths in Oxford city center, UK. Because the vehicle's position during the second traversal of the path is determined only via localization of the



(a)



(b)

Figure 6.1: Teach-and-repeat demonstration in Oxford, UK. The full path is shown in (a) while (b) shows a portion of the path in more detail. Although the second path is obtained by localizing against scans from the first path, both trajectories are very smooth.

incoming scan against scans observed during the first traversal, performance can be gauged by the smoothness of the second trajectory. Applying this reasoning to Figure 6.1, the second trajectory of the teach-and-repeat experiment is very smooth, indicating that the adaptation of our pipeline for localization works well.

6.6 Summary

Several radar artifacts and characteristics must be considered when designing and constructing algorithms that utilize radar data. We seek to do so by developing a framework that is highly robust to noise, false detections, and other appearance changes. Our pipeline is designed to be easy to adapt across settings and operational under variable conditions. We successfully demonstrate accurate radar-only ego-motion estimation and localization. The work in this thesis centers around three core observations about radar and its application to odometry, explained in Section 6.4. There are natural limitations of our approach, which we identify and discuss. Finally, we verify that our pipeline is compatible with localization.

Chapter 7

Conclusion and future work

In this thesis, we propose a radar-only pipeline for precise and robust ego-motion estimation and localization. We demonstrate our results using an FMCW scanning radar. Radar is a promising alternative sensor for mobile autonomous systems that operates well under variable lighting and weather conditions as well as indoors and outdoors. Its long range capabilities and information-rich scans make it suitable for tasks, such as odometry, localization, path planning, and object detection. However, in comparison to other popular sensor systems, radar presents several challenges. In particular, it is a noisy, low-resolution, and low-speed sensor that often returns false positives and compresses 3D information into 2D scans.

We develop a system that is capable of handling these difficulties and requires minimal fine-tuning between datasets. It consists of two modules: keypoint extraction and data association. We demonstrate highly accurate ego-motion estimation, comparing our radar-only approach against visual odometry and GPS. We give an in-depth system analysis that also discusses the vulnerabilities of our approach.

There are several promising directions for future work. Two ways to im-

prove our pipeline would be to streamline the unary candidate proposal method, which is the most computationally expensive operation, and to address under-constrained scene geometries. For the former, using deep learning to identify candidate matches, for instance, would take advantage of a neural network's ability to encode inputs as feature vectors and perform quick computations at run time. Possible directions for the latter issue include investigating tertiary graph matching and providing a metric indicating whether certain motion parameters are observable given the scene geometry.

This thesis also verifies that localization using radar is possible. However, these results are limited in scope, and a more in-depth examination of radar localization should be done. For example, two difficulties that may arise are insufficient scene uniqueness (e.g., incorrectly matching two scans that look similar but are from different locations) and perspective change (e.g., not recognizing the same location due to a different viewpoint). A similar objective would be to investigate loop closure with radar, which provides the additional benefit of correcting current beliefs about the trajectory.

While this thesis focused on radar, future work could factor in other sensors. Sensor fusion with lidar or vision sensors—which are short-range and highly detailed—would complement radar, which has long range but low resolution. These sensors could be incorporated in a variety of ways, such as (1) straightforwardly using each to verify the information of the other, (2) combining the estimates provided by two independent sensor systems to produce a more refined estimate, or (3) using one sensor to learn how to better extract more accurate information from the other. An example of the third is [73], which uses lidar's

precision to train a neural net that generates occupancy grids from radar scans. In general, one major obstacle to sensor fusion is calibration. Alternatively, due to the modularity of the framework presented, another way to explore other sensors would be to adapt and test the pipeline for similar data, like lidar pointclouds or sonar scans. It would be productive and interesting to compare the results across sensors to evaluate the applicability of our method.

We also suggest extending our work in the following ways. First, although a discussion of our system's sources of error is included in this thesis, an important next step is performing a rigorous uncertainty analysis. This might involve testing the system's response to removing salient information or adding noise. Another way to naturally incorporate uncertainty is via filtering (e.g., Kalman or particle), which would additionally provide a principled approach for correcting the trajectory after loop closure. Second, as the scan matching method naturally removes objects it perceives to be dynamic, another extension would be to modify the data association algorithm to identify and track moving objects. Third, our keypoint extraction approach returns keypoints and their positions, discarding all other scan information. It may be possible to deduce other keypoint characteristics, like the object material or size, that can aid in unary candidate matching or other tasks. Finally, another limitation of scanning radar is that it collapses 3D information onto 2D scans. However, there may be ways to infer 3D information. For instance, the change in an object's appearance as it moves relative to the radar could indicate its height, and large changes in intensity in certain portions of a radar scan can indicate that a vehicle is rolling or pitching.

