Exploring Novel Sensing Paradigms for Existing Wireless Infrastructure

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Diana Zhang

B.S., Electrical Engineering, The Pennsylvania State University
B.S., Computer Engineering, The Pennsylvania State University
M.S., Electrical and Computer Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

August 2021

Dedication

To my father –

whose American dream is three children with PhDs.

Here is the first.

Acknowledgements

I would like to acknowledge my advisor, Swarun Kumar, who used to sit and debug kernel code with me into the evening. Though he is now six students and two children busier than he used to be, he has remained available for late-night panicked calls three nights before a paper deadline. I am so grateful to have had continuous access to the calm and optimistic guidance that earned him MobiCom 2017's Nicest TPC Member award.

I also would like to thank my committee: Anthony Rowe, Peter Steenkiste, and Stephanie Gil, for their kindness, feedback, and support towards this dissertation and throughout the course of my Ph.D.

Next, my amazing co-authors, without whom these projects would never exist: Akarsh Prabhakara, Jingxian Wang, Junsu Jang, Junbo Zhang. Thank you all for sharing the burden of late nights and the endless power cycling trying to get our systems to do what they were *just doing* fifteen minutes ago! They're fantastic and you should hire them. Of course, WiTech Lab overall has also been a great technical and emotional support, providing generous feedback and a safe place to bounce ideas around.

Additionally, I would like to thank my surrogate upperclassmen mentors, Adwait Dongare, Alexei Colin, and Elahe Soltanaghaei – thank you for being available as senior students to me since my lab group had none.

I also acknowledge my family – my parents who came to this country with no technical or English skills to give me every opportunity they could. Mom, I'm sorry I did not become a pharmacist. My brothers: Tony, who inspired me to pursue Electrical Engineering and Brian, a mechanical engineering student whom I call with any problem that requires spatial skills.

Thanks to my partner, Tom, who deserves a tome of gratitude the size of a second thesis but is getting this instead. Thank you to the roommates who made my time here more enjoyable: Victor, thank you for lending me your chairs so I could run experiments on them; Ben, I'm sorry I almost dropped a weight on your head; Gordon, Katherine, and Thomas, I will miss staying up until 4am talking about who-knows-what (although my aging body will not). Thanks to my floormates for those fun days at CIC I miss, and to our amazing admins Grace and Dawn for all they did to support my work.

My work was funded by the National Science Foundation through CRII Grant #1657318 and CRI Grant #1823235, as well as the ARCS Foundation, the Google Women Technmakers scholarship program, the Michel E. and Kathy Doreau Fellowship in Electrical and Computer Engineering, and the Marshall Fellowship.

Abstract

Over the past few decades, we have experienced the permeation of commodity wireless devices into every facet of our existence: our lives at home, our commutes and community infrastructure, and our workplaces. Wireless devices are already on board many sensing platforms, thus making them appealing to explore for adding sensing capabilities without increasing bulk. For example, consider the ability to use WiFi, already aboard a UAV for transmitting video data, to additionally provide information on the environment. In this dissertation, we explore opportunities in re-engineering these already widely deployed technologies (specifically, WiFi, mmWave RADAR, and cameras) for sensing object material and long-range depth imaging – challenges that traditionally require large additional sensor loads (e.g., X-band and K-band RADAR) and struggle with occlusions like fog and walls (in the case of cameras and LIDAR). The ability to sense object material and image depth, in both line-of-sight and non-light-of-sight settings, would enable future vehicles and cities to better respond to environmental obstacles. For example, we envision a future disaster-response drone that can navigate a scene with responses guided by awareness of occlusion *type* – e.g., detecting a tree and flying lower to avoid getting caught in branches, or detecting a human under rubble sending a distress call. Beyond object type, the ability to image at a long range is also helpful for responding quickly to objects in the environment, reducing the flight path (and thus, battery) necessary to survey a scene, and providing environment dimensions to better guide rescue crews on equipment needs.

We approach this vision first from the perspective of exploring novel sensing capabilities of WiFi, since it is comparatively less studied than RADAR and vision. Specifically, we develop and evaluate a WiFi-based approach to both localize and identify object type (specifically composing material) of obstacles in the environment in our system IntuWition. IntuWition uses polarimetry – the measurement of how the polarization of signals is changed as it passes through or reflects off objects – as a position- and orientation-robust way to measure the material type of objects in the environment. This leaves the challenge of imaging objects in a scene, ideally at a long range and at high resolutions. mmWave RADAR and cameras emerge as the best candidates for this, but they individually face some challenges: mmWave RADAR has extremely poor angular resolution (about 1 degree in azimuth and about 11 degrees in elevation for a top-of-the-line model), while cameras have poor depth resolution, particularly at long ranges. We present the development and evaluation of a hybrid mmWave/camera system, Metamoran, that uses camera information to guide the de-noising, angular positioning, detection, and imaging by a mmWave RADAR. These approaches help to demonstrate that new approaches to sensing that exploit the proliferation of wireless devices can enable increased environmental awareness.

Contents

List of Figures

Chapter 1

Introduction

Commodity wireless devices have permeated our daily lives, continuing to change how we work, commute, and live. In 2020, the average number of connected devices per household in the United States was 10.37 [10]– coupled with growing industrial and infrastructural connectivity, we see connected devices truly becoming widely deployed.

We ask the question – "What opportunities exist in re-engineering commodity wireless devices, without changing their hardware, to enable new capabilities?" The prevalence of these wireless devices makes them appealing to explore as a sensing modality – after all, they are already deployed with such density. If they could additionally be exploited to sense as well as communicate, they could provide additional functionality without additional hardware and infrastructure cost. In addition to the advantage of its availability, wireless has additional advantages over common sensing modalities. It is less power consumptive than RADAR or LIDAR. Unlike camera/visible light-based sensing solutions, wireless signals operate well in occluded environments – including through-wall to through-fog sensing, depending on the frequency. As a result, commodity wireless for sensing has been explored significantly in the literature, including for localization[54], tracking[108], and health applications [117]. The benefits of being a commercially available device, produced at a large scale for everyday consumers, is also important: the large scale and competition with which commodity wireless devices are produced for communication quickly encourages both technological improvement and cost reduction strategies, and regulators are incentivised by widespread use to maintain and expand upon the wireless spectrum available for unlicensed use[30]. Over time, these trends would drive down the cost of commodity wireless sensing systems while increasing the bandwidth available to them, thus improving their resolution and performance.

In Fig. 1.1, we show some of the spectrum with commodity wireless devices on the market, as well as the strengths of each of these frequency bands. In general, the larger the wavelength, the wider the

Frequency	Wavelength	Permeability	Bandwidth	Range Resolution	Noise	Example Applications	Pervasiveness
< 400MHz	> .75m	High	Low	Low	High	NFC	Limited
433 MHz	.7m	7 cm	1.74 MHz	86m	High	Ham Radio	Limited
915 MHz	.3m	3 cm	26 MHz	5.8m	Low	Zigbee, LPWAN	Emerging
2.45 GHz	12cm	1.2 cm	100 MHz	1.5m	Low	WiFi, Bluetooth	High
5.8 GHz	5cm	.5 cm	150 MHz	.75m	Low	WiFi, RFID	High
24 GHz	12.5mm	1.25 mm	250 MHz	.6m	High	ADAS	Emerging
61 GHz	5mm	.5 mm	500 MHz	.3m	High	WiGig	Emerging
77 GHz	4mm	.4mm	4GHz	.08m	High	Vehicular	Emerging
122 GHz	2mm	.2 mm	1 GHz	.15m	High	Emerging	Emerging
244 GHz	1mm	.1 mm	2 GHz	.08m	High	Emerging	Emerging
500 THz	600nm	Low	320 THz	High	High	Cameras	High

Figure 1.1: Benefits and downsides of commodity wireless spectrum in the US.

obstacles it can ignore (rule of thumb is an object an order of magnitude smaller than the wavelength can be ignored). More bandwidth provides more range resolution, and bandwidth is more available at higher frequencies. We note the region around 1GHz-10GHz, commonly referred to as the "microwave window"[4], has the most ideal noise characteristics: lower frequencies include galaxy noise, and higher frequencies face noise due to atmospheric absorption of water and oxygen. However, unlicensed bands in this range are sensitive to noise sources such as microwave ovens and electrical noise. Finally, we consider how widely distributed the given frequency already is – with technologies such as WiFi, Bluetooth, and cameras already widely deployed in many homes and workplaces.

There are many possible frequencies to choose from, as well as many possible new approaches to using object-specific characteristics for identification. For our work, we chose three frequency bands of electromagnetic spectrum to explore further based on current and projected prevalence: WiFi, Automotive mmWave RADAR (77GHz), and cameras. These frequencies also benefit from substantial past research to form a foundation. WiFi has the benefit of propagating through and around walls due to its lower frequency, but suffers from limited range and resolution. Automotive mmWave has excellent range resolution, but struggles in the angular domain. It can work well in partially occluded environments, as well as penetrate very small obstacles such as fog and smoke. Camera has the best angular resolution – each pixel can provide a fraction of a degree – but it can only work in unoccluded environments and it struggles with providing depth.

We explore two questions in the context of these three most prevalent electromagnetic sensing technologies: (1) What new properties would be beneficial to sense that cannot already be sensed? (2) What

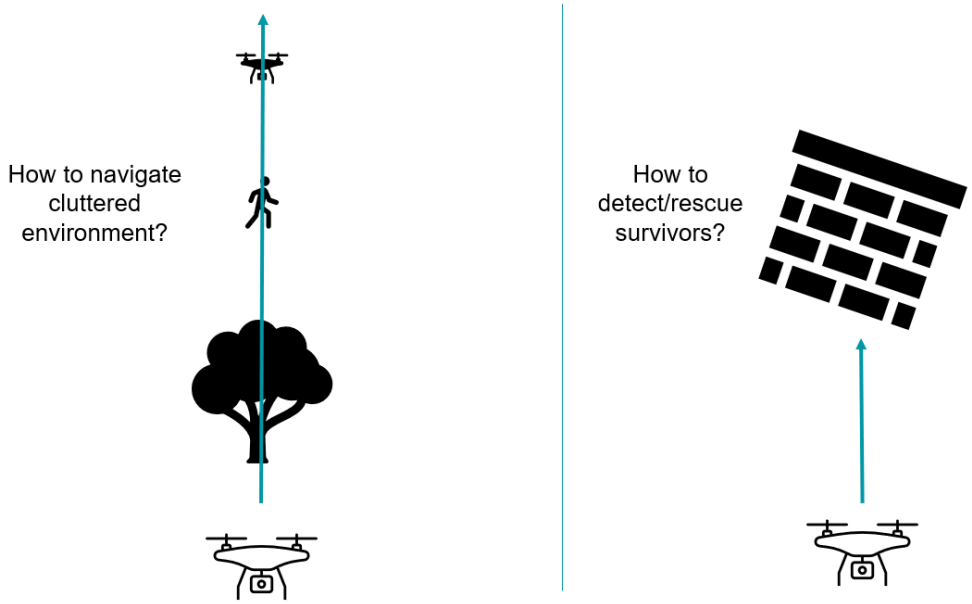

Figure 1.2: There are many challenges that a drone at a disaster site might face: in particular, we consider navigating a cluttered environment and detecting survivors. Existing sensing hardware for these challenges space have significant limitations, including not working in occluded environments or being bulky.

missed opportunities have yet to be explored in commodity wireless sensing? These questions are each expanded upon further in the following sections. The exploration of missed opportunities in section (2) further outlines the two systems whose design, implementation, and evaluation form the core contributions of this dissertation.

1.1 Vision: New Sensing Capabilities of Commodity Wireless

Some challenges that today's UAVs face, such as navigating cluttered environments and detecting objects in non-line-of-sight, are shown in Fig. 1.2. We envision a future UAV tasked with searching for survivors at a disaster site. It can navigate an area without human control, even in visually occluded environments like smoke and fog. It can respond to obstacles beyond just avoiding collisions: with object-specific responses, the UAV might fly lower near a tree to avoid getting caught in its branches, quickly swerve to avoid collision with another UAV, or give a person a large berth to avoid startling them. The UAV has the ability to sense survivors even under collapsed building materials, and can even provide information on the location, shape, dimensions of the collapsed building to inform rescue teams of what type of equipment is necessary to respond to this situation, shown in Fig. 1.3. The ability to sense the environment despite

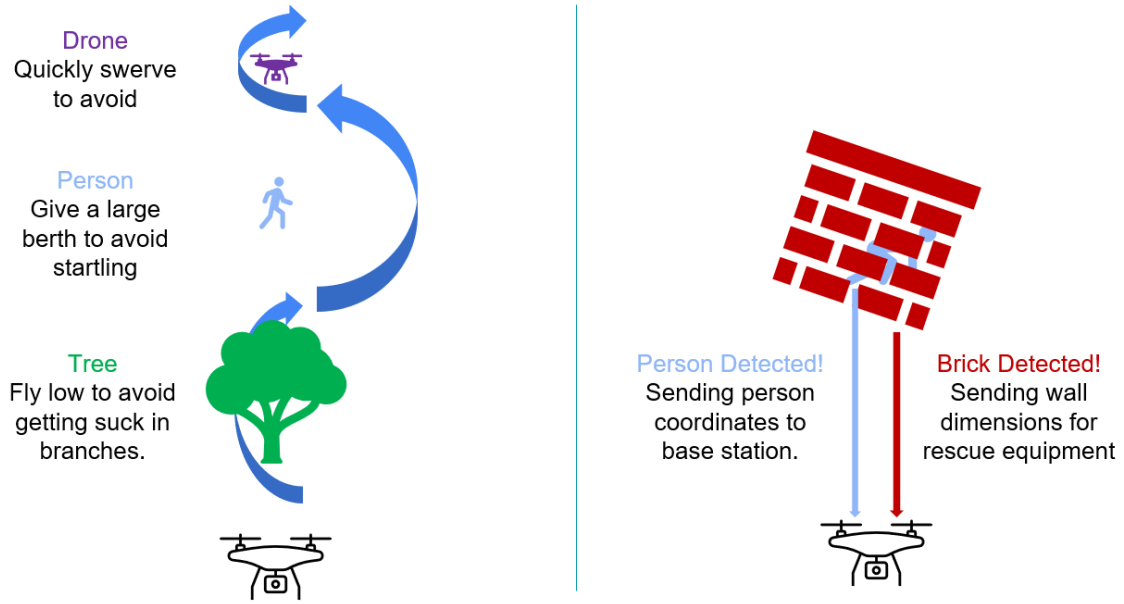

Figure 1.3: We envision new sensing capabilities that can provide navigational guidance based on object *type*, detect reflectors even in non-line-of-sight, and image the environment to help rescue teams.

occlusions, to sense object *type*, and to image objects from a long distance are the three goals that motivate this dissertation. They can be expanded beyond disaster relief to applications such as package delivery, surveillance, and autonomous vehicles.

To work toward this vision, this thesis explores opportunities in using wireless communication channels as a sensing modality, both in isolation and hybridized with existing dominant sensing modalities. Specifically, our goal is to enable and enhance object *type* detection and imaging, via three capabilities:

1. Localization of objects, including behind obstructions
2. Material sensing the determine the type of objects
3. z-location and contour sensing for a 3-D depth map of the obstructed environment

These capabilities can provide autonomous vehicles with more environmental knowledge.

Consider a scene as depicted in Fig. 1.4, being imaged with a combination of common sensors: a camera, WiFi, and an automotive mmwave RADAR. In this scene, we seek to determine the position of reflectors, the object type of reflectors, and a contour map of the area. How can we leverage these sensors and process this image to address this challenge? We note that one approach is to simply collect data from all three sensors and fuse via machine learning model. To preserve simplicity and transparency, we instead address this open problem in parts.

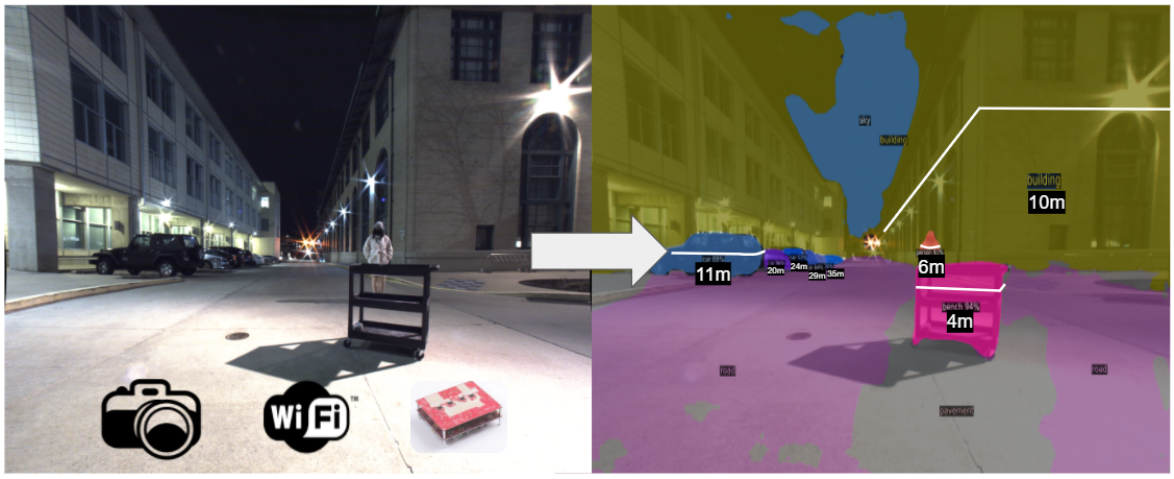

Figure 1.4: Above, we show the vision for our project. This envisioned system can detect the position, depth profile (contour map), and object *type* of objects in its environment. Such a system would enable autonomous vehicles to make obstacle-specific responses to objects in its environment.

1.2 Technique: New Sensing Approaches to Commodity Wireless

In this dissertation, we focus on two missed opportunities in commodity wireless sensing. The first is inspiration from traditional radar techniques, specifically RADAR Polarimetry, to develop a WiFi sensing system that can detect material-specific properties and distinguish materials. The second is a hybrid approach to commodity wireless sensing with camera sensing – currently the most prevalent electromagnetic sensing technology – which takes advantage of physical properties to improve long-range depth sensing and imaging.

1.2.1 Polarimetry for WiFi-Based Material Sensing

First we consider the question: what can we do with just WiFi alone? WiFi is a widely available technology whose frequency makes it suitable for propagating around corners and through thin (few centimeters) objects. This makes it an attractive platform for localization of objects in non-line-of-sight at decimeter-scale. For our project *IntuWition*, we further build on WiFi's inherent sensing opportunities to determine material *type* of an object using RADAR polarimetry, detailed further in Ch. 3.

For IntuWition, resilience to objects fully and partially occluded by fog is critical for system performance. For thinner reflectors, this is a capability intrinsically available to WiFi. However, this is complicated by the need to have sensors aboard a single system. Because our sensing system must exist upon a single platform – e.g., a drone – the environment must be sensed using signals reflected from the environment, akin to RADAR or LIDAR. This is in contrast to other wireless sensing approaches, which might

sense material of an object penetratively [27] or image an area using devices set up on both sides of the scene [48]. This complicates how to isolate individual target objects, as commodity wireless devices are not typically designed transmit and received from a specific direction as RADAR and LIDAR samples do. We describe our approach for isolating individual objects based on measured time-of-flight windows in Sec. 3.5.