Bibliography

[1] S. Stergiopoulos, *Advanced signal processing handbook: theory and implementation for radar, sonar, and medical imaging real time systems.* CRC press, 2000.

[2] M. I. Skolnik, "Radar handbook," 1970.

[3] P. Dong and Q. Chen, *LiDAR Remote Sensing and Applications.* CRC Press, 2017.

[4] R. P. Hodges, *Underwater acoustics: Analysis, design and performance of sonar.* John Wiley & Sons, 2011.

[5] M. Adams and E. Jose, *Robotic navigation and mapping with radar.* Artech House, 2012.

[6] "A430s15 radiation pattern e-plane," https://www.diamondantenna.net/, accessed: 2018-09-30.

[7] M. Golio and J. Golio, *RF and Microwave Applications and Systems*, ser. The RF and Microwave Handbook, Second Edition. CRC Press, 2007. [Online]. Available: https://books.google.co.uk/books?id=fNJLcL1LBpEC

[8] "Navtech radar technical specifications," https://navtechradar.com/technical-specifications/, accessed: 2019-02-20.

[9] "Bosch lrr4," https://cds.bosch.us/themes/bosch_cross/amc_pdfs/LRR4_292000P0ZH_EN_l accessed: 2019-02-20.

[10] "Delphi electronically scanning radar," https://www.autonomoustuff.com/wp-content/uploads/2016/08/delphi-esr.pdf/, accessed: 2019-02-20.

[11] "Velodyne lidar hdl-64e," https://velodynelidar.com/hdl-64e.html/, accessed: 2019-02-20.

[12] "Flir bumblebee 21394a," https://www.ptgrey.com/support/downloads/10132/, accessed: 2019-02-20.

[13] "Novatel span-cpt," https://www.novatel.com/assets/Documents/Papers/SPAN-CPT.pdf/, accessed: 2019-02-21.

[14] "Navtech radar," https://www.navtechradar.com/, accessed: 2018-09-30.

[15] C. Latham, A. Stobbs, and R. S. of Great Britain, *The Birth of British Radar*. Radio Society of Great Britain, 2012. [Online]. Available: https://books.google.com/books?id=dPwlmgEACAAJ

[16] P. Zarchan, *Tactical and strategic missile guidance*. American Institute of Aeronautics and Astronautics, Inc., 2012.

[17] K. Tsipis, "Cruise missiles," *Scientific American*, vol. 236, no. 2, pp. 20–29, 1977.

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

[19] T. East, "Location estimation and guidance using radar imagery," *IEEE Transactions on Aerospace and Electronic Systems*, no. 6, pp. 1383–1388, 1975.

[20] R. G. Caves, P. J. Harley, and S. Quegan, "Matching map features to synthetic aperture radar (sar) images using template matching," *IEEE Transactions on geoscience and remote sensing*, vol. 30, no. 4, pp. 680–685, 1992.

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

[22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[23] T. Deissler and J. Thielecke, "Uwb slam with rao-blackwellized monte carlo data association," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–5.

[24] H. Rohling, "Ordered statistic cfar technique-an overview," in *Radar Symposium (IRS), 2011 Proceedings International*. IEEE, 2011, pp. 631–638.

[25] D. Vivet, P. Checchin, and R. Chapuis, "Localization and mapping using only a rotating fmcw radar sensor," *Sensors*, vol. 13, no. 4, pp. 4527–4552, 2013.

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

[27] ——, "Relative radar cross section based feature identification with millimeter wave radar for outdoor slam," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 1. IEEE, 2004, pp. 425–430.

[28] 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 *Radar Conference (EuRAD), 2013 European.* IEEE, 2013, pp. 471–474.

[29] J. Callmer, D. Törnqvist, F. Gustafsson, H. Svensson, and P. Carlbom, "Radar slam using visual features," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 71, 2011.

[30] F. Schuster, C. G. Keller, M. Rapp, M. Haueis, and C. Curio, "Landmark based radar slam using graph optimization," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on.* IEEE, 2016, pp. 2559–2564.