Another challenges is, when using a wireless signal to image an object, very small adjustments can significantly modify the signal at the receiver. How can we take advantage of this to sense composing material in a way independent of factors like position and orientation? For example, a small positioning difference in orientation/location could be the difference between constructive interference, destructive interference, or even reflections that redirect the majority of the signal away from our device. Other qualities of a material, such as surface texture, also affect the reflected signal received. And indeed, there are many cases where it is difficult to decouple positioning characteristics from object characteristics – how might a strong reflector (e.g. a metal sheet) be distinguished from a weaker reflector (e.g. a wooden sheet) that is larger, or positioned closer, or oriented at a more favorable cross-sectional area? We present in Sec. 3.4 an approach that takes advantage of polarimetry of the reflecting object to provide positioning-agnostic material identification.

1.2.2 Combining mmWave and Camera sensing for long-range depth imaging

Next, we consider sensing opportunities for two devices that are already commonly available in advanced driver assistance systems: camera and mmwave. Using cameras provide access to the tools created in the well-developed field of camera vision, for techniques including segmentation. Furthermore, cameras have pixel-scale angular resolution and can detect object position to a tenth of a degree. However, it struggles with depth resolution. mmWave, in contrast, provides the ability to image and detect objects several tens of meters away at a cm-scale while struggling with angular resolution (1.4 degrees in azimuth and 18 degrees in elevation for a top-of-the-line model). The natural complementary abilities of these two sensing modalities, in addition to their joint prevalence in vehicles, makes this an interesting hybridization to explore. We discuss this in Ch. 4.

One challenge is that a fundamental physics limitation of commodity wireless is resolution. Bandwidth is limited, and even with additional antennas, state-of-the-art commodity wireless RADARs are limited to degree-scale angular resolution. Compare this to a camera, which can easily provide tens of pixels per degree. Cameras, however, struggle with depth detection – and indeed, there remains a gap for a depth imaging, at high resolution, and at a long range. We discuss our approach for a long-range depth imaging

solution, a hybrid mmwave/camera sensing system, in Sec. 4.4.

Finally, we consider the open challenge of how to best hybridize different technologies. We approach this using the simplest case of two technologies with complementary capabilities – selecting mmwave and camera. Most sensor fusion occurs at higher levels – i.e., cross-referencing objects that have already been detected by RADARs and cameras to confirm their existence and pair depth and angular information. We instead ask – can we do better if we use one sensor to inform the others' processing? Specifically, we consider the case of a known angular breadth and position of objects-of-interest – as identified by computer vision algorithms – to better identify which reflector on our mmwave RADAR is a given object-of-interest. This is further discussed in Sec. 4.6.

1.3 Thesis Statement and Contributions

Our objective for this work is to explore new capabilities for existing wireless infrastructure. The commodity wireless sensing techniques we explore in this work, as well as their deployment and evaluation, supports the following thesis:

New approaches to sensing using existing, widely deployed wireless devices can enable increased environmental awareness for future cities.

Our contributions, the sensing paradigms explored and evaluated to serve as evidence for this thesis, are as follows.

1. We develop and evaluate a novel system that sensing modality to extend capabilities of commodity wireless sensing to include material identification.
    - Development of, to our knowledge, the first system that explores the use of commodity Wi-Fi radios in discerning objects in the environment using polarization.
    - Demonstration of a system that detects the material and location of occluded obstacles, opening up applications for autonomous UAVs and beyond.
    - System implemented and evaluated indoors and outdoors.
2. Fusion of wireless sensing with camera for improved depth resolution
    - Development of a novel approach to integrating mmWave and camera sensing to expand beyond each sensor's individual capabilities
    - System implemented and evaluated on 151 outdoor scenes

These contributions will be further highlighted in the rest of this thesis.

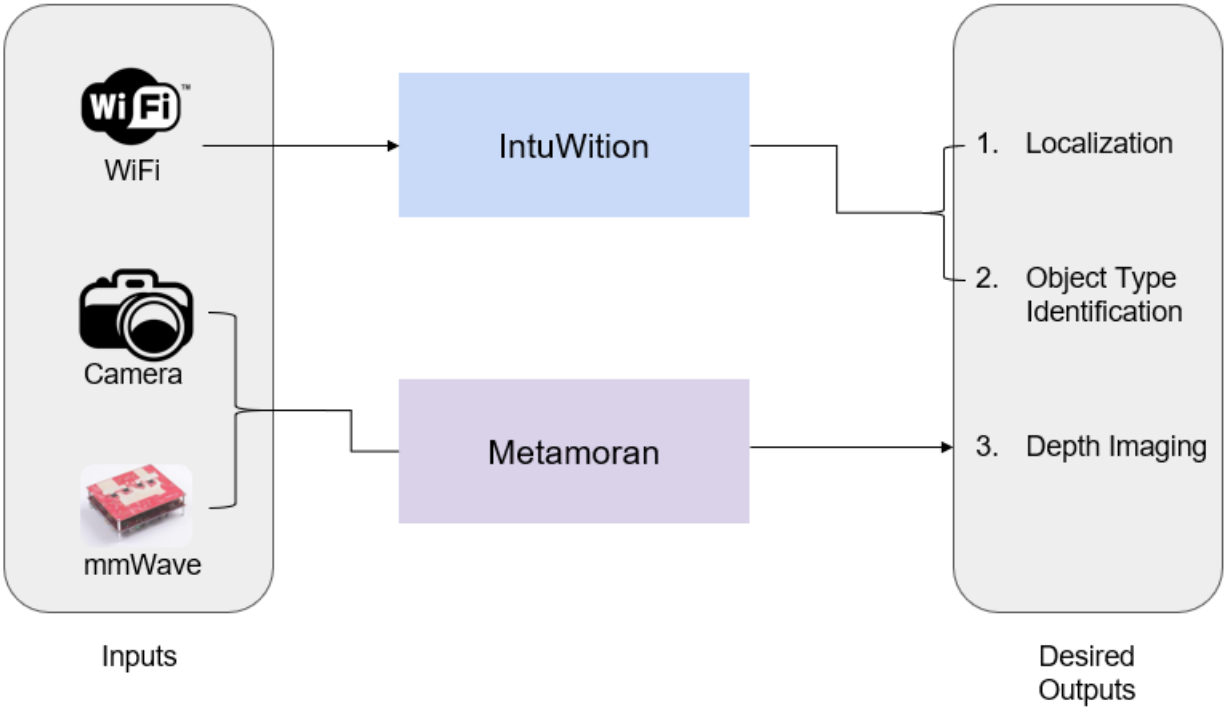

Figure 1.5: A system overview of our project, which takes as input WiFi, Camera, and mmWave to localize, determine object type, and image depth.

1.4 Thesis Overview and Outline

In this thesis, we will explore the scenario in Fig. 1.4. For a sensing system outfitted with some of the most ubiquitous wireless communication and sensing technologies: mmWave, WiFi, and cameras, how do we approach the challenges of determining the location of occluded objects, determining the type of an object, and developing depth maps for both a scene and objects-of-interest in the environment. In Fig. 1.5, we show a block diagram with the inputs and outputs of our scenario.

The rest of this thesis is organized as follows. Chapter 2 provides background information about capabilities of WiFi as a sensor and mmWave as a sensor to lay the foundation for this thesis and provides context for this work with respect to existing literature. In Chapter 3, we present our development and evaluation of IntuWition, a WiFi sensing system inspired by polarimetry to identify the composing material of objects in the environment. In Section 4, we present our development and evaluation of Metamoran, a mmWave and camera hybrid sensor that can provide long-range depth imaging of a scene. Finally, in Sections 5 and 6, we present our discussion, conclusions and future work.

Chapter 2

Background & Related Work

Though wireless signals are ubiquitous for communication, we instead explore these same signals for sensing. In this section, I provide some context on how wireless signals have generally been used for sensing.

2.1 The Communication Channel

A *channel* is the medium through which a transmitted signal is mapped to a given received signal. It has certain measurable properties, including noise, power attenuation due to distance, and other impairments. We define the communication channel $h(t)$, for a given transmitted signal $x(t)$, received signal $y(t)$, and noise $n(t)$, as follows:

$$y(t) = h(t) * x(t) + n(t) \tag{2.1}$$

One channel impairment particularly relevant to this dissertation is multipath. Multipath occurs when a transmitted signal undergoes multiple paths – that is, reflects off objects in the environment for paths of different lengths and attenuation. At the receiver, this results in the same signal received multiple times at different delays and different magnitudes, analogous to an echo. When considering multipath in isolation, the channel can be regarded as an impulse response:

$$h(t) = \delta\ (t) + a \times \delta\ (t - \alpha) + b \times \delta\ (t - \beta) + ... \tag{2.2}$$

Where attenuation factors $a, b, ...$ are less than one and delay constants $\alpha, \beta, ...$ are positive. Of course, the typical channel might be significantly more complex due to additional impairments beyond multipath.

The channel model applies regardless of the type of signal that is transmitted, and indeed even regardless of what the communication medium is: for example, multipath can occur due to internal reflections in a copper wire as well as due to environmental obstacles in a wireless channel, such as walls or vehicles. In this dissertation, we focus on the wireless channel, specifically for two technologies: WiFi and mmWave.

2.2 WiFi as a Sensor

WiFi is defined in the IEEE 802.11 family of standards. For our work, we specifically use IEEE 802.11ac, which provides high-throughput communication in the 5GHz band. In the 5GHz band, there is 150MHz of bandwidth available for communication, typically split into 20MHz channels[33]. Using these measured received wireless channels, relative position can be determined for a given transmitter/receiver pair. One common flavor of relative position is fingerprinting, where a receiver's position relative to a stationary access point is measured using known information about the environment – either from past measurements or modeling of waves. The disadvantage of this approach is its struggles with new and changed environments: to develop a new model takes some time. A helpful reference to better understand how fingerprinting works for WiFi systems is the WiFi system *RADAR* [11].

Another approach to positioning is angle-of-arrival (AOA), which is the relative angle between a transmitter and a receiver. AOA can be calculated using phase offsets between each element of a receiving antenna array. This receiving array can be multiple antennas connected to a single card or otherwise synchronized in some way, or a synthetic array emulated by motion (SAR). The use of beamforming algorithms, such as Bartlett[13] and MUSIC[88], further improve the performance of angle-of-error measurements. Multiple measurements need to be calculated to determine position rather than just an angle, through a process called triangulation. The disadvantage of this approach is that the antenna arrays for WiFi can become large and bulky: with an optimal antenna separation distance of $\lambda/2$, each antenna would have to be spaced a few cm apart (more is more accurate), and the distance traversed to calculate AOA via SAR might be fuel-consumptive depending on the vehicle. Arraytrack[114] and Ubicarse[54] are two contrasting WiFi systems that implement WiFi-based AoA, which would be helpful to better understand WiFi AoA in practice. Arraytrack uses a stationary array and the MUSIC algorithm, while Ubicarse uses SAR and a multipath power-based approach for determining angle-of-arrival. SpotFi [53] develops an approach to measuring AoA that relies on treating each WiFi subcarrier as a separate antenna element, with likelihood estimations assisted by time-of-flight.

Time-of-flight/time-of-arrival (TOF/TOA) is a approach to positioning that calculates the distance between a transmitter and a receiver. It has the advantage of requiring the least infrastructure and set-up

time of these approaches, requiring only one transmitter-receiver pair. The naive approach to time-of-flight: directly measuring how long it takes for a signal to reach an object and bounce back, typically requires synchronization much higher than a typical WiFi card is able, particularly at short indoor distances. Consider that detecting a sub-meter displacement would require measuring a time difference of a few nanoseconds – on a device that can take up to 122 μ S to transmit a single packet. Thus, WiFi localization systems often use intra-pulse modulation to measure time-of-flight. The resolution of time-of-flight using intra-pulse modulation is limited by bandwidth: resolution = $\frac{c(speed\ of\ light)}{2*Bandwidth}$. For a single 20MHz WiFi channel, this is a resolution of only 7.5m – much too coarse for most indoor applications. There have been many different approaches to improving time-of-flight performance in WiFi systems. Tonetrack [115] hops across 80MHz, uses multiple access points to improve performance, and develops MUSIC algorithm in the time domain. Chronos [105] uses a single access point but sweeps across all WiFi bands, developing how to handle non-consecutive WiFi bands.

The deployment and evaluation of WiFi-based sensing systems is further complicated by other factors, such as hardware offsets. In particular, even slight differences in timing and phase between devices can be a significant source of error in calculated distance. Some sources of error are constant sources of phase error, such as the time necessary to physically propagate through a wire and antenna, which can be removed with calibration. Other sources of error vary over time, such as carrier frequency offset, which occurs due to a transmitter and receiver having slightly different clocks. This is very common in WiFi cards, and indeed, in commodity wireless overall, since highly accurate clocks are expensive and bulky. Due to parameters such as temperature and intrinsic physical differences, crystal oscillators (a common cheap clock) vary in frequency within their given tolerance bounds. This small difference in timing becomes huge differences in distance measured when multiplied by the speed of light, and worse, the error accumulates over time. Most WiFi localization papers (included those cited above) will discuss how to accommodate these hardware offsets, but common solutions include using a reference antenna and using a forward-backward channel.

The ability to localize using WiFi has opened a wide variety of applications. Past work has leveraged variations in Wi-Fi signals for activity recognition [80, 107], occupancy sensing [26, 25, 6], imaging [42, 84, 48] and location tracking [47, 5, 113, 112, 53]. Indeed, a range of device-free applications have been proposed by fine-grained motion tracking of users, sensitive enough for even keystroke sensing [9] and UAV identification [69]. Much of this related work assumes multiple access points or bulky infrastructure in the environment and are therefore ill-suited for our application. Our work instead seeks to discover the location and material of multiple potentially occluded objects in the environment using moving Wi-Fi radios on a mobile platform without supporting infrastructure. Key to our approach is the use of

polarization of reflected signals to detect the material of objects using commodity Wi-Fi radios.

2.3 mmWave radar as a Sensor

The mmWave radar we use is in the automotive band – 77GHz. It transmits a frequency-modulated continuous wave (FMCW), or 'chirp.' This is a signal whose frequency varies with time. To detect range with an FMCW signal, a chirp is transmitted, reflected off an object in the environment, and then mixed with the original signal to get an "intermediate frequency" (IF) signal, which is a single tone. The initial frequency of the IF signal is the difference between the initial phase of the received signal first received at time t and the phase of the transmitted signal at the same time t. This phase difference, $\phi = 2\pi\ f\tau$, can be used to calculate the time of flight τ to objects of interest, given initial frequency f. FMCW has the same bandwidth limit to resolution: resolution = $\frac{c(speed\ of\ light)}{2*Bandwidth}$, but because mmWave has much more bandwidth available (4GHz at the 77GHz band), this translates to a resolution of 3.75cm. The full-duplex nature of FMCW radar removes offsets as an issue since the signal is received by the same device that is transmitting, at the same time that it is transmitting, making it a straightforward solution to ranging. Velocity can also be measured by transmitting multiple chirps and calculating how much range has changed in a known amount of time. Finally, angle-of-arrival is less bulky to calculate than is the case with WiFi, since antenna elements need to be spaced only a couple millimeters apart for an mmWave array. The principles of FMCW for radar have been largely unchanged for several decades, but this guide[19] released by TI might be helpful for understanding the basics of mmWave radar sensing.

Because mmWave radar already builds on a strong body of radar sensing principles and is well-developed as an approach, related research in this space tends to be focused on novel mmWave systems to address new applications. For example, recent work has explored mm-wave radars to sense the immediate environment of UAVs and autonomous cars [46, 124]. mmWave has also been explored for location tracking [108, 72] and breath/heart rate tracking [117]. Additional prior work has also explored high-resolution mmWave radar systems for through-obstruction imaging [34], security scanning [96] and predictive maintenance [66]. While complementary, these solutions are not designed to measure high-resolution depth images at extended distances, primarily due to the limited azimuth resolution of radar platforms.

2.4 Past work in Material Identification

Additional past work exists that uses commodity wireless devices [106, 27] that can detect material by sensing how signals are attenuated by objects placed between the transmitter and receiver. We distinguish ourselves from this work by focusing on reflection-based versus penetrative sensing, which makes our

approach easier to deploy on a single platform. WiFi has also been used to detect metal [110] by sensing how signals are reflected by environmental objects, though this is limited exclusively to metallic and non-metallic objects.

Our approach in is inspired by polarimetry, the measurement of how polarization of a signal changes upon interaction with objects in the environment. RADAR systems use synthetic aperture radar polarimetry for topography imaging [78, 90, 89, 16]. Aircraft and satellites can even use polarization of received RADAR signals to measure soil moisture of farmland [93]. Recent years has seen the persistence of polarimetry largely unchanged as a sensing modality but used with new materials [56] and for new applications [119]. In contrast to this rich literature, our approach seeks to bring radar polarimetry to light-weight commodity Wi-Fi cards. For this, we overcome many challenges such as the near-far effect and imperfections of such radios. We further distinguish ourselves from existing polarimetric systems in that our proximity is much closer than an airplane or a satellite, where polarimetry systems are commonly deployed. We choose this approach over other radar-based object recognition techniques, such as Project Soli, Google's RADAR-based system for object & material recognition [118] which assumes extremely close proximity between the object and the RADAR system.

2.5 Past work in Long-Range Depth Imaging

Current vision-based imaging solutions include Cameras [65], LIDARs [82] and depth imaging [41], which are often used in diverse outdoor 3-D imaging applications. Some depth camera systems (e.g. monocular depth estimation [15]) struggle at extended distances, some (e.g. stereo-vision [94]) require extended baselines for high accuracy, while others (e.g. IR structured light [85]) function poorly under ambient light. More broadly, systems struggle to measure depth at a high resolution at long range, with about meter-scale accuracy at up to 80m range in monocular depth estimation cases [122] and only operating up to around 20m in the case of depth cameras [98]. Some LIDAR systems [76] offer higher accuracy at extended ranges, however face other significant limitations stemming from the power consumption of the laser as well as robustness to dust, weather conditions and coexistence with other LIDAR platforms [17, 51]. There also exist infrastructure-based solutions [55, 31, 81] which we seek to avoid to preserve simplicity.

For our approach, we draw from past work in Camera/RF fusion for our hybrid system, Metamoran. Camera and RF fusion has been proposed for automatic re-calibration [116], industrial workplace [86], localization [7], person identification [29] and fall detection [50]. Radar-Camera fusion has also been studied for diverse vehicular applications including attention selection to identify objects-of-interest [123, 37, 20], tracking mobile objects [63, 92, 121] better object perception and classification under

poor weather [49, 45, 38], detecting vehicles and guard rails [99, 8, 44] and generating obstruction-resilient 2D images [57]. Vision-based sensing has also been used for more effective communication using mmWave [35, 70]. Beyond radar and vision, prior work has used multi-modal fusion across a variety of sensors for tracking human activity [58], autonomous driving [22] and beyond. We distinguish ourselves from this body of work by focusing on combining mmWave radars and camera for high-resolution depth imaging at long ranges, including under partial occlusions.

Chapter 3

IntuWition: On the Feasibility of WiFi-Based Material Sensing

IntuWition addresses the vision of this thesis from the perspective of: what can we do with WiFi alone? Cameras and RADAR have comparatively well-explored capabilities, given decades of research in large fields of study. WiFi as a sensing modality has been less explored, although as discussed in Ch. 2, it has also been explored for our first desired output: localization. So, how can we expand WiFi's sensing capabilities to address our second two desired outputs: object type identification and depth imaging? Unfortunately, the resolution we can provide with WiFi (around a meter), makes it impractical for depth imaging. So in this chapter, we present our approach to sensing object *type* via detecting material. The parts of our overall vision contributed by IntuWition are shown in Fig. 3.1. It takes as input the measured

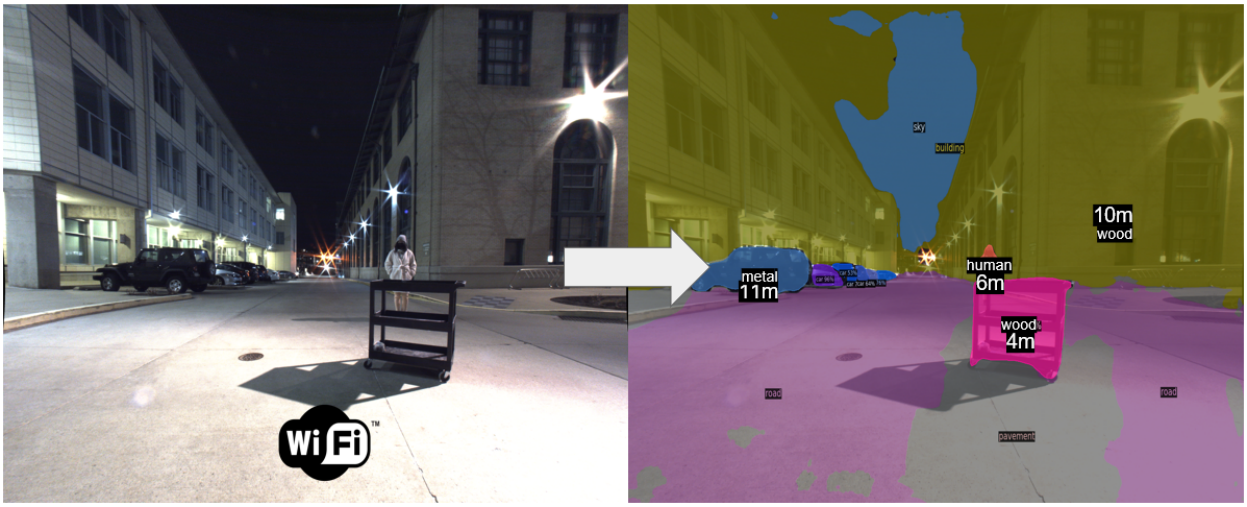

Figure 3.1: IntuWition is a first attempt to achieve our vision of localization and object type identification using WiFi alone.

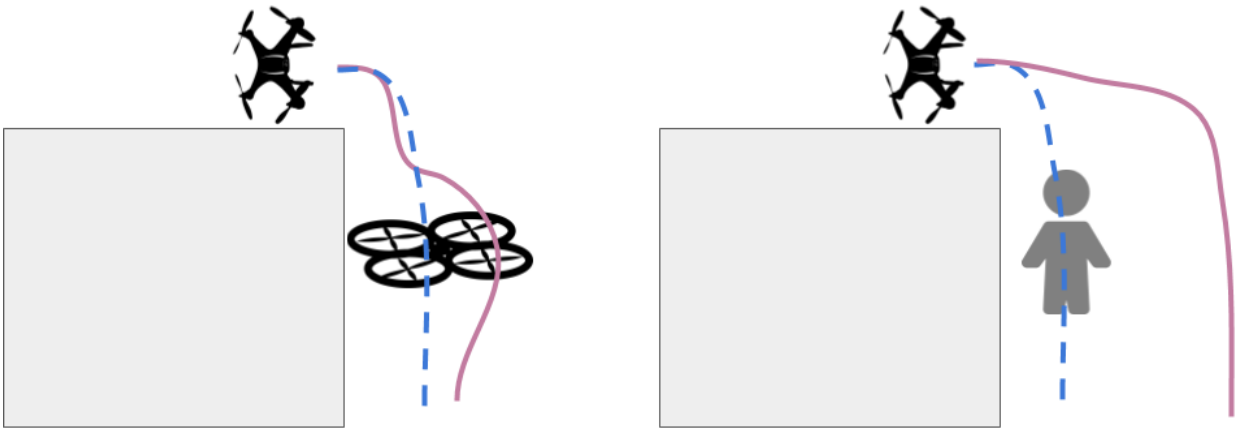

Figure 3.2: IntuWition helps detect the type of material of an object hidden from view. This could assist, say a drone's path planning algorithm, to change its planned path (blue) to a new path (purple). For example, it could swerve upwards to avoid another drone, but would give a human a wide berth.

WiFi channels of a scene, and produces an output of the location of objects in the environment, as well as assigning the object one of three type labels: human, wood, or metal.

3.1 Motivation and Approach

Recent years have seen an explosion in wireless sensing and tracking research, from people to emotions to objects. However, most existing work in the wireless sensing space do not sense the precise material composition of the objects they track. Yet, in many occasions wireless radios may need to be aware of the type of objects in the environment, beyond their location. For instance, consider autonomous UAVs that may choose to use their onboard Wi-Fi radios to detect if an object around the corner is a person or another drone. This could allow it to respond differently in either case – for example by giving the human a larger berth versus moving up vertically to avoid the drone (see Fig. 1). Beyond UAVs, object type sensing could transform ordinary Wi-Fi devices into sensors that identify objects around the corner, with applications in search-and-rescue, smart occupancy sensing, vehicular safety and beyond.

Unfortunately, state-of-the-art sensing solutions fall short of localizing and discerning occluded obstacles. For instance, cameras can both localize and identify objects, but operate solely in line-of-sight. This could be addressed with high-resolution aircraft RADARs; however, these bulky devices are not portable and are highly expensive [3]. Further, newer radars designed for flight are often at high frequencies for higher resolution, but this comes at the cost of reduced ability to propagate through and around materials and 6x the power consumption over a typical Wi-Fi router [2]. Wi-Fi based sensing systems [5, 80] can penetrate walls and obstacles unlike light to track hidden objects, but have not been comprehensively

explored for material sensing. In this chapter, we ask: "Can commodity Wi-Fi radios sense the location and type of moving objects?"

We present IntuWition, a system that explores the feasibility of identifying both the material and location of surrounding objects using commodity Wi-Fi. We envision IntuWition as a complementary sensing modality to add non-line-of-sight and materials-sensing capabilities to existing sensing systems. We deploy Wi-Fi radios indoors and outdoors on a college campus to explore the materials sensing capabilities of WiFi. IntuWition processes wireless channels between pairs of commodity Wi-Fi radios to accurately estimate and track the location and material of different objects in its surroundings, including wooden, metal, and human obstacles. We further present a proof-of-concept application demonstrating IntuWition's performance on autonomous UAVs in distinguishing between occluded humans vs. other drones both indoors and outdoors.

At the heart of IntuWition is an approach that infers material properties of objects in the environment by measuring the wireless signals that reflect off them. In particular, it captures a specific property: the polarization of the reflected waves from an object. When a polarized incident wireless signal reflects off the surface of different objects, it behaves differently based on the texture and material of the objects. Specifically, metal absorbs and re-emits the wave, resulting in a change of polarization pattern; smooth surfaces reflect the waves with polarization intact; and rough surfaces scatter the waves and diffuse polarization. In effect, this causes a change in the observed polarization of the reflected wave. We demonstrate that by measuring the polarization of reflected waves from surrounding objects, one can infer the material they are composed of. We draw inspiration from the remote sensing community, where aircrafts image the topography using RADAR polarimetry [78, 93] to distinguish between trees, open fields and even infer soil moisture using reflected signals.

To bring RADAR polarimetry to commodity Wi-Fi radios, IntuWition measures the Wi-Fi signal from a vertically-polarized transmitter antenna to three mutually-perpendicular polarized receiving antennas, all of which are atop a Wi-Fi enabled device. We then compare the measured power of the signal across these antennas to infer its material composition. IntuWition builds a multi-layer perceptron model to achieve this independent of other parameters that influence received signals such as the size, thickness, texture, distance, and angle of an object. A key challenge in IntuWition's design is the ability to detect and discard signals that are reflected and scattered by multiple objects sequentially and therefore experience a hybrid change in polarization that corresponds to none of their underlying materials. IntuWition achieves this by reverse-engineering and recognizing the unique ways in which a Wi-Fi signal's polarization compounds when it bounces off many objects (see Sec. 3.4).

A second challenge IntuWition addresses is to separate the signals that reflect exclusively off each

individual object in the environment in 3-D space, prior to analyzing its material composition. IntuWition does so by processing these signals to measure the time-of-flight they experience as they reflect off each obstacle over brief slices of time. Additionally, mobility of the sensing autonomous systems can allow for views the same object from multiple perspectives. These samples further allow us to triangulate its 3-D physical position and to filter out the object's reflected signal for material classification. A key challenge in doing so pertains to the hardware non-idealities of commodity wireless radios that induce random errors in the measured time-of-flight owing to arbitrary timing offsets between the Wi-Fi transmitter and receiver, both of which are two distinct radio chips with different clocks. IntuWition addresses this challenge by simultaneously sending signals from the transmitter along two pathways: a wired pathway and a wireless pathway, only the latter of which is impacted by the environment. We then estimate timing offsets from the wired pathway to correct for its effect on the wireless pathway. The rest of this chapter describes how IntuWition addresses this and other hardware non-idealities such as carrier frequency offsets, sampling offsets, and phase shifts between the transmitting and receiving RF chains. Further, we discuss the limitations of IntuWition in Sec. 4.10: It often misses small, fast-moving or well-shielded objects, and cannot tell apart materials with similar polarization characteristics (e.g. two different humans).

We implement IntuWition on Intel Galileo boards equipped with commodity Intel 5300 Wi-Fi cards [36]. We perform a detailed feasibility study in large indoor and outdoor spaces in a university campus and distinguish between common material types: copper, aluminum, humans, plywood and birch, in a variety of line-of-sight, non-line-of-sight, stationary, and mobile settings. We further mount our Wi-Fi platform on a UAV (DJI Matrice 100) to classify between humans and drones for indoor and outdoor testing. Our results reveal the following:

- Our system classifies between 5 types of material (copper, aluminum, plywood, birch, and human) of a variety of sizes and orientations at an accuracy of 95% in line-of-sight and 92% in non-line-of-sight settings.
- Our system can classify objects that are 0.42 m apart in line-of-sight and 0.55 m apart in non-line-of-sight.
- Our system classifies between humans and drones with accuracy averaging 89% at UAV speeds of up to 2 m/s in dynamic indoor and outdoor settings.

Contributions: This chapter presents IntuWition, to our knowledge, the first system that explores the use of commodity Wi-Fi radios in discerning objects in the environment using polarization. We demon-

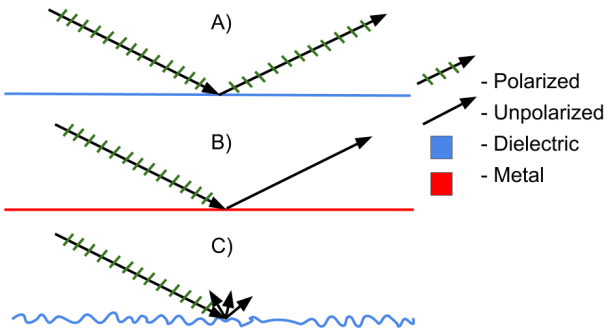

Figure 3.3: Scattering off (A) a smooth surface, (B) a smooth metallic surface, and (C) a rough surface produces differently polarized signals.

strate how IntuWition detects the material and location of occluded obstacles, opening up applications for autonomous UAVs and beyond. IntuWition is fully implemented and evaluated indoors and outdoors.

3.2 Polarimetry

This section provides a brief primer on how capturing the polarization of signals that scatter off an object provides information about the material it is made of. Fig. 3.3 illustrates three extreme instances of scattering responses for a perfectly polarized incident wave: (A) The incident wave reflects off a perfectly smooth non-metallic surface, maintaining linear polarization. (B) The incident wave reflects off a perfectly smooth metallic surface, eliminating linear polarization [52]. (C) The incident wave scatters off a perfectly rough surface, which may or may induce polarization based on the material. Hence, by measuring the polarization and power of the received waves, we can distinguish objects based on texture and material (metallic or otherwise).

To better understand this phenomenon, we consider a 2-D case for simplicity, a surface defined by parallel and perpendicular vectors p and s, respectively. As described in [39], an incident s-polarized signal can be denoted as $E_s(t) = Ae^{j\omega t}\hat{e}^s$. Upon reflection off a surface, amplitude, phase, and orientation can all be affected, resulting in a reflected signal $E_r(t) = Ae^{j\omega t}(r_{ss}\hat{e}^s + r_{sp}\hat{e}^p)$, where r_{ss} and r_{sp} are complex numbers describing how the s-oriented incident wave has changed in phase and power along each of the s and p components. Note that depending on the two phase value shifts, the locus of our signal might change from linear polarization to circular or elliptical – if they change by the same amount, or by a factor of π, linearly polarization is preserved.

Because of perpendicular linearly-polarized receiving antennas, IntuWition measures $\hat{e}^s$ and $\hat{e}^p$ separately, thus being able to measure material-specific parameters r_{ss} and r_{sp}. Having antennas that allow us to receive different polarizations of data opens the door to polarimetry. Furthermore, receiving these

signals separately allows us to take the ratio of these signals to remove power and phase shifts due to propagation, which affects signals similarly regardless of their orientation.

Consider the incident vector $E_I(t) = acos(2\pi ft + \phi_a)\hat{x} + bsin(2\pi ft + \phi_b)\hat{y}$, where a and b are amplitude quantities, f is our frequency, t represents time, and ϕ_a and ϕ_b are phase offsets that determine the locus of our signal – that is, its polarization. Because the incident signal is vertically polarized, this simplifies our form to $E_I(t) = bsin(2 + \phi_b)\hat{y}$. Upon interaction with a material, the transmitted signal

Mathematically, we capture the effect of scattering using vertically (V) and horizontally (H) polarized antennas at the transmitter and receiver [28]. Let the channels between the two transmit and receive antennas be h_{HH}, h_{HV}, h_{VH} and h_{VV} forming a matrix:

$$H = h_{HH}h_{HV}h_{VH}h_{VV}$$

The above matrix captures the polarization of the received signal, for instance should polarization be perfectly preserved the H matrix would have close to zero off-diagonal elements, while an unpolarized signal has elements of similar power on average. Polarimetry captures this effect using a quantity α which is defined as the weighted mean of the orientation of the eigenvectors of the covariance matrix of H weighted by its eigenvalues (see [79] for a detailed definition and derivation). Intuitively, α measures the angle of how far the scattering mechanism is from an ideally smooth non-metallic reflector. A small α indicates a smooth non-metallic reflector, $\alpha = \pi/2$ indicates an ideal metallic reflector, while α takes intermediate values for all other materials.

We note that in practice this property is influenced by the location, geometry, orientation and texture of reflecting surfaces in the environment besides material composition. The following sections present models to process wireless channels to decouple the reflector's material composition from all the remaining quantities.

3.3 Overview

IntuWition's objective is to allow a Wi-Fi enabled mobile devices to obtain both the 3-D location and the material composition of surrounding objects, including those occluded from view. It aims to do this using existing commodity Wi-Fi radios on the drone, without requiring supporting wireless infrastructure in the environment. In other words, an IntuWition device must both transmit and receive Wi-Fi signals and analyze them to report the location and material of surrounding reflectors. Fig. 3.4 presents an illustrative workflow of IntuWition.

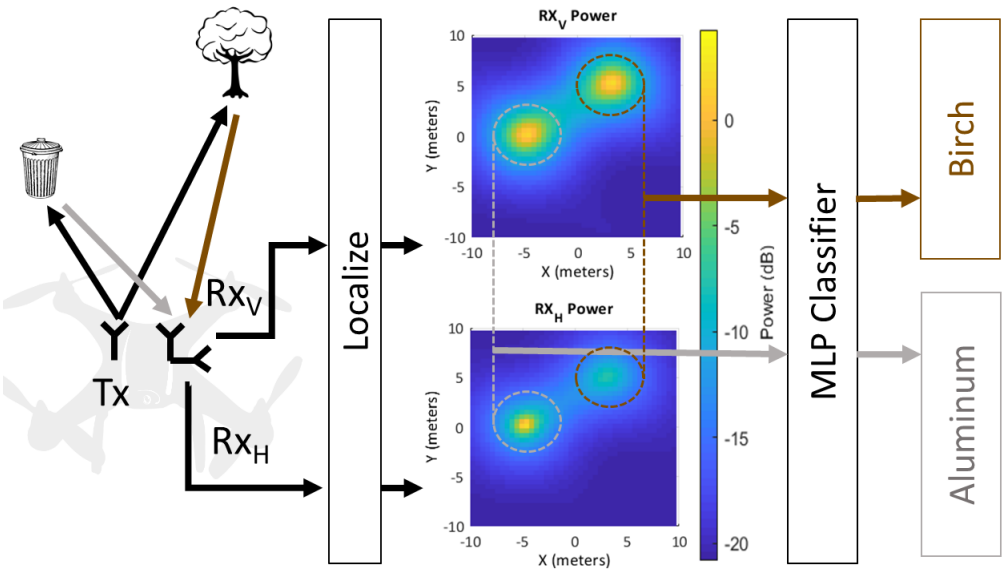

Figure 3.4: IntuWition Workflow: An example in 2-D (for simplicity) shows how IntuWition isolates locations of two objects and compares powers across horizontally and vertically polarized antennas to infer material type.

3.3.1 Learning Material Composition

IntuWition determines the material composition of any object-of-interest in the environment by studying the polarization of the signal reflected from it. Specifically, IntuWition first uses the known time-of-flight of the reflected signal from an object to develop a filter that outputs the wireless channel component that corresponds solely to the reflected path. It then processes this signal component on a horizontal and vertically polarized receiving antenna and measures their ratio to determine the extent to which the reflection impacted the polarization of the transmitted signal. IntuWition measures this ratio across a range of frequency bands, developing a vector that captures the impact of the material on signal polarization. We then build a model based on multi-layer perceptrons to classify the material and type of the object-of-interest. We also compare our model with five other typical machine learning models.

Extracting Features: Among the key challenges in material classification is identifying the best features that are unique to materials. IntuWition relies on the change in polarization of the received signal from the transmitted signal across different frequencies. Specifically, we filter different *subsets* of the wireless channels, each spanning a different range of frequencies to obtain the wireless channel component reflected from the object-of-interest. We then compare these channels at the horizontal and vertical antennas to study their polarization across subsets of frequency bands. Finally, we feed the ratio between channels at the horizontal and vertical antennas into our model. IntuWition leverages a Multi-Layer Perceptron model to process this feature vector and infer the object material and type, while remaining robust to

device movement and signal multipath. Sec. 3.4.1 details IntuWition's learning models.

Sensing Mobile Objects over Time: Perhaps the most challenging objects to track are objects that are themselves moving with time. For instance, consider two objects whose trajectories intersect. While IntuWition can identify an approximate relative location of each moving object along their respective trajectories, it may erroneously swap the trajectories of the two when they intersect. Sensing materials of objects provides a natural solution to this problem. Specifically, IntuWition can use the polarization of reflected signals from each object to identify it and accurately track its position over time without ambiguity. Sec. 3.4.2 describes how we exploit the synergy between localization and material sensing to improve each other's performance.

3.3.2 Locating Surrounding Objects

To accurately locate surrounding objects, IntuWition analyzes the wireless channel-state-information available at a compact commodity 3-antenna Wi-Fi receiver from a co-located Wi-Fi transmitter on the drones. These channels are a combination of signals propagating along different paths as they reflect off various surrounding objects, as well as the direct path between the transmitter and receiver. IntuWition processes these wireless channels across Wi-Fi frequency bands and measures the time-of-flight experienced by the signals along each path. When multiplied by the speed of light, this provides the distance traversed by the signal from the transmitter to the reflector and then back to the receiver. By computing these distances from multiple perspectives as the drone flies in the 3-D space, IntuWition triangulates the 3-D position of the object-of-interest.