[31] K. Werber, J. Klappstein, J. Dickmann, and C. Waldschmidt, "Interesting areas in radar gridmaps for vehicle self-localization," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on.* IEEE, 2016, pp. 1–4.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[34] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang, "Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3982–3991.

[35] G. Bertasius, J. Shi, and L. Torresani, "Deepedge: A multi-scale bifurcated deep network for top-down contour detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4380–4389.

[36] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *arXiv preprint arXiv:1606.00915*, 2016.

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

[38] M. Rapp, K. Dietmayer, M. Hahn, F. Schuster, J. Lombacher, and J. Dickmann, "Fscd and basd: Robust landmark detection and description on radar-based grids," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on.* IEEE, 2016, pp. 1–4.

[39] 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.

[40] P. J. Besl, N. D. McKay *et al.*, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

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

[42] O. Sorkine, "Least-squares rigid motion using svd," *Technical notes*, vol. 120, no. 3, p. 52, 2009.

[43] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on.* IEEE, 2001, pp. 145–152.

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

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

[46] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition," in *European conference on computer vision.* Springer, 2000, pp. 18–32.

[47] R. Fergus, P. Perona, and A. Zisserman, "A sparse object category model for efficient learning and exhaustive recognition," in *null.* IEEE, 2005, pp. 380–387.

[48] R. Carraghan and P. M. Pardalos, "An exact algorithm for the maximum clique problem," *Operations Research Letters*, vol. 9, no. 6, pp. 375–382, 1990.

[49] I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo, "The maximum clique problem," in *Handbook of combinatorial optimization.* Springer, 1999, pp. 1–74.

[50] J. Neira, J. D. Tardós, and J. A. Castellanos, "Linear time vehicle relocation in slam," in *ICRA.* Citeseer, 2003, pp. 427–433.

[51] Q. Wu and J.-K. Hao, "A review on algorithms for maximum clique problems," *European Journal of Operational Research*, vol. 242, no. 3, pp. 693–709, 2015.

[52] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *International journal of pattern recognition and artificial intelligence*, vol. 18, no. 03, pp. 265–298, 2004.

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

[54] B. Jain and K. Obermayer, "A necessary and sufficient condition for graph matching being equivalent to the maximum weight clique problem," *arXiv preprint arXiv:0912.4584*, 2009.

[55] 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 *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on.* IEEE, 2009, pp. 4925–4930.

[56] 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.

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

[58] 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.

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

[60] V. Mohanty, S. Agrawal, S. Datta, A. Ghosh, V. D. Sharma, and D. Chakravarty, "Deepvo: A deep learning approach for monocular visual odometry," *arXiv preprint arXiv:1611.06069*, 2016.

[61] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2938–2946.

[62] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[63] A. Kendall and R. Cipolla, "Modelling uncertainty in deep learning for camera relocalization," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 4762–4769.

[64] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[65] S. H. Cen and P. Newman, "Radar-only ego-motion estimation in difficult settings via graph matching," YouTube Spotlight Video for ICRA, 2019. [Online]. Available: https://www.youtube.com/watch?v=OHcj2A4cPMU

[66] ——, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions," in *Robotics and Automation, 2018. Proceedings. (ICRA 2018). 2018 IEEE International Conference on.* IEEE, 2018.

[67] W. Churchill, "Experience based navigation: Theory, practice and implementation," Ph.D. dissertation, University of Oxford, Oxford, United Kingdom, 2012.

[68] "Novatel oem7720," https://www.novatel.com/assets/Documents/Papers/OEM7720-Product-Sheet.pdf/, accessed: 2019-02-21.

[69] S. H. Cen and P. Newman, "Radar-only ego-motion estimation in difficult settings via graph matching," YouTube Spotlight Video for ICRA, 2019.

[70] ——, "Radar-only ego-motion estimation in difficult settings via graph matching." Under review, 2018.

[71] "Lord microstrain 3dm-gx4-45," https://www.microstrain.com/sites/default/files/3dm-gx4-45_datasheet_8400-0059.pdf, accessed: 2019-02-21.

[72] Y. Latif, C. Cadena, and J. Neira, "Realizing, reversing, recovering: Incremental robust loop closing over time using the irrr algorithm," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 4211–4217.

[73] R. Weston, S. Cen, P. Newman, and I. Posner, "Probably unknown: Deep inverse sensor modelling in radar," *arXiv preprint arXiv:1810.08151*, 2018.