While there has been much work on localizing surrounding objects using RADAR [32, 105] and recent work on wireless material sensing applications [110, 106, 27], bringing it to inexpensive and commodity Wi-Fi radios on mobile systems without external supporting infrastructure brings forth several new challenges that IntuWition must address:

Disentangling Signal Paths: First, IntuWition must separate signal paths from various objects surrounding the mobile device. IntuWition achieves this by actively exploiting the mobility of the device itself (e.g. a moving UAV or tablet). Sec. 3.5.1 shows how by processing the received wireless channels across spatial locations and frequency of operation of Wi-Fi radios, one can simultaneously obtain time-of-flight and angle-of-arrival of surrounding objects.

Locating Occluded Objects: Second, IntuWition must develop algorithms that can analyze weak reflections from distant objects and those that are occluded by other objects (e.g. walls, trees, etc.). IntuWition achieves this by sensing changes in the wireless channels from weak reflections to detect the presence of

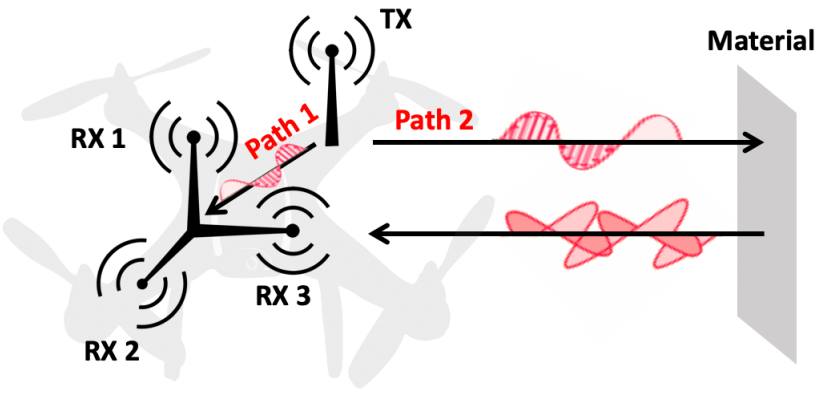

Figure 3.5: IntuWition transmits signals from a vertically polarized antenna to 3 mutually perpendicular receive antennas and processes polarization changes due to reflections from materials.

the moving objects that produce them. A key challenge in implementing this is to filter out the movement of the drone itself, as it also changes wireless channels, over and above the movement of the object-of-interest. IntuWition therefore develops a novel background subtraction algorithm that eliminates the effect of the drone's movement in the wireless channels to detect the movement of surrounding objects. Sec. 3.5.2 describes our approach in greater detail.

3.4 Material Sensing using WiFi

This section describes how IntuWition senses the material properties of a surrounding object. We assume the received signal has been pre-processed to localize and isolate the signal component from a specific object of interest (detailed in Sec. 3.5). We describe how we bring RADAR polarimetry to the Wi-Fi context, while mitigating the limitations of commodity Wi-Fi radios.

Hardware Setup: IntuWition's setup consists of four Wi-Fi radios – a transmitting device broadcasts packets from a vertically polarized antenna to three receiving light-weight Wi-Fi chips, each with a mutually perpendicularly polarized receiver antenna on port one. Note that inexpensive omni-directional whip antennas can be used as vertically polarized antennas because of their natural propagation pattern. We connect a wire to the transmitter and, via a splitter and attenuated cable, connect this to antenna port 2 on each receiving chip. This provides resilience to phase errors from hardware offsets (details in Sec. 3.5.1).

Fig. 3.5 shows an illustration of how the polarization pattern of signal is changed after reflections from a material. If no reflector is present in the environment, the three mutually perpendicular receiver antennas $RX1$, $RX2$, $RX3$ will only receive the single signal polarization pattern of the direct path ($Path1$). In the presence of a reflector, the receive antennas also obtain signals along $Path2$ whose pattern of

polarization depends on the material of the reflector.

IntuWition uses three receiving antennas instead of two as is used in the traditional RADAR polarimetry because our transmitter and receiver are at about the same elevation level of the obstacle we are sensing. As a result, modeling the sensed object as a 2-D reflecting plate (as is the case from a satellite or a high plane) is no longer valid. Even when movement is along four controllable degrees of freedom (say on a quadcopter drone), three antennas are sufficient to collect information about scattering from various surrounding objects (including the floor), despite the movement of the platform itself.

Approach: IntuWition employs Multi Layer Perceptrons to classify different materials for a given object. Our system takes as input the polarization of the reflected signal from a given object across frequencies. It then extracts features via a multi-layer perceptron that closely correlate with material composition of the object. A key challenge we face in designing our material classification algorithm is to decouple various other factors that influence the wireless channels, such as object size, texture and orientation. The rest of this section details our solutions to deal with these challenges as well as our choice of features and classification algorithm.

3.4.1 Building the Machine Learning Model

Prior to extracting features, IntuWition uses its object localization algorithm (details in Sec. 3.5) to filter the wireless channel from an object of interest. In particular, it measures this channel at all receive antennas across OFDM subcarriers. Further, IntuWition repeats the localization algorithm over many different subsets of Wi-Fi operating frequencies to obtain the wireless channel component reflected off the object. In doing so, IntuWition retrieves a vector of wireless channels per-antenna from the object versus frequency – both across different subsets of frequency bands and subcarriers.

Next, IntuWition computes the ratios of the channel received between pairs of antennas (Fig. 3.5). By doing so, we eliminate the effect of distance of the object on signal power and focus instead on polarization, which is highly correlated with the material composition of the object. The resulting three channel ratios sampled across frequencies form a vector of over a thousand elements. Feeding all of these as features into our machine learning model would be counter-productive, given that it requires a complex higher-dimensional classification model that is highly susceptible to over-fitting.

However, we find that channel ratios sampled across all frequencies are not all equally important as features, owing to differences in the quality of channels (e.g. channel fading and interference on some frequencies) or the number of measurements available. Hence, IntuWition selects a subset of features from the wireless channel ratios across frequencies. We use a greedy algorithm to rank features according

to number of successful measurements across the whole channel and select the top 200 ones. Then we remove the samples which have missing values among those 200 features. In doing so, we can reduce the number of features to avoid over-fitting.

While a range of machine learning classification models are available, IntuWition employs Multi-Layer Perceptrons (MLPs) to train its classifier based on a theoretical analysis and an empirical comparison. MLPs have been widely observed to generalize well compared to hand-crafted features in recent deep learning literature [95], since they automatically learn features from input data. We construct two versions of MLPs which take raw feature values as input and outputs either the material class or an object type. The latter pertains to our proof-of-concept demonstration of our system on a UAV, which classifies between two classes of objects – drones and humans who may have multiple (yet different) classes of materials on their surface.

Besides MLP, we considered alternative machine learning models. Sec. 4.9 presents an empirical comparison of various machine learning models: MLP, RBF-SVM, *k*-NN, PCA, and Naïve Bayes, and shows that MLPs provide the maximum classification accuracy for us.

Training and Testing: We train the model by feeding the top 200 features to Multi-Layer Perceptrons. These features were selected as the most impactful out of all the power ratios between pairs of antennas captured at frequencies across large bandwidth. To avoid the overfitting of the training model, we collect the training data in different radio environments (indoor and outdoor), across multiple days. We collect the data of both sheets of different materials and different textures of real objects like chairs. We train the model for 500 epochs, where one epoch is a full pass of the data through the neural net. We randomly shuffle the training data after each epoch. Our batch size is 32 and we randomly select one batch at a time from the training data sets to feed into the neural nets. We divide the entire dataset into separate training and testing dataset to evaluate the classification network. Our MLP contains three hidden layers with each layer having 200 units and using a rectified linear unit as the activation function. The network is trained to optimize for cross-entropy loss using the Adadelta algorithm [120].

3.4.2 Tracking and Sensing Multiple Objects

Our discussion so far has exploited localization to isolate and find the material that objects in the environment are made of. However, sensing the materials of objects can also improve the performance of localization of these objects. To see how, we consider a challenging problem in device-free location tracking: tracking multiple objects whose trajectories intersect.

Ambiguous Trajectories: Fig. 3.6 shows an example of two objects – a pedestrian and a bicycle whose

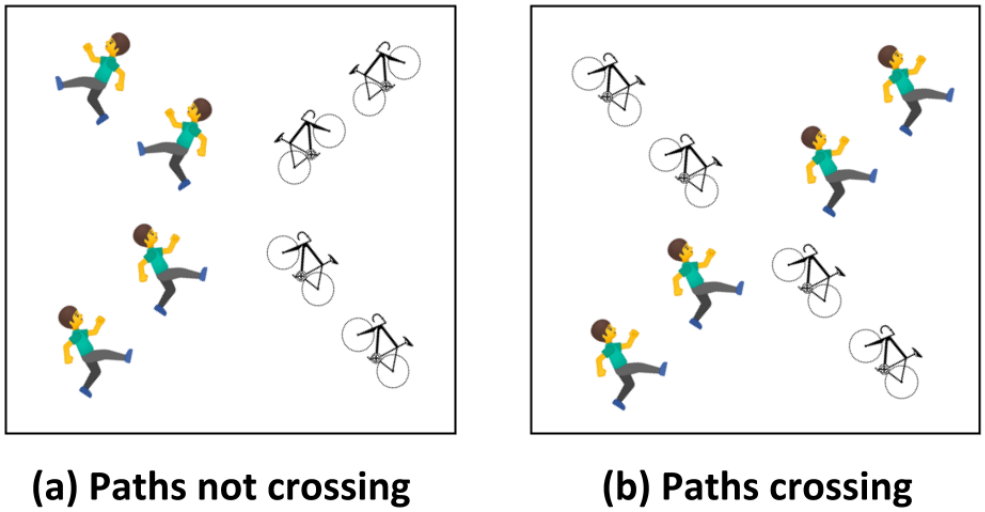

Figure 3.6: IntuWition can track objects with ambiguous intersecting paths by tagging objects with their respective materials.

trajectories intersect at a crosswalk. Traditional wireless device-free localization systems struggle to distinguish between the two trajectories in Fig. 3.6(a) and Fig. 3.6(b) – both of which result in an identical set of observed times-of-flight of reflectors.

IntuWition can resolve this challenge by exploiting a very simple property – while the location of an object can change over time, its material composition cannot. As a result, IntuWition can use the measured material composition of the two objects that are tracked to distinguish between the trajectories in Fig. 3.6(a) and (b). An important limitation of this approach is that it cannot disambiguate objects of precisely the same material. We believe that sensing other properties (e.g. size, shape, etc.) along with material can help address this limitation – a task for future work.

Spurious Multi-Object Reflections: A second challenge that IntuWition resolves is to eliminate spurious reflections off multiple objects before reaching the receiver. The resulting polarization observed is a combination of the properties of all materials through which the signals reflect. Specifically, the phase shift of $r_s s$ and $r_s p$ values of the signal received (see Sec. 3.2) that captures signal polarization is the sum of the $r_s s$ and $r_s p$ values of each material. We note that signal absorptions cause limited impact on polarization [21].

IntuWition addresses this challenge by actively modeling multi-object polarization. In particular, IntuWition sorts objects based on their observed-value of time-of-flight (i.e. a measure of distance) relative to the vehicle. It then progressively labels each object as legitimate, only marking objects as spurious if its phase shift is the sum of subsets of legitimate phase shifts. At a high level, this is because the only propagation effects that affect the horizontally and vertically polarized signals asymmetrically at the re-

ceiver are due to reflections from obstacles, and the effect of the final received signal can be modeled as the product of the transmitted signal along with each reflector. This translates to a sum in phase effects. Thus, IntuWition quickly eliminates spurious reflections off multiple objects.

3.4.3 Material Properties vs. Other Effects

From atop a mobile platform, our system must deal with materials at various ranges, sizes, orientations, and textures. Our system must therefore extract the effect of material properties of the object from wireless channels from the variety of other properties of surfaces that influence channels: object size, orientation, texture and location.

Effect of Location: In handling objects of different sizes, ranges and orientations, measuring changes in polarization actually emerges as an advantageous approach when compared with measuring pure power. While reflected power could certainly distinguish a wooden and metal sheet at the same distance, it remains vulnerable to multipath and is subject to ambiguity, e.g. between a nearby wooden sheet and distant metal sheet. We instead use power ratios of the channels from a vertically polarized transmitter received by three perpendicular linearly polarized antennas. This ensures that the distance traversed to the reflector is effectively factored out of our ratio reading.

Effect of Object Size and Orientation: The challenges of varying reflector sizes and angular orientations would pose a similar issue for power-based material identification. These properties are captured in radar ranging by a property σ which they call the Radar Cross Section (RCS). Also known as the effective area, RCS is the value associated with what size an object's reflection, independent of what its dimensions actually are, appears to a radar sensing system. Angle of orientation and physical size of this object both affect this, as intuitively, a 4'x4' sheet oriented at 30 degrees would appear larger than one oriented at 45 degrees. Our multi-antenna approach to measuring rotations in polarization would be affected by the same effective area across all three antennas, thus removing the effect of size and angular orientation from isolating the material.

Effect of Surface Texture: We note that surface texture and material composition can be decoupled by looking at a window of reflected values, instead of a single point. Polished surfaces show a distinct peak at the time of flight of the reflected object (along with a few later peaks due to multipath effects). On the other hand, in rough volume reflectors, signals transmitted through the initial surface but have multiple opportunities to reflect through the width of the object, resulting in a shorter, wider peak.

Given the entire spectrum of possible surface textures, we could not exhaustively test this space, but to help account for this effect, we train our machine learning models (Multi-Layer Perceptron) on diverse

textures per material - such as different finishes of metal, or a wooden sheet sanded by different grit sandpaper. We present our results in 3.8.3.

Effect of Incident Angle: The incident angle affects polarization by the well-studied Fresnel Equations.[52]. We train our models with many different angles to help account for this effect.

3.5 Localization using WiFi

Prior to sensing the material of an object-of-interest in the environment, IntuWition needs to process the received wireless channel to separate the wireless signal component reflected off this object alone. IntuWition achieves this by first finding the location of the object-of-interest, using signals from an object, collected from different perspectives, to triangulate its position.In the case of multiple objects, we remove peaks using successive interference calculation and then remove spurious multi-object reflections based on their r_{ss} and r_{sp} values, as described in 3.4.2. IntuWition then develops a filter to extract the wireless signal component that emerges from that particular location. The rest of this section describes IntuWition's approach to locate surrounding objects.

3.5.1 Separating Signal Paths in Mobile Settings

IntuWition's first task is to separate the signal components arriving along multiple paths as they reflect off each object in the environment. At first blush, IntuWition may achieve this using prior work on RF localization using commodity Wi-Fi radios [105]. Specifically, past systems seek to separate the time-of-flight of different signal paths from the received signal, as they reflect off various objects in the environment. An important factor that determines the resolution of measured time-of-flight of various signal paths is the available bandwidth (20 MHz for Wi-Fi). To mitigate this, past work [105] stitches together wireless channel measurements across multiple frequency bands to emulate signals from a wide-band receiver. However, an important challenge in combining frequency bands in mobile contexts such as the UAV is that the device moves in the 3-D space so that its location changes significantly between channel measurements. Hence, channels across packets change both due to change in Wi-Fi frequency as well as the change in device location between measurements.

IntuWition addresses this challenge by actively modeling both the frequency of operation and the movement of the device (e.g. on a drone) in its analysis of the wireless channels. In doing so, it retrieves both the angle-of-arrival and the time-of-flight of reflecting surfaces simultaneously. The rest of this analysis makes two simplifying assumptions for ease of exposition, which we relax later in this section: (1) The device moves in 2-D space; (2) phase errors due to frequency and timing offsets are accounted for.

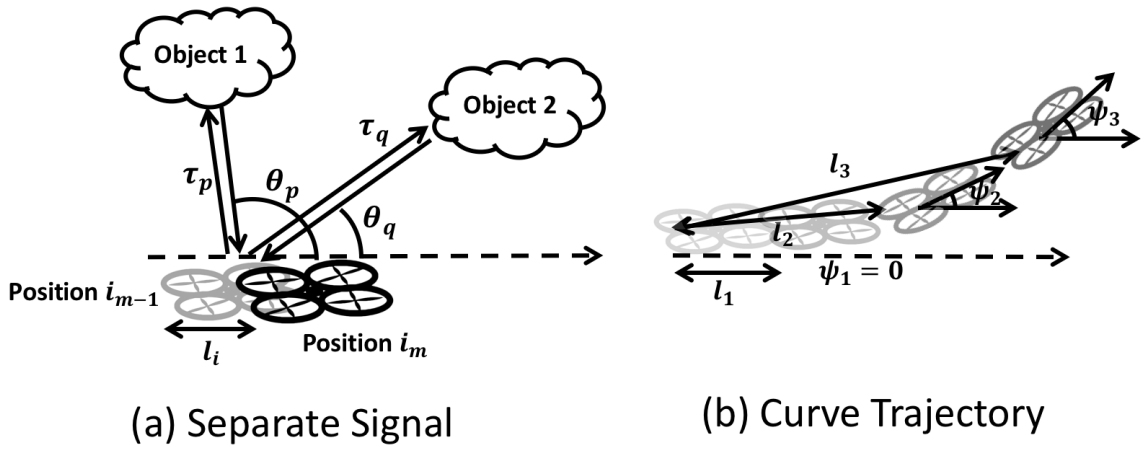

Figure 3.7: (a) IntuWition measures the distance and orientation of objects of interest using the wireless channels that reflect off them. (b) Notation for device (e.g. on a UAV) location along curved trajectories.

Mathematically, assume that the device (Fig. 3.7) moves on a 2-D trajectory that is at a distance l_i relative to its initial position, ψ_i relative to its initial orientation and on frequency f_i when receiving packet i. We assume that the distance between the reflector and the device is much greater than the displacement of the vehicle across packets, allowing for us to approximate that the angle of the reflector does not change. To recover the time-of-flight and location of objects in the environment, IntuWition builds upon the Bartlett algorithm [54] in 2-dimensions. Specifically, we can write the power of the wireless channel received along a distance r and angle-of-arrival θ as: $\mathrm{P(r,}\,\theta) = \left|\sum_i h_i e^{2\pi j f_i \left(\frac{2r}{c} + \frac{l_i \cos(\theta - \psi_i)}{c}\right)}\right|^2$

The above power-profile will have peaks corresponding to the polar coordinates (r, θ) of various objects in the environment relative to the vehicle. We can then extract the wireless channels corresponding to any specific object of interest at (r, θ) as: $\mathrm{h(r,}\,\theta) = \sum_i h_i e^{2\pi j f_i \left(\frac{2r}{c} + \frac{l_i \cos(\theta - \psi_i)}{c}\right)}$ We show in Sec. 3.4 how one can detect the material of an object at (r, θ) by analyzing $h(r, \theta)$ at the horizontally and vertically polarized antennas across frequency.

While the above analysis assumes that the available Wi-Fi frequencies and distances moved by the device between packets are uniformly spaced, in practice, this may not be true, leading to spurious peaks in $P(r, \theta)$. IntuWition builds on past work on wireless localization [105] to eliminate these peaks by leveraging the sparsity of signal multipath. Specifically, we assume that the signals that reflect off the environment emerge from a small number of dominant paths, leading to a sparse $P(r, \theta)$ with a few distinct peaks. Mathematically: $\min_{\{\forall p: A_p\}} \sum_p |A_p|$,
$\sum_i \left|h_i - \sum_p A_p w_p\right|^2 = 0$
$w_{i,p} = e^{-2\pi j f_i \left(\tau_p + \frac{l_i \cos(\theta_p - \psi_i)}{c}\right)}$ This optimization resembles a non-convex 2-D Non-Uniform Discrete Fourier transform (2-D NDFT). IntuWition solves it numerically using proximal gradient descent algorithms [40]

generalized to 2-D to accurately localize objects, without being misled by spurious artifacts.

We make a few important generalizations of our approach: (1) While the above optimization assumes the reflector is far from the object, IntuWition solves the equivalent problem for close reflectors through a maximum likelihood approach that iterates over coordinates of closeby reflectors. As we only care about nearby objects, the search space is relatively small, enabling efficient optimization. (2) To mitigate the effect of hardware offsets, we draw from past works [91, 115] which connects two Wi-Fi chips by a wired pathway (via RF attenuators) to enable synchronization. Specifically, we connect a cable from pin 3 of the transmitter, split to pin 3 of each of the three receivers, and divide each communication channel by the reference cable channel. We then use this ratio of the wireless channels between wireless and wired pathways, across both horizontal and vertical antennas, to eliminate frequency and timing offsets. (3) It is easy to see that our analysis readily generalizes to 3-D by iterating over the polar angle ϕ_p of any reflector p in the optimization as well, i.e setting: $w_{i,p} = e^{-2\pi j f_i \left(\tau_p + \frac{l_i \cos(\theta_p - \psi_i) \sin \phi_p}{c}\right)}$.

3.5.2 Detecting Occluded Objects

Among the most significant challenges IntuWition faces is detecting occluded objects, whose signals are received very weakly at receiver. More problematically, the weak signals of these objects are often overwhelmed by reflections from the object in front of them. This is the classic problem of the near-far effect faced by several wireless sensing solutions [6, 80]. Past solutions address this problem in several ways for moving objects in the environment, the most common of which is background subtraction [6, 5], where these systems subtract the observed wireless channels between two time instances to filter out what changed between them.

While IntuWition can benefit from these past solutions if the device is static, they do not apply when the device moves. This is because the wireless channels between measurements change both due to changes in the environment and the movement of the device itself.

IntuWition addresses this challenge by developing a background subtraction algorithm that accounts for the movement of the device itself. Specifically, let us assume that $h(r, \theta)$ and $h'(r, \theta)$ are the wireless channels along any given direction (r, θ) measured at two time instances where the device is displaced by a distance Δd and re-oriented by $\Delta\theta$. It follows that any object at (r, θ) at the first time instance is now at $(r - \Delta d \cos\theta, \theta - \Delta\theta)$ (provided $r \gg \Delta d$). As a result, if the reflector at (r, θ) remain static, it is easy to see that $h(r, \theta) = h'(r - \Delta d cos\theta, \theta - \Delta\theta)$. One can therefore perform background subtraction to detect moving object at each (r, θ) by subtracting the channels as follows:

$$\Delta h(r, \theta) = h(r, \theta) - h'(r - \Delta d \cos\theta, \theta - \Delta\theta)$$

IntuWition can then process $\Delta h(r, \theta)$ for each (r, θ) to detect the material of moving objects at each location, even if their signals are weakened due to occlusion.

3.5.3 Role Of Localization

Localization plays an enabling role to material sensing in our system. It is not our primary technical contribution, but localization is what allows us to isolate objects in the environment for materials classification and determines the resolution of the system. Localization also removes occlusions and determines the operational range of our system. After the object is isolated, it is passed to the material sensing algorithms described in §3.4 to isolate object characteristics.

3.6 Limitations of IntuWition

Given that our system is a feasibility study, we note some important limitations of IntuWition(see §4.9 for evaluation): It performs poorly when detecting weak reflectors whose signals are attenuated due to distance or size. Its resolution in separating multiple objects is limited by the total aggregate bandwidth of Wi-Fi: about 0.4 m in our experiments. This means that a human leaning on a wall may be misclassified. It often misses fleeting reflections due to fast moving objects. It cannot tell apart certain materials that have similar polarization characteristics or objects composed of the same material (e.g. two different humans). We are also limited in our ability to resolve a signal after significant multibounce effects, as each reflection significantly reduces the power at the receiver.

In addition, our system as a material classifier may respond most strongly to surface characteristics and misclassify clothing as humans, for example. In cases of extensive attenuation, such as around-the-corner coupled with through-wall sensing, our localization accuracy drops further. Finally, we note that because of the use of background subtraction, our system is most effective for systems where an object is introduced or where an object is moving. However, objects of sufficient size, reflectivity, and/or visibility can still be detected without background subtraction.

3.7 Implementation

System Setup: We implement IntuWition using four WiFi cards, one for transmission and three for receiving. Each includes a linearly polarized antenna (Ettus Vert2450) connected to each Intel 5300AGN card on the Intel Galileo Boards, which run the Linux 802.11n CSI tool [36] to obtain channel state information (Fig. 3.8). The transmitter also has an SMA splitter and wire on the first port of the 5300AGN card that

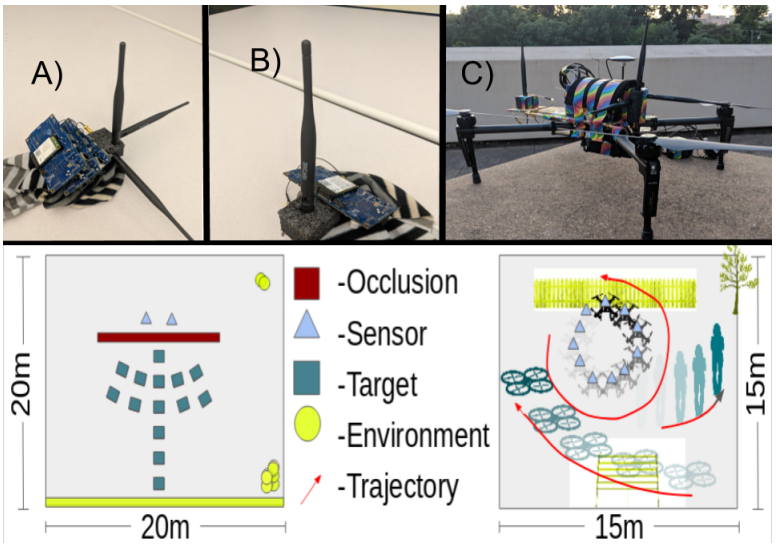

Figure 3.8: Setup: (A) 3 receiving and (B) 1 transmitting antennas, (C) Augmented UAV setup. Indoor (left) & outdoor testbeds (right)

carries a reference signal to the second port of the receivers to synchronize all radios and remove phase errors caused by hardware impediments (see Sec. 3.5.1). The wired line is attenuated by 50 dB to avoid saturation. To measure the polarimetry in 3D space, the three receiving antennas are oriented perpendicular to each other. Note that we chose to use three Wi-Fi cards, instead of one which would suffice in principle, because we observed that the first port of the Intel 5300AGN card had a consistently higher receive gain on average when compared to the other ports (a peculiarity of the chip). The overall weight of our sensing system is 439g, which can be significantly reduced (to just the antennas) should future Wi-Fi cards on mobile platforms (e.g. onboard Wi-Fi on drones) report wireless channels as the Intel 5300 does. We place our WiFi receiver and transmitter, pictured in 3.8A and B, on two setups for testing: a rolling cart, with a separation of 42cm, and atop a drone as shown in 3.8C, with a separation of 31cm.

We note that the use of WiFi chips with a future sensing feature could reduce the chips needed to two (if one chip could accept three receivers), would eliminate the need for the Galileo boards, and could speed our channel hopping process. Further, as we are only using channel measurements for sensing, our transmissions are arbitrary packets which could instead be data for the autonomous system.

Software and Run-Time: We implement IntuWition's algorithms in MATLAB/C++ (MLP implemented in Python) in the cloud and track location at the rate of 8 Hz and sense materials at the rate of 2 Hz. The main bottleneck in ensuring faster, real-time analysis is the amount of time it takes to sweep through all available WiFi channels in the ISM band supported by the Intel 5300 including 5-GHz bands – a problem that may be remedied by future chipsets. In its current state, a full frequency sweep for each receiver

radio results in, at most, a file 18.9kB. This requires a backhaul of 37.8kBps for material sensing. For localization, the radios hop across 1/4th of the channels, so the backhaul required is the same.

Experimental Evaluation: Unless stated otherwise, our evaluation across experiments are conducted in settings as described below. We evaluate the system in a 225 sq.m outdoor garage space, a 500 sq.m outdoor plaza, and four different indoor spaces ranging from 43-400 sq.m in a large university campus (see Fig. 3.8).

Our evaluation consists of diverse types of objects of various shapes, sizes and thicknesses. For our material tests, we use sheets of copper(3'x3'), aluminum(2'x3', 3'x3'), birch(2'x4'), plywood(4'x2'), maple(4'x4'x.25", 4'x4'x.5", 4'x4'x.75"). The wooden sheets also include a variety of textures: semi-gloss, roughed up with steel wool, and sanded by 80 and 220 grit. Furthermore, we additionally test more realistic objects in our evaluation, including a chair, a table, a filing cabinet, and a car. We also recruited ten human volunteers of different body types, with half our testing performed in summer clothing (thin t-shirts and shorts) and half of it performed in a winter clothing (jackets/coats). We collected data at different orientations (40 to 150 degrees of tilt) and distances (2m to 10m away from our setup), indoors and outdoors. We perform thirty trials for each material at each configuration. All our experiments are in indoor/outdoor multipath-rich settings (walls, furniture etc.), including non line-of-sight where objects were behind wooden partitions. We emphasize that for each experiment, we train and test our system on completely different subsets of both materials and testbed locations. All of our experiments consider the presence of multiple objects whose location and material we explicitly track (up to four). We explicitly evaluate the effect of mobility of the system in Sec. 3.8.5, which describes a proof-of-concept evaluation of our system to sense obstacles hidden from view around a UAV. Our UAV was measured to hover at 0.05 m/s by default and moves at speeds up to 2 m/s (fast walking human). Note that graph error bars show standard deviation of location/sensing accuracy across multiple experiments.

3.8 Evaluation

In this section, we describe the methods and results for our experiments. We note that localization is a precondition for material sensing, and both material sensing and localization are required for object tracking, so the results are inherently integrated but have been separated into material sensing and localization to observe trends.

3.8.1 Microbenchmarks

IntuWition relies on two hypotheses: that polarimetry can provide significant gain over spatial diversity

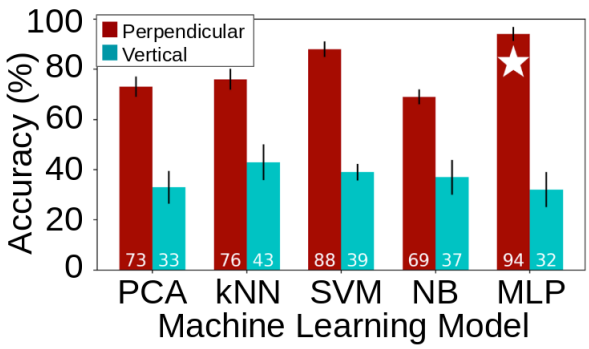

Figure 3.9: Three vertical antennas vs. three mutually perpendicular antennas, compared in material identification accuracy across 5 common classifiers. Note significantly higher performance for perpendicular antennas across the board, showing importance of polarization.

as a material sensing modality, and that Multi-Layer Perceptrons are the best model for our application due to their analysis of higher-dimensional features (compared to say, kNN) and their deep architecture (allowing for more efficient and intelligent behavior than shallow ones [14].)

To test these hypotheses, we ran a preliminary experiment. To confirm polarimetry, rather than spatial diversity, was contributing to material sensing gain, we ran identical trials with three vertical receiving antennas as well as three perpendicularly oriented antennas. We collected data (specifically, power ratios of receiving antennas) in a line-of-sight lab setting to distinguish between three material types: birch, aluminum, and human, at a variety of angles and distances, to create our training set. Then, we rearranged the setup and background furniture in the lab to create the test set. Using this training and test data, we found material classification accuracy for five different common machine learning models, as shown in 3.9.

From this experiment, we draw two major takeaways: The use of perpendicularly oriented antennas significantly increases material sensing accuracy across all machine learning models, and that Multi-Layer Perceptrons are indeed the ideal model for our experiment. We chose to build the Multi-Layer Perceptron to have three hidden layers based on our experimental analysis.

3.8.2 Object Localization

While localization is not the main focus of our chapter, we include localization accuracy to isolate different materials and distances in Fig. 3.10 for completeness. For ground truth, we affix fiducial markers to objects and implement a camera-based fiducial tracking system using ARToolKit [1], which reports a sub-millimeter baseline accuracy. The primary takeaway from our localization results is that our system can typically isolate a line-of-sight (LOS) or non-line-of-sight (NLOS) object that is at least a meter separated

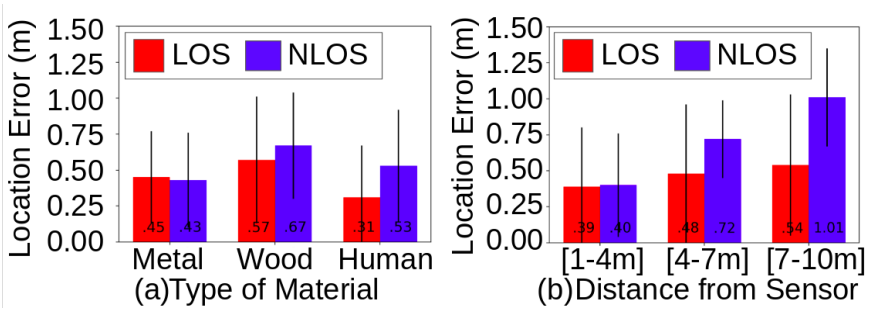

Figure 3.10: (a) Effect of Material Type on Localization Error: We observe our system localizes wood and metal best, which we expect is due to the larger surface areas of our material samples on average when compared to humans, for an overall mean error of 0.49m. (b) Effect of system to reflector one-way distance on Localization Error: we see error increase with distance, which we expect is due to less power being received from reflectors as distance increases.

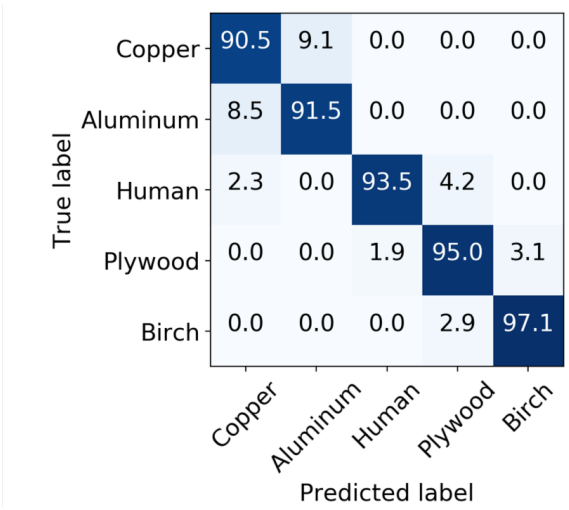

Figure 3.11: This confusion matrix shows our efficacy in classifying between five different materials using our Multi-layer Perceptron Classifier.

from other reflectors.

3.8.3 Material Sensing

In this section, we discuss our core material sensing results and polarimetry's robustness to material thickness, surface area, range, and texture.

(1) Overall Observations In fig. 3.12(a), we can see IntuWition has overall classification accuracies of 95% in LOS and 92% in NLOS. The confusion matrix in fig. 3.11 further shows that our system tends most often to confuse objects of the same class– metals copper and aluminum, for example. We surmise our system can only distinguish them at all due to the thin layer of oxide that forms on the surface of a metal – aluminum oxide is an insulator, while copper oxide is a semiconductor, or perhaps due to polymer coatings used to prevent rusting. We don't anticipate this confusion to be of concern in some contexts

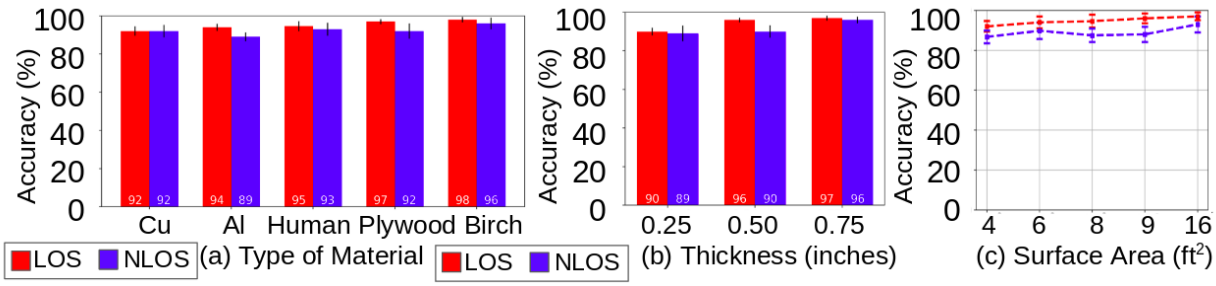

Figure 3.12: (a) The effect of material type on classification accuracy: we see an average accuracy of 93.5%, with higher errors for copper and aluminum (since they are more easily confused). (b) The effect of thickness of wood on classification accuracy: we see lower accuracy as thickness decreases and fewer layers of material exists to reflect its signature. (c) Material Sensing vs. Surface Area: we see accuracy increases with surface area as scattering occurs at a larger scale.

such as UAV path planning, since the exact type of metal isn't critical to the functionality of the system.

When comparing system performance with multiple tracked objects, we see a negligible performance decrease for the two-object case, a 5% decrease in classification accuracy in the three-object case, and a 6% decrease in classification accuracy in the four-object case.

(2) Material Sensing vs. Thicknesses We use wood to evaluate our system's robustness to thickness, as metal only has a skin depth of roughly 1μm at Wi-Fi frequencies (mostly acting as a surface reflector) and the material composition of humans is heterogeneous.

In Fig. 3.12(b), we observe that while 0.25" wood has less layers of water to refract compared to the other two, our system still maintains a robust classification accuracy higher than 89% even in the NLOS scenario. Further, we observe that a thickness increase of half an inch resulted in a 7% increase in classification accuracy in line-of-sight settings. We surmise that this is because there is more available material to reflect and refract its unique polarization signature. This leads us to believe that for non-metallic objects, the material that composes the largest part of a reflector may impart the most significant share of signature, as would be necessary for identifying humans regardless of clothing.

(3) Material Sensing vs. Surface Area We use a variety of material surface areas to evaluate our system's resilience to different sizes. Specifically, we used various sizes of aluminum, copper, maple, and birch, as well as a human. These ranged in surface area from 4ft^2 to 16 ft^2 at distances ranging from 0-15 m.

Fig. 3.12(c), shows that as surface area increases from 4ft^2 to 16ft^2, the classification accuracy increases from 88% to 93% in NLOS and from 92% to 98% in LOS scenarios. That is, reflective surfaces with larger cross-sectional areas are easier to detect and classify, as expected. Note that lower surface areas will progressively lower accuracy (a limitation of IntuWition).

(4) Material Sensing vs. Range To evaluate our system's robustness to range, we collected measure-

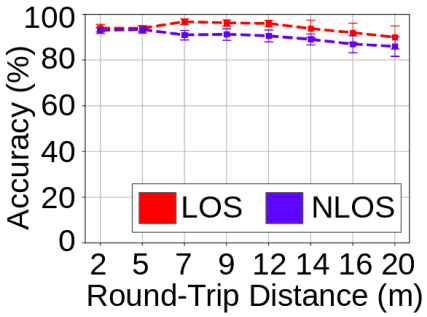

Figure 3.13: Effect of Round-trip Distance on Material Sensing: we see accuracy slightly increases with distance at the beginning, as the reflector becomes more distinguishable from the strong line-of-sight signal, and eventually falls with distance.

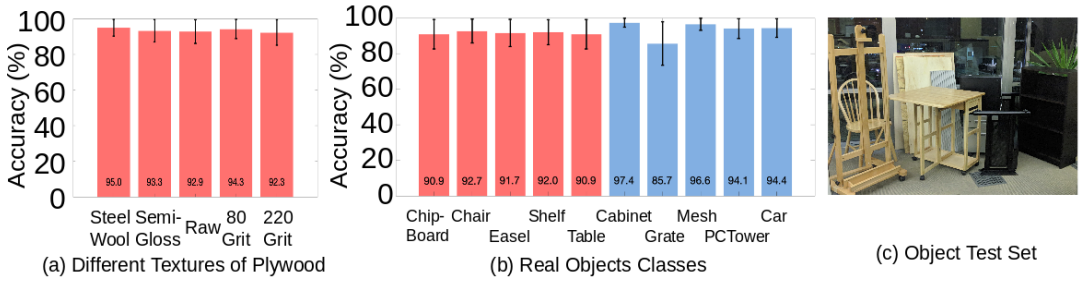

Figure 3.14: (a) We show our classification network accuracy dealing with different surface textures of the same material (wood). (b) We show our classification network accuracy dealing with real objects used in our daily life of different material (wood is showed as red, metal is showed as blue). (c) We show the boards, furniture, and objects used for these experiments.

ments at up to 20 m in round-trip distance from the system to the sensed object.

As shown in Fig. 3.13, we observe a slight increase in our system's accuracy in the beginning, which we surmise is due to the increased ability of our system to resolve the reflector given the strong line-of-sight signal. Then, there is minimal impact in our system's performance with up to a round-trip distance of 12 m, following which we see a gradual decrease as expected due to the reflector's signal being received more weakly.

(5) Material Sensing vs. Surface Texture Finally, to evaluate our system's robustness to texture, we collect measurements from five surface textures of plywood of fixed size and thickness: unfinished, roughed up with steel wool, painted with semi-gloss, sanded to 80 grit (particle size of 201 μm), and sanded to 220 grit (particle size of 70 μm).

As shown in Fig. 3.14a, the surface texture of wood does not play a statistically significant role in classification accuracy.

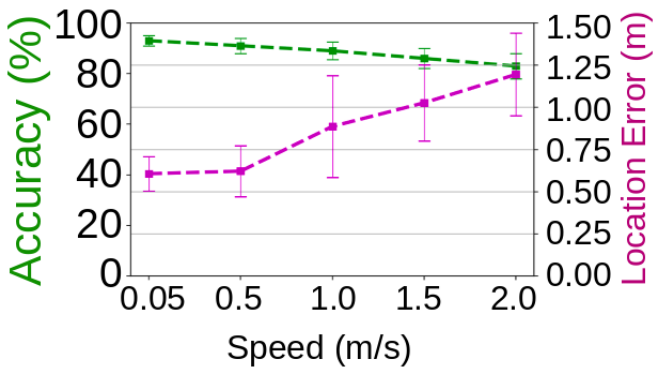

Figure 3.15: Measures the accuracy of (a) Object recognition (drone vs. human); and (b) Localization at varying UAV speeds.

3.8.4 Object Recognition of Real-Life Obstacles

To test our system's coarse applicability to real-life objects, we train a new model to classify wood from metal using all collected test and training data up until this point to create the best model possible. We test this on new, unseen objects. These include five of each wooden and metal objects of varied material types (various wood and metal types): A rough chipboard sheet, a table, a chair, an easel, and a bookshelf; a filing cabinet, a shelving grate, a mesh furniture stand, a matte PC tower, and the passenger-side-door of a car. These were collected in a lab setting using a 4'x4'x1" wooden sheet to occlude the setup for NLOS settings, with the exception of the car which was collected in a yet-unseen parking garage environment and was occluded by a 2' concrete wall for NLOS readings.

As seen in Fig. 3.14 we see relatively high accuracy across the board given that our system had only been trained on metal and wooden sheets and these new objects were unseen. We note that classification accuracy for the metals appears to be slightly higher, with the exception of the metal grate, whose structure likely introduces a lot of destructive interference.

3.8.5 Application: Object Recognition from a Mobile UAV

In this experiment, we demonstrate our system's utility in avoiding two types of obstacles: humans and other UAVs for a proof-of-concept application: a delivery drone navigating indoor and outdoor spaces. We are particularly interested in capturing how the drone's speed and vibrations affect our system (see Sec. 3.4.2). Fortunately, recall that IntuWition only needs the relative trajectory of the UAV, not its absolute location at any point. Further, IntuWition applies a sliding window filter (N=6) on the output of inertial and motion sensors on UAVs across multiple measurements to mitigate spurious readings. Since our drone is operating around 100 revolution per second which can be translated to 100 Hz of the vibration of

the drone [59], we additionally apply a high-pass filter to remove this noise. We classify between drones and humans from atop a drone at a few different speeds: 0, 0.5, 1.0, 1.5 and 2.0 meter per second along a circular trajectory with a 5 m radius. Note that we do not test at higher speeds owing to limitations of the UAV platform and to ensure user safety. We further localize the reflector using 3-D MUSIC to measure localization error in three dimensions.

We notice that increasing UAV speed decreases the accuracy of material sensing. One hypothesis is that because the rate of sweeping Wi-Fi frequencies is around 2 Hz, increased vibration of the propellers impacts our error in localization which in turn influences material sensing accuracy. Our MLP model has a prediction latency around 0.04 second (25 Hz). We observe that our accuracy falls off with speed, but we are still able to achieve 89% mean accuracy in classifying between humans and drones at speeds of up to 2 m/s and mean localization accuracy of 0.87 m.

3.9 Summary

This chapter presents IntuWition, a system that explores sensing the material and location of hidden objects in the environment using commodity Wi-Fi radios. IntuWition analyzes the change in polarization of the wireless signals as they reflect off different objects surrounding a Wi-Fi device to infer their material composition. A detailed evaluation demonstrates promising accuracy in both localization and material identification.

While IntuWition categorizes between five types of materials, we believe future work can take this much further. We believe high-bandwidth radios such as 802.11ad can greatly improve material sensing resolution. We also leave for future work developing optimized hardware that performs all material processing on-board the device, as opposed to the cloud. In the context of UAVs, fusing information with on-board cameras and dealing with object recognition system design challenges on robotic platforms remains an important task for future work.

Chapter 4

Metamoran: A Hybrid mmWave and Camera System for Long-Range Depth Imaging

After considering opportunities in improving WiFi as a sensor towards the vision in Fig. 1.4, we turn our attention to opportunities in the other two sensing modalities. Because RADAR and camera are so thoroughly developed as sensors individually, we instead focus on exploring opportunities at their intersection for high-resolution (both angular and depth) imaging of a scene, our third system output target. We note that localization and object type identification are largely implicit to these sensors: mmWave RADAR is well developed to detecting the location of objects, and camera vision techniques can provide object labels for obstacles in a scene. We exploit both of these capabilities to image obstacles in the environment with

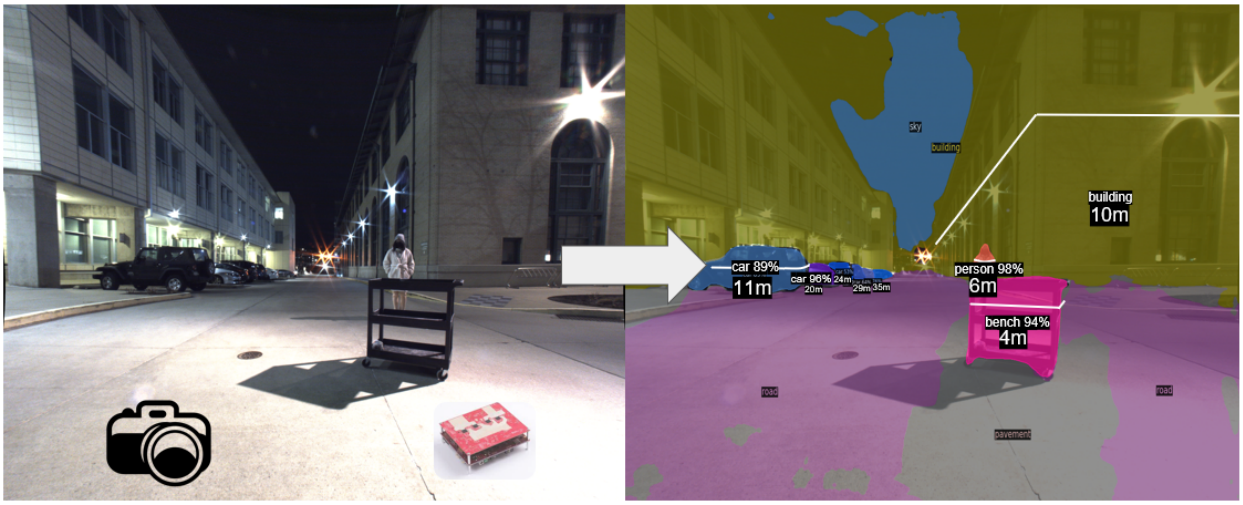

Figure 4.1: IntuWition is a first attempt to achieve our vision of localization and object type identification using WiFi alone.

a depth metric. The parts of our overall vision contributed by Metamoran are shown in Fig. 4.1. It takes as input the camera capture and mmWave I/Q values of a scene, and produces an output of the location of objects in the environment, a semantic label of object type, and a depth profile for the given object.

4.1 Motivation and Approach

One of the most appealing features of mmWave radar systems arises from its high bandwidth and carrier frequency, which enables precise depth estimation at long depth ranges, often as large of 60 meters, and at cm-scale resolutions. This finds application in a wide range of areas, including security [18], automobile safety [109], industrial sensing and control [24]. For comparison, most RGB camera solutions of the same physical form-factor (e.g. monocular depth estimation [15], depth cameras [98], stereo-vision [94], etc.) struggle to reach such resolutions for objects at extended distances and are about an order-of-magnitude worse. Yet, mmWave radars, by themselves, are not a capable 3-D imaging solution as their angular resolution along both azimuth and elevation is extremely poor — with the best radars of the market at least $10\times$ poorer than camera systems. This has led to mmWave radars being restricted to niche applications – for instance, in airport security [18] or physical collision sensing [109] — where their impressive depth range and resolutions are not fully utilized. This naturally leads us to the question: *Can we fuse cameras and mmWave radar sensor data to provide the best of both worlds and build a rich 3-D depth imaging solution?*. In doing so, we seek a 3-D imaging system that can be readily deployed from a single fixed vantage point to enable applications as long-range road-side safety systems, surveillance and security applications, wide-area mapping and occupancy sensing.

This chapter presents Metamoran[1], a hybrid mmWave and camera-based sensing system that achieves high angular and depth resolution for objects at significant distances – up to 60 meters (see Fig. 4.2). It achieves this through a *novel specular radar processing algorithm* that takes information from computer vision algorithms such as deep neural network-based image segmentation as input. While efforts have been made to fuse radar and camera data in the past, primarily for short range object detection and tracking [20], imaging under physical [73] or weather-related occlusions [62], this chapter considers the unique problem of hybrid mmWave/camera sensing for long-range outdoor depth imaging.

A key contribution in our system is improving depth sensing capabilities beyond what is typically achievable by a mmWave radar alone using a novel radar processing algorithm that provides high depth resolution (along the z-axis) *guided* by computer vision techniques that have high spatial resolution (in x and y). First, we detect and identify an object using a camera-based image segmentation algorithm, which

[1] A fictional race from the Dragon Ball Universe that taught Son Goku the Fusion technique [103].

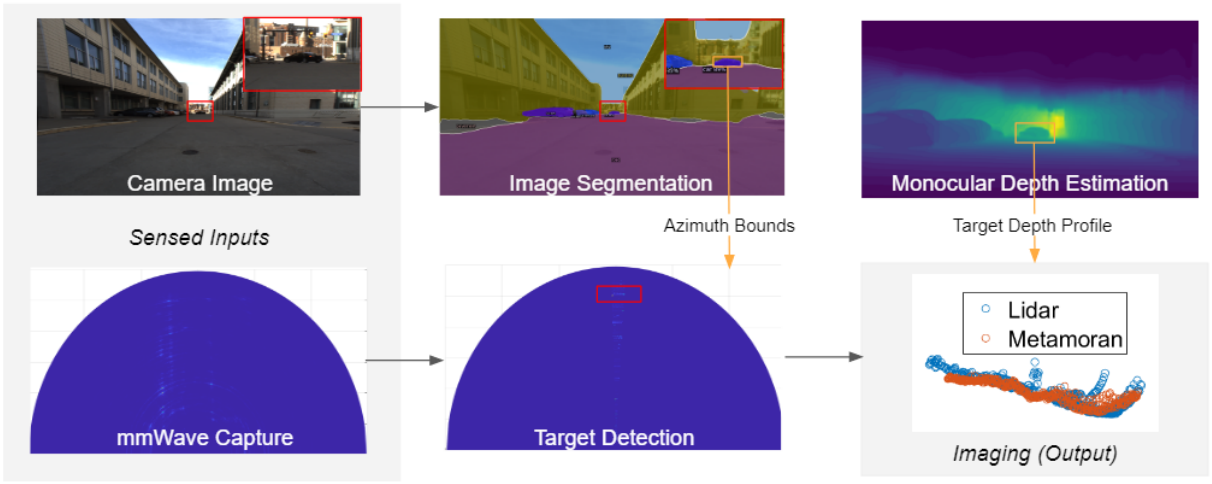

Figure 4.2: Metamoran devises a novel mmWave specular beamforming algorithm that forms high resolution depth-images 60 m away from objects-of-interest, using inputs from vision techniques such as image segmentation.

gives us the angular position (in the x-y plane) of objects in the environment as well as their spatial outline. Our key technical contribution is a novel specular radar beamforming algorithm (see Sec. 4.6) that returns high-resolution depth estimates by processing radar signals along the angular span and shape outline for each object in the image identified using segmentation. We then show how such a system could be combined with dense monocular depth estimates to create robust depth images of individual objects, capturing depth variation within the object itself, even at extended distances away from the radar-camera platform. In other words, we show how semantic inferences on vision data collected by the camera can help declutter and provide useful priors to obtain high-resolution depth images that are better than standalone radar or camera algorithms.

Our second contribution is to address various challenges in making Metamoran robust in case of cluttered environments, unfavorable object orientation, extended distances and partial occlusions that impede the radar, camera, or both. We address this problem specifically for stationary objects: this form radar's worst-case scenario (Doppler can help to detect moving objects) as objects that are not moving (e.g. traffic signs, parked cars, children at a bus stop) can also be important to detect. We narrow down objects whose spatial bounds are consistent across both camera and radar images, thereby allowing for increased robustness by reducing clutter. We also observe strong reflections from out-of-spatial bound reflectors leak into our spatial bound of interest and design cancellation techniques to detect weak reflections which would otherwise be masked by spurious objects. Further, we design and show how Metamoran continues to operate well, even amid partial occlusions, e.g. due to fog or partial occlusion from other objects. We document instances where radar systems can actively be used to improve camera image segmentation by

identifying objects that were initially missed by segmentation.

We implement Metamoran with a TI MMWCAS-RF-EVM radar and a FLIR Blackfly S 24.5MP color camera. Due to the relative lack of rich public mmWave radar I/Q datasets over long distances, we collected extensive data (200 scenes totalling 100 GB of I/Q samples and camera data) in diverse scenes outdoors at a major U.S. city. A few highlights from our results include:

- An evaluation of the effective median depth of an object-of-interest at distances of up to 60 meters, in diverse outdoor settings, at a median error of 28 cm. This is an improvement of about 23× versus state-of-the-art monocular depth estimation and 13× versus a naive camera + radar beamforming solution.
- Dense estimation of the azimuthal/depth profile of a single object-of-interest, for an imaging error of 80 cm at distances up to 60 meters. This is an improvement of about 4× versus state-of-the-art monocular depth estimation and 6× versus a naive camera + radar beamforming solution.
- A demonstration of resilience to various classes of partial occlusions and blockages.

Contributions: We makes the following contributions.

- Metamoran, a novel system that combines camera and mmWave sensing to achieve high resolution depth images at long ranges.
- A specular beamforming algorithm that leverages the output of image segmentation algorithms from computer vision to declutter and retrieve depths of objects-of-interest from radar I/Q samples.
- A detailed implementation and evaluation of Metamoran in varied environments to demonstrate substantial improvements in long range depth imaging.

Limitations: We concede that our system is limited by more significant occlusions that impact camera observations and discuss the limitations of our system in Sec. 4.10 as well as present an evaluation of both successful and failure modes with various types of occlusions in our results in Sec. 4.9.

4.2 mmWave and Camera Fusion

Radars, once only limited to military applications, are today used ubiquitously in a variety of applications from airport security [18], automotive applications [102], human-computer interfaces [60] and industrial automation [64]. A key factor which enabled this trend was the usage of mmWave frequencies which allowed for compact antenna arrays and wide bandwidths, both of which are crucial for radars' target ranging and imaging capabilities. mmWave radars, as the name suggests, use radio waves of millimeter

scale wavelengths in either 60 GHz or 77-81 GHz by first actively illuminating an environment and then processing the reflections from various objects in the environment. This is noticeably different from modern image sensors which purely rely on passively sensing rays which make their way to the sensor. The reflections from the objects encode useful information such as objects' range, azimuth, elevation and velocity with respect to radar. The transmitted illumination and radar hardware are the main factors which limit the radars' ability to generate high resolution 3D images of the scene.

Advantages of mmWave Radar: Most commodity radars transmit a Frequency Modulated Continuous Wave (FMCW) signal which is a waveform that continuously changes its frequency over time to span a significant bandwidth B. A radar's range resolution is fundamentally limited by this effective bandwidth of the transmitted signal as $\frac{c}{2B}$ (c is speed of light). In the 77 GHz band, we have a theoretical range resolution of 3.75 cm over tens of meters. In this regard, radars are on par with time of flight LIDARs which report a similar range accuracies. However, unlike LIDARs, radars work in all weather conditions (rain, snow, fog) and extreme ambient lighting (sunlight) [67].

Limitations of mmWave Radar: However, radars unfortunately have worse azimuth and elevation resolutions compared to both cameras and LIDARs. While range resolution is limited by the bandwidth of the radar signal, angular resolutions are dictated by the number of antenna elements that are packed on a radar. As the number of antenna elements increases, so too does the resolution. The best state-of-the-art commercial mmWave radar available [100] with as many as 86x4 antenna elements has a 1.4°x18° angular resolution. In contrast, state of the art LIDARs today achieve 0.1°x2°, atleast 10x better angular resolution than radars [61]. With a poor angular resolution, 3D radar images look very coarse and blobby in the angular domain. While more antenna elements can be added, they come at significant increases in device cost and form-factor – bridging the 10× gap is simply not an option with today's state-of-the-art hardware. We make the observation that even commodity cameras, because of their dense focal planar array image sensors, are better than radars in terms of angular resolution at about 0.02°x0.02° [68]. This observation leads us to study combining the high angular resolution of camera systems with the high depth resolution of mmWave radar – an approach we describe in the next section.

4.3 Overview

4.4 Metamoran's Approach

Metamoran at a high level, takes as input camera and 77 GHz mmWave radar data from a scene. We use these inputs to fuse and return a high-resolution depth image for specific objects-of-interest at distances

of several tens of meters away. We specifically consider cars and persons – key to surveillance, industrial and occupancy sensing applications. Our key contribution is a novel radar processing algorithm that produces refined depth estimates for specific objects-of-interest, based on priors obtained through image segmentation of camera images. We choose a radar-based processing approach rather than an exclusive deep-learning based approach on all underlying data (images + raw I/Q), due to better explainability of the inferences. Besides, the resolution obtained from our system in depth is close to the physical limits that can be obtained owing to the bandwidth of the radar. Nevertheless, our solution benefits heavily from state-of-the-art deep neural network based image segmentation algorithms that operate on image data.

System Architecture and Outline: Fig. 4.2 depicts the architecture of our system that we elaborate upon in the following sections. First, we apply two state-of-the-art pre-processing steps that operate on image data (Sec. 4.5): (1) image segmentation, i.e. identify the spatial (x and y) bounds of objects-of-interest – cars, people and traffic signs; (2) Monocular depth estimation to obtain an approximate estimate for the shape of these objects, albeit prone to error at large distances. We then design a novel specular beamforming algorithm in Sec. 4.6 that uses priors along one dimension (x and y) from image segmentation and monocular depth estimation which provide a coarse shape of the object of interest to then obtain a fine-grained depth image. (3) Our final step (Sec. 4.7) is to build resilience to occlusions and clutter into our system, to improve performance in a variety of circumstances.

4.5 Image Pre-Processing

Metamoran's first step is to process camera image data to learn about the approximate span in azimuth and elevation of objects-of-interest, as well as an approximate silhouette or outline along the x-y plane, i.e. parallel to the depth axis. We specifically consider three specific classes of objects-of-interest that are ubiquitous in outdoor sensing – cars, pedestrians and roadside infrastructure (traffic signs). As mentioned in Sec. 3.2, we exploit the high angular resolution of camera systems that are at about 0.02°x0.02° [68] – orders-of-magnitude better than mmWave radar systems. 's vision pre-processing steps below are therefore crucial in providing prior information on the shape and location of objects-of-interest along the x-y plane so that mmWave data can be used to focus on these objects and improve resolution along the z-axis.

4.5.1 Image Segmentation

To find the spatial bounds (along x-y) of objects of interest, we perform state-of-the-art image segmentation which labels objects by their type and creates masks that capture the outline of these objects (see Fig. 4.3

Figure 4.3: **Image Segmentation:** Metamoran uses image segmentation to identify the spatial bounds along the x-y axes of objects-of-interest – cars, pedestrians, traffic signs – with semantic labels assigned.

for an example).

We perform image segmentation using Detectron2 [111] trained with KITTI dataset. This model has been previously trained on several objects including cars, pedestrians and traffic signs in outdoor environments. We use these types of objects as our primary test subjects without additional model tuning. This image segmentation combines the best of both worlds from semantic segmentation and instance segmentation, by providing a segmentation mask (outline), a semantic label for the mask and instance ID for each detected object as shown in Fig. 4.3. The segmentation mask directly provides the spatial bounds and precise shape of the object along the x-y plane and is fed as a prior for mmWave specular beamforming in Sec. 4.6 below.

4.5.2 Monocular Depth Estimation

As a second step, we perform state-of-the-art monocular depth estimation specifically on objects-of-interest filtered through image segmentation above. We use this scheme both as a baseline for comparison and to provide a coarse range of depths (depth profile) that the object spans. We use AdaBins [15] for monocular depth estimation of the objects-of-interest as detected by the image segmentation step. We note that state-of-the-art monocular depth estimation is poor in terms of accuracy and resolution at extended distances, with errors of about 19.5 meters for objects that are 60 meters away (see Fig. 4.13). Nevertheless, we see that monocular depth estimation provides useful prior information on the approximate range of depths that the object spans and combined with image segmentation provides a rough 3-D shape (outline) of the object that serve as inputs for our mmWave specular beamforming algorithm in Sec. 4.6 below.

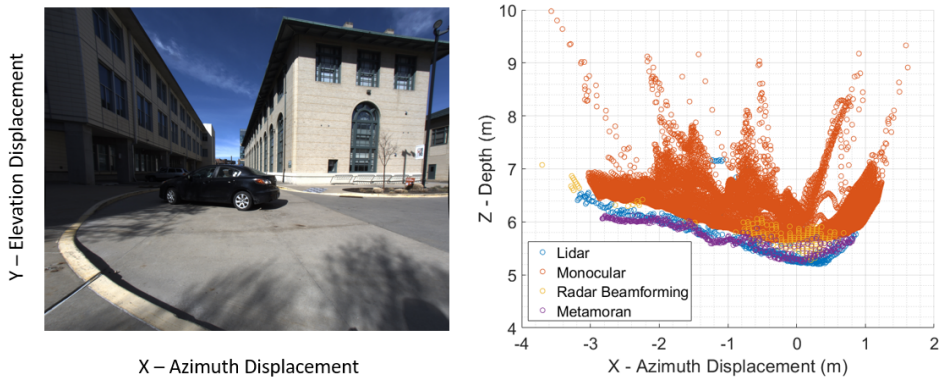

Figure 4.4: **Metamoran vs. Radar Beamforming and Monocular Estimation:** A qualitative comparison of the depth images shows standard radar beamforming to be very coarse in azimuth resolution, monocular to have significant absolute depth offsets but great azimuth diversity, and Metamoran which leverages rich shape information from image pre-processing to generate an accurate, dense depth image.

4.6 mmWave Specular Beamforming

Metamoran's specular beamforming algorithm processes the complex I/Q samples received from the mmWave radar platform, coupled with the shape outlines of objects-of-interest in the scene, obtained from the image pre-processing steps in Sec. 4.5 above. In traditional mmWave beamforming [97], received I/Q samples are effectively projected along all spatial angles (azimuth and elevation) to obtain the signal time-of-arrival between the object to the radar. This quantity, when multiplied by the speed of light, obtains the depth of the object. Unfortunately, this approach relies on the azimuth resolution of the radar, which is fundamentally limited by the number of antennas on the radar itself – at best 1.4° in state-of-the-art radar systems. The end result is a coarse radar image.

4.6.1 Depth Super-Resolution

Metamoran's key technical contribution is a novel specular beamforming solution, a super-resolution algorithm that overcomes the poor azimuth resolution of mmWave radars by using priors from the image pre-processing steps in Sec. 4.5. At a high level, Metamoran attempts to build a mmWave wireless signal called the *object template* that captures the influence of an object of a particular shape (as determined by camera pre-processing) on mmWave radar receptions. Further, Metamoran also knows the precise azimuth and elevation angle that this object template appears at, owing to the high angular resolution of camera systems. Metamoran then identifies the best-possible depth one could apply to this object template to best fit the observed radar signals. The end result is a finer resolution depth image of the object-of-interest as shown in Fig. 4.4(b).

Detailed Algorithm: Mathematically, Metamoran's algorithm extracts the approximate shape contour inferred from image pre-processing, coupled with a mmWave ray-tracing model to estimate the expected I/Q samples of reflections from such an object – i.e. the object template. Essentially, the object-template is obtained by modeling each point on the surface of the shape of the object $S(x, y, z)$ as a point reflector shifted to some depth value d that results in an overall distance of d relative to the radar. In its simplest form, one can then obtain this point's contribution to the received signal as at each wavelength λ as [104]:

$$h_{template}(d) = \frac{1}{d} e^{-j4\pi d/\lambda}$$

Where the 4π rather than the traditional 2π stems from the fact that radar signals are reflected or scattered back round-trip. We can then denote $h_{template}(d)$ as the total channel experienced across the entire bandwidth over all the points in the template. Metamoran then applies a matched-filter to obtain $P(d)$ – the correlation of the object template at each possible depth d relative to the radar by processing the received signals across frequencies. Mathematically, if h is the received channel, we have:

$$P(d) = h^*_{template}(d)h$$

We then report the depth estimate of this object as the value of d that corresponds to the maximum of $P(d)$, i.e.

$$d^* = \arg\max_d P(d)$$

Algorithm 1 provides a more elaborate description of the steps of Metamoran for FMCW mmWave radar signals.

Metamoran's design of object templates overcomes the azimuth and elevation resolution limits of mmWave radar. To see why, note that one could intuitively view our design of templates as effectively performing a form of sparse recovery – i.e., Metamoran assumes that objects of a particular shape are unique at a certain range of azimuth and elevation in the radar reception. This sparsity assumption is key to Metamoran's super-resolution properties.

4.6.2 Intra-Object Depth Profiling

We note our current description of Metamoran's algorithm provides only one depth value per object template, i.e. one depth per object. In practice, we deal with extended objects and we would require multiple depth values across the object. We could use local peaks from the specular beamforming output near the peak depth value. But, the point cloud so obtained is very sparse and only becomes sparser with increasing object distances. In an ideal world, we would like an output similar to monocular depth estimation (see Fig. 4.5 for an example). In monocular depth estimation, pixel color and other image

Algorithm 1: Specular Beamforming Algorithm

Input : Image Segmentation Object Mask, P
Monocular Depth Estimation, M
Raw I/Q Radar capture, h

$S = M \cdot P$ // Approximate 3D shape of object
$C(x,z) = \text{GetShapeContour}(S(x,y,z))$
for *depth* d **do**
  $h_{template}(d) = \text{ShiftByDepth}(C(x,z),d)$
  $P(d) = h^*_{template}(d)h$ // Matched Filtering
$d^* = dargmaxP(d)$ // Depth Estimate
/* Choose local peaks near d^* to generate Metamoran's sparse point cloud */
$MM_{sparse} = \text{GenerateSparseImage}(d^*, P(d))$
/* Nullify large absolute errors from monocular estimation */
$C = \text{ShiftToDepth}(C, d^*)$
/* Reject outliers which occur along the edges of the image */
$C^* = \text{RejectOutliers}(C)$
$MM_{dense} = \text{Fuse}(MM_{sparse}, C^*)$
Output: $MM_{dense}(x,z)$ // Dense Depth Image

Figure 4.5: Monocular depth estimation gives a dense RGB-D depth image which is promising for fusing with sparse Metamoran's specular beamforming point clouds.

features are used to identify objects at various depth levels resulting in a dense RGB-D image as shown in Fig. 4.5. Our key idea is to make use of the dense monocular depth estimation in conjunction with the sparse point cloud from specular beamforming described so far. However two problems persist in realizing this fusion: (1) First, while monocular depth estimation may often correctly return the *relative* depths between different parts of a large object such as a car, it often makes large errors in *absolute* depths, particularly for objects at extended distances [87, 83]. (2) Second, monocular depth estimation often struggles with objects that do not have significant variation in color with respect to the background or sharp edges that intuitively simplifies depth estimation [87, 83]. The rest of this section describes how we address both these challenges to fuse Metamoran's depth images with off the shelf monocular depth estimates (see Fig. 4.5) that offer superior accuracy to monocular depth estimation.

Correcting Absolute Errors: To address the first challenge, we can simply shift the monocular depth estimates for any given object-of-interest so that they line up with the sparse point cloud obtained from Metamoran's specular beamforming algorithm. This ensures that absolute errors for any given object-of-interest are minimized. A key point to note is that for large objects (e.g. a car), there may be some ambiguity on which exact point on the monocular depth estimate should be shifted to line up with Metamoran's estimate. To remedy this, we correlate the object template used in Sec. 4.6.1 from image segmentation with the image that resulted from monocular depth estimation. Recall that this very object template was used to estimate the object's depth in Metamoran's super-resolution algorithm. The correlation process therefore allows us to identify the pixel on the image that best corresponds with the depth estimates from Metamoran's super-resolution algorithm.

Correcting Relative Errors: After aligning the monocular depth estimates with the sparse point cloud from Metamoran's beamforming, a naive way to fuse this would be consider all points from both modalities. But, as seen in Fig. 4.4(b), edges of monocular estimates tend to deviate a lot from the primary contour outline of the object. If fused as is, one would experience errors expected from monocular depth estimation. It's therefore important to select points from the aligned monocular depth estimates that only lie along the primary contour outline and reject outliers. We note that the number of points detected per azimuth bin in monocular estimates fall off sharply at the edges where our outliers of interest lie. By using a simple threshold based outlier detection, we identify points which actually lie along the primary contour. Upon fusing selected monocular depth estimate points and sparse point cloud from Sec. 4.6.1, we obtain a depth image that outperforms different algorithms using either of the two modalities in terms of depth and azimuth resolution and depth accuracy.

4.7 Resilience

The effective imaging of a reflector relies first on effective detection of the desired object. Improving the ability of a mmWave radar to detect and find the depth of a given reflector in cluttered conditions thus becomes a critical enabling piece. This falls into three broad categories: reducing false positive rate from spurious peaks and unwanted reflectors, increasing the ability of our system to detect weak reflectors, and providing resilience to occlusions. We discuss how the introduction of a camera allows Metamoran to improve in all of these categories when compared to radar alone.

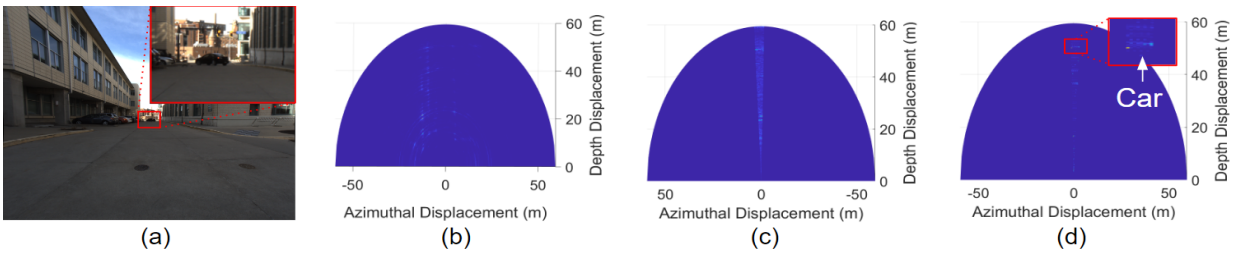

Figure 4.6: **Metamoran vs. Clutter:** Metamoran can help identify objects-of-interest despite environmental clutter. (a) shows our scene, a narrow parking lot bound by buildings with a lot of cars, as well as our target, a car that is 50m away. (b) shows the raw radar beamforming of the area, with very prominent out-of-span peaks from nearby cars and buildings. (c) shows the slice of the radar beamforming bound by azimuth span determined from image segmentation of the image. (d) shows the same azimuthal slice with side lobes of out-of-span reflectors removed, with only one peak remaining that corresponds to the reflected power profile of a car.

4.7.1 Reducing Clutter

To improve the robustness of Metamoran's algorithm, we present a key optimization that was pivotal in identifying the true depth of objects-of-interest. In particular, our focus is in cluttered environments where reflections from a large number of objects impede identifying the depth of the true object. At first blush, one might assume that even with a large number of objects in the environment, the number of objects at the desired azimuth angle – as specified by image segmentation, would be relatively few. Further, given that the object is in direct line-of-sight of the camera, it can also be expected to correspond to the first peak observed along this 3-D angle.

However, we observe in practice that peaks from extremely strong reflectors leak significantly in azimuth as well, often into our desired angle. This is due to the poor angular resolution of the radar. This is a problem due to two factors: (1) these leaks can appear as a false peak closer to our detector, corrupting a first peak approach, and (2) these strong reflectors are often three orders of magnitude larger than our desired reflector, and thus have leaks that can dwarf our targets-of-interest. One must therefore perform a declutter phase prior to applying Metamoran's specular beamforming algorithm that discounts and eliminates spurious results at depths that correspond to these spurious peaks. Doing so would prevent Metamoran's algorithm from being misled by such peaks. Fig. 4.6 provides a qualitative comparison of the impact of Metamoran's algorithm in decluttering the radar image and identifying the true peak. The plots (b)-(c) is this figure represent $P(d, \theta)$, which we call *radar profiles*, that represent the power of signals received at different depths d and azimuth values θ, measured through the standard Bartlett-based radar beamforming algorithm [77]. Our objective is to remove unwanted clutter in these profiles to focus on the object's of interest by masking out unwanted regions. This allows us then apply Metamoran's mmWave

super-resolution algorithm from Sec. 4.6 by ignoring unwanted clutter.

Specifically, in Metamoran we look for peaks in the regions of our radar profile that fall outside of the azimuth span of our target, as expected from image segmentation. For each peak, we generate an object template that is the scale and position of that peak – including its side lobes – and subtract it from our profile. We iterate many times until the magnitude of the peaks in the area outside of our focus are comparable to the expected magnitude of the target reflector. This is analogous to successive interference cancellation in wireless communications [74], or the CLEAN algorithm in radio-astronomy [23], with the distinction that we only remove peaks outside of our desired sensing azimuthal span. What this process accomplishes is the removal of side lobes from very large peaks in our azimuth of interest – which is critical for the performance of our system.

4.7.2 Addressing Weak Reflections

In this section, we explore ways to amplify extremely weak reflections from objects-of-interest, either due to their material properties, poor orientation or extended range from the radar. Indeed, the precise level to which radar reflections weaken depends on a combination of all of these properties and we evaluate this further for a diverse set of objects in Sec. 4.9.1. While radar typically uses Doppler to detect weak reflectors that are mobile, in varied applications (surveillance, mapping, security, etc.) it is important to detect objects that are not moving as well (e.g. a parked car or road sign). While doppler can of course still be a practical solution for detecting relatively few moving objects, we instead focus on what can be done to improve a single capture.

We note that while background subtraction is a naive solution to this problem, because of the the many orders of magnitude larger a noise reflector might be than our given target, even slight positional or power fluctuations between captures can leave very large peaks that make our target difficult to find. Further, background subtraction only addresses this problem for moving objects, not stationary objects that might also be dangerous.

Our approach instead relies on the fact that – because of image segmentation – we are certain that the object we are looking for exists in a given azimuth span, and we also know its object type (e.g. car or person). As a result, we can determine a received-signal-strength upper bound based on the object type and each distance. Thus, in-span reflectors that are significantly higher than expected (and their side lobes) can also be removed as clutter as described in 4.7.1 and target peaks can be detected.

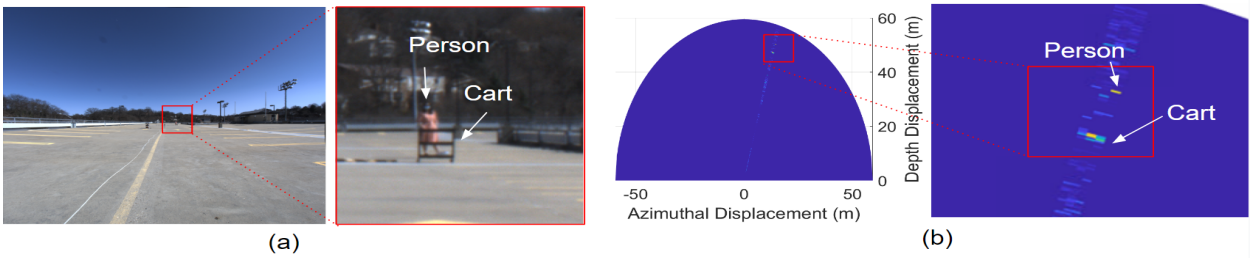

Figure 4.7: **Metamoran vs. Partial Occlusions:** Metamoran can help identify objects-of-interest despite partial occlusions. (a) shows an image of our scene, a person behind a cart, located approximately 45m away. (b) shows Metamoran's capture of the person and the occluding (left) half of the cart. Since image segmentation detected both an unlabeled object and a partially covered person, Metamoran takes the farther reflector as the target.

4.7.3 Impact of Partial Occlusions

Metamoran is also designed to be robust to – and even account for – partial occlusions such as fog or physical obstructions. In the case of physical obstructions, such as the cart in front of a person pictured in Fig. 4.7, image segmentation will generate a mask for both the obstruction and the target. For a known obstruction type, the obstruction can be detected as a target object and then removed as clutter, using techniques explained in 4.7.1 and 4.7.2. In the case of an unknown obstruction, we instead look for two peaks in our azimuth span and take the farther one as our target.

While in some instances of partial obstructions, image segmentation can be fairly robust, it could fail in other instances. However, mmWave radars are known to be fairly resilient to partial occlusions [34] – and we evaluate instances where Metamoran can leverage radar peaks to actively improve segmentation in Sec. 4.9.3. Our discussion in Sec. 4.10 also captures failure modes of this approach, especially for severe occlusions (e.g. heavy fog).

4.8 Implementation

System Hardware: Metamoran is implemented using a FLIR Blackfly S 24.5MP color camera and a TI MMWCAS-RF-EVM RADAR (see Fig. 4.8). We operate the radar at 77-81 GHz with a theoretical range resolution of 3.75-17.8 cm, depending on max range. The radar also has 86 virtual antennas spaced out along the azimuth axis which provides a theoretical azimuth resolution of 1.4°. As explained in Sec. 3.2, this is at least an order of magnitude worse than cameras and lidars. Unlike fusion approaches which rely on processed point clouds [71], this radar supports logging raw complex I/Q samples which is critical for our processing.

Testbed and Data Collection: We test this system in a variety of 200 outdoor scenes such as parking

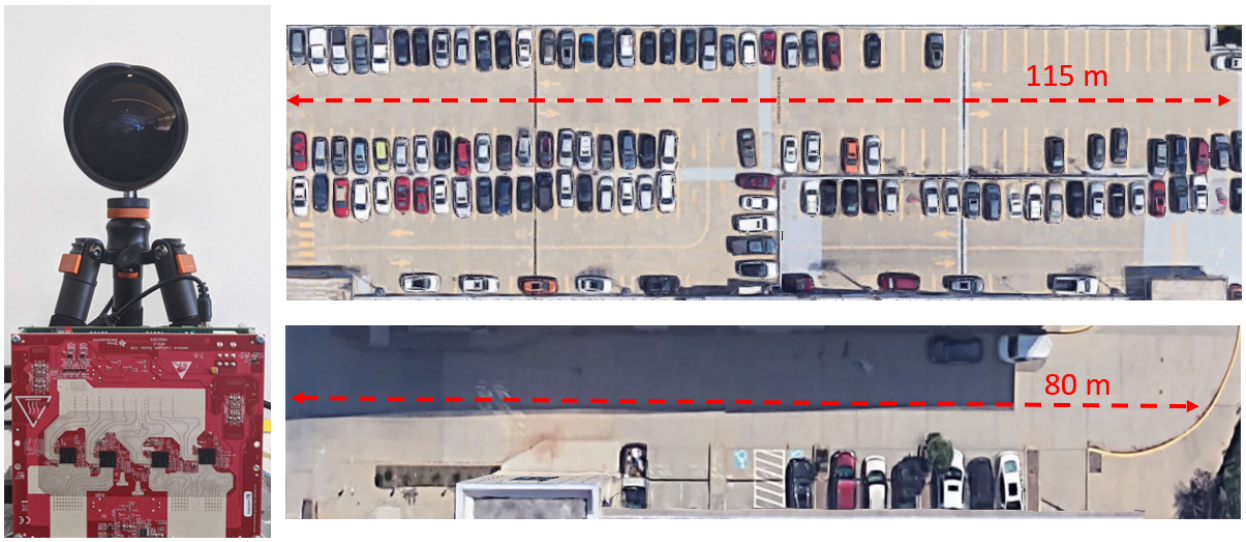

Figure 4.8: **Metamoran's Sensing Platform:** Metamoran is implemented using a FLIR Blackfly S 24.5MP color camera and a TI MMWCAS-RF-EVM mmWave radar. **Evaluation:** Metamoran was evaluated in outdoor spaces like roads and parking lots with rich multipath from buildings, fences, lamp posts, other cars.

lots and roads at distances ranging from 1 m to 60 m from objects-of-interest. These environments have rich multipath arising due to buildings, street lamps, fences, out-of-interest parked cars and pedestrians. Fig. 4.8 shows two candidate locations in the area surrounding a university campus in a major U.S. city.

Ground Truth: We collect ground truth data using a Velodyne Puck LIDAR (VLP-16), which generates 3D point clouds, with fine azimuth and elevation resolutions and 3 cm ranging error. While this lidar is rated for up to 100 m, in practice, on a sunny day, we found the Puck collected data with sufficient point cloud density only until about 30 m. Therefore, for ranges beyond 30 m, we surveyed a point closer to the object-of-interest and placed the lidar at that point.

Baselines: We compare Metamoran with two baselines that use the same hardware platforms: (1) *Naive fusion of Camera and Radar:* We use image pre-processing to obtain the azimuth spanned by object-of-interest. We perform standard radar beamforming for FMCW radar, and bound the output to the azimuth span and then pick the strongest reflector as the target. (2) *Monocular Depth Estimation:* We use state-of-the-art monocular depth estimation algorithm [15] trained to report depth values up to 80 m.

Objects-of-interest Selection: We select a car, a person, and a stop sign for use as our targets, because these are useful for a variety of applications, including smart city and surveillance. Further, these provide a variety of reflectors in size, shape, and reflectivity to evaluate our system. We note that while it is necessary to sense people and cars while they are moving, they are also important to sense when they are stationary – in the case of a delivery truck, an uber, or a child at a bus stop, for example. Indeed,

static objects are much more challenging versus moving objects to detect in radar processing because Doppler-based filtering or background subtraction cannot be used to remove clutter. We therefore focus our evaluation on imaging static objects.

Calibration: We note that Metamoran requires both internal calibration of the components as well as external calibration between the camera and the radar. Internally, our mmWave radar is calibrated using a corner reflector placed at 5m, as described in the TI's mmWave Studio Cascade User Guide [43]. The camera intrinsics are measured by taking many photos of a checkerboard to remove fisheye distortion (using Matlab's Computer Vision Toolbox [101]) and for image segmentation and monocular depth estimation.

Externally, Metamoran requires a consistent understanding of object shapes between the mmWave platform system and the camera system. While both of these are co-located in Metamoran, they are at a small relative distance of 15 cm, which could lead to inconsistencies in the images produced by the two modalities. Metamoran accounts for this using a joint calibration of the mmWave radar and camera using a feature-rich metallic surface that is viewed from both the camera and radar platform to capture a Euclidean transform between their frames of reference. The object is chosen to be feature-rich for both platforms, with stark differences in both color and the presence/absence of strong mmWave reflectors (metallic structures). We note that the transform obtained from calibration is applied, prior to fusing measurements from either platform to ensure consistency.

4.9 Results

4.9.1 Microbenchmarks

Comparing Object Reflectivity

Method: To empirically determine expected power thresholds for detecting target objects in an occluded object, we measure the peak value from radar beamforming for our three target reflectors: car, person, and a road sign, across different distances in 81 line of sight settings.

Results: Our results for this are shown in Fig. 4.9. We observe that power falls off significantly with distance. From about 10 m to 50 m, the reflections attenuate: 16.7× for a car, 63× for a person, and 4.4× for a sign. We note that the sign is a significantly weaker reflector than a person despite being a .762m × .762m metal sheet outfitted with optical retro-reflectors: past work indicates that this may be due to the majority of incident signal being reflected specularly off planes and thus not received by our radar [12].

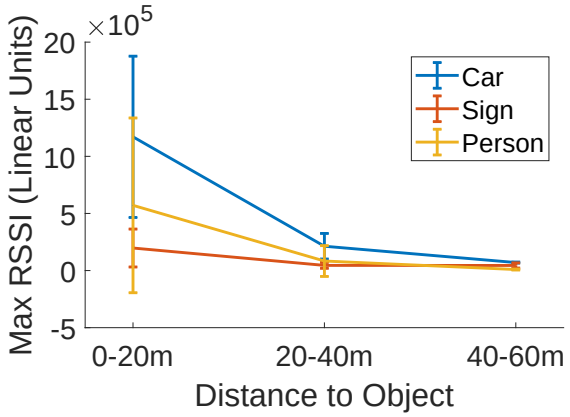

Figure 4.9: Range Attenuation: Reflectivity of an object in line-of-sight conditions after out-of-span SIC has been applied.

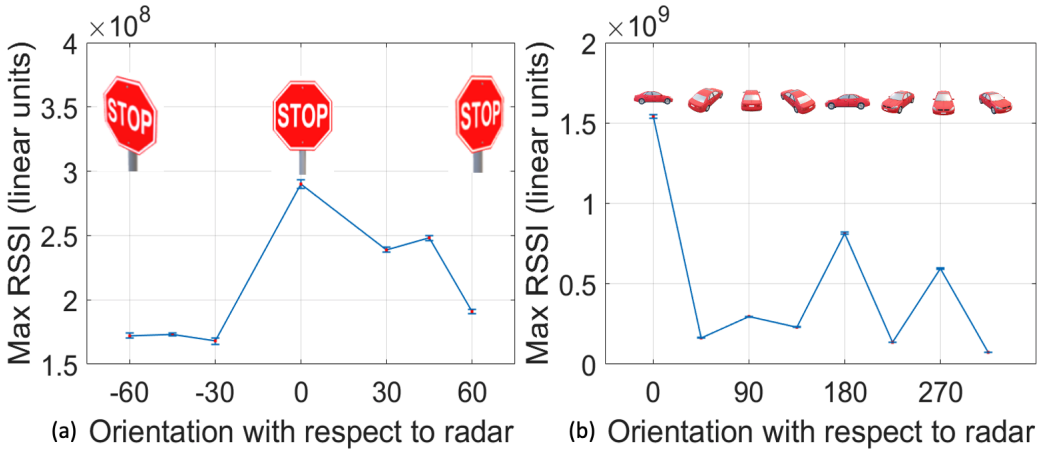

Figure 4.10: Orientation: The magnitude of reflected signal varies with the orientation of our planar targets (sign and car), with peaks at the highest effective area

Impact of Object Orientation

Method: To evaluate the impact of orientation on the reflectivity of our more planar reflectors, we collected data across 7 angles of the front of a stop sign and 8 angles of a car. This data was measured at a fixed 4m away from the object.

Results: The peak values from radar beamforming at different orientation are shown in Fig. 4.10. We find that the peaks correspond, as expected, with the largest effective area: the face of the stop sign, and the side of the car. We find the stop sign peak reflectivity degrades 1.68× at poor orientation, and the car can degrade 21× depending on orientation.

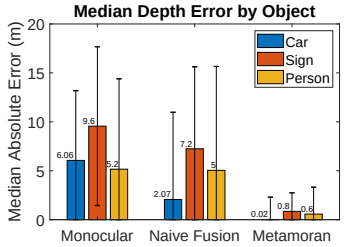

Figure 4.11: Across all algorithms, we see car with the lowest depth error, followed by person, followed by sign. This correlates with each object's reflectivity.

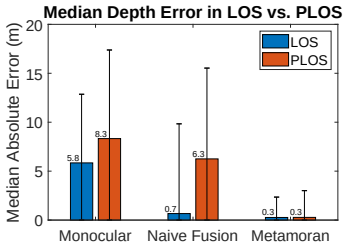

Figure 4.12: Across all algorithms, we see degraded performance in PLOS compared to LOS, particularly in our naive fusion baseline.

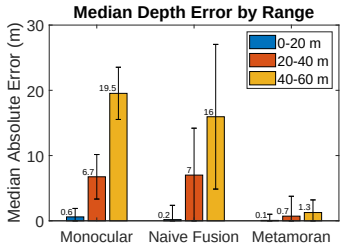

Figure 4.13: Across all algorithms, we see median depth error rise with increased range, with Metamoran showing better accuracy.

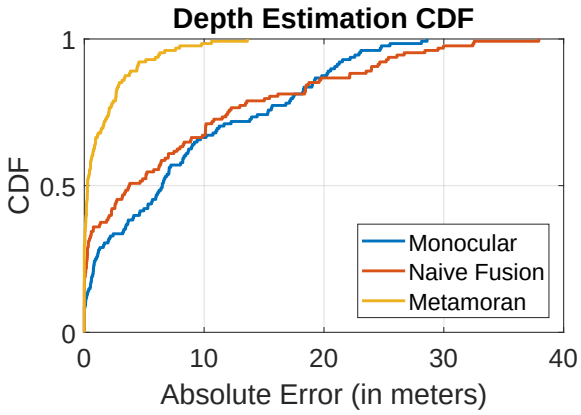

Figure 4.14: CDF of absolute error shows Metamoran is superior to our two baselines in median accuracy.

4.9.2 Depth Resolution

Method: For our range results, we collected 146 data samples in varying lighting conditions at 2 obstacle-rich sites. We collected both line-of-sight (LOS) captures of targets as well as captures of partial line-of-sight (PLOS) occluded by carts, fog, and other environmental objects. Targets were positioned from 3 m to 58 m.

Data was collected in 2 range/resolution buckets: 4.2cm at 0-20m, 11.6cm at 20-60m. The primary bottleneck of range resolution for this system is the TDA2SX SoC capture card that is on the MMWCAS board – it can handle at most a data width of 4096, corresponding to 512 complex samples per receiver. This may be improved with hardware research and advancements, but improvements in that domain are complementary to our approach.

Depth error is measured from one point in each of these approaches (Peak value obtained with naive fusion of radar beamforming and camera, Metamoran estimate and, most repeated value over an object mask for monocular depth estimation) to the depth span provided by the LIDAR.

We compare median error in depth across objects-of-interest for Metamoran and the two baseline systems: naive fusion and monocular depth estimation. We include error bars corresponding to +/- the standard deviation of our collected data. We note that we present median over mean due to the long tail often found in RF localization and sensing that affects both Metamoran and the baseline: slight variances in noise and power can result in disproportionately large errors if the second-largest peak overtakes the first. For systems with a low median error, this effect can be ameliorated by taking multiple snapshots and removing outliers.

We represent three sets of results: (1) three different reflector objects; (2) Partial occlusions including fog and other objects preventing a complete direct view of the object; (3) three different range buckets. Across all experiments, we find that Metamoran significantly outperforms the baselines. We elaborate the performance across each axis below.

Object Results: Fig. 4.11 shows the median error in depth across objects-of-interest for Metamoran and the two baseline systems. We see lowest error for the car across the board due to a combination of factors: the car is our strongest reflector and also offers multiple points on its surface to reflect radar signals due to its size (4.66m x 1.795m). We see performance further degrade with the progressively weak reflectors as measured in Sec. 4.9.1: person is the next most accurate, followed by the sign.

Occlusion Results: Fig. 4.12 shows the median error in depth in line-of-sight (LOS) and partial-line-of-sight (PLOS) for Metamoran and the two baseline systems. We see a particularly significant degradation in our naive fusion baseline for PLOS, which frequently takes the occluding object as the strongest reflector, unlike Metamoran, which can detect and account for occlusions using image segmentation.

Range Results: Fig. 4.13 shows the median error in depth across range for Metamoran and the baselines. As expected, accuracy across all approaches, objects, and occlusion settings deteriorates with range due to weaker received signals.

CDF Results: Fig. 4.14 shows CDF of the median error in depth for Metamoran and the baselines. Metamoran has a median error of **0.28m** across all collected data, compared to 6.5m for monocular depth estimation and 3.75m for naive radar and camera fusion. These correspond to mean values of 1.42m, 8.48m, and 7.89m respectively due to long tail effects.

4.9.3 Depth Imaging

Method: To compute high resolution depth images, we implement the method in Sec. 4.6.2. In contrast to Sec. 4.9.2 which only computed depth errors, here we want to characterize system performance for a point cloud obtained from the baselines monocular depth estimation and naive fusion of camera and

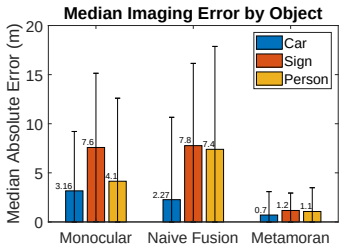

Figure 4.15: Imaging Errors increase with decreasing object reflectivity across algorithms.

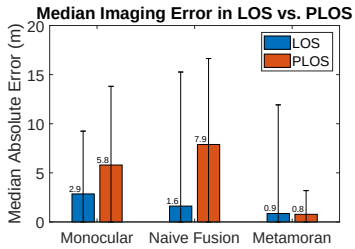

Figure 4.16: Imaging Errors are degraded in partial line of sight scenarios across all algorithms.

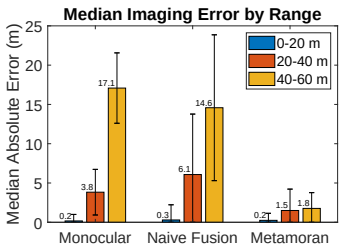

Figure 4.17: Imaging Errors vs. increasing range.

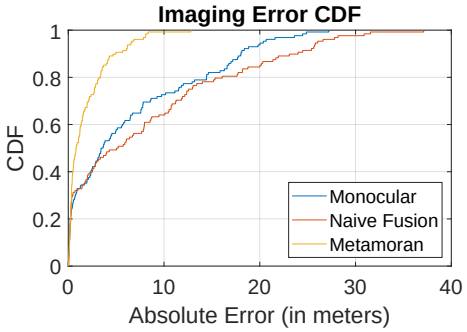

Figure 4.18: This CDF shows that Metamoran significantly outperforms the baselines. The tail in the case of Metamoran is much smaller than that for baselines.

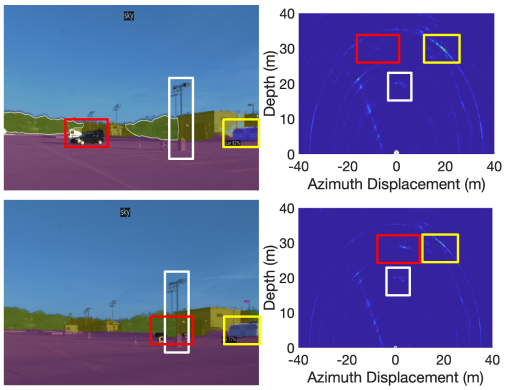

Figure 4.19: Similarly colored boxes contain similar objects across segmentation and radar. While cars in the red boxes are missed by camera, radar still detects them.

radar, and our system against lidar point clouds. Data collection is as similar to that explained in Sec. 4.9.2.

To compare two point clouds A and B, we use a modified version of Hausdorff distance [75] as follows: $\min\left\{a \in Amedian\{b \in Bmin\{d(a,b)\}\}, b \in Bmedian\{a \in Amin\{d(b,a)\}\}\right\}$ where $d(a,b)$ is the distance between points a and b. Hausdorff distance is popularly used in obtaining similarity scores between point clouds. Intuitively, this metric measures the median distance between any two points in the point cloud. The lower the distance, the more similar the point clouds are. We report this distance as imaging error in meters.

Results: Trends in imaging results largely follow those in depth imaging, as problems with detection propagate through the system. We note that shape error is larger than the depth error across the board due to additional pairwise distances being calculated. Figure 4.15 shows the imaging errors against different object types for the 3 different algorithms, Figure 4.16 shows the median error in imaging in line-of-sight and partial-line-of-sight for Metamoran and the two baseline systems, and Figure 4.17 shows the median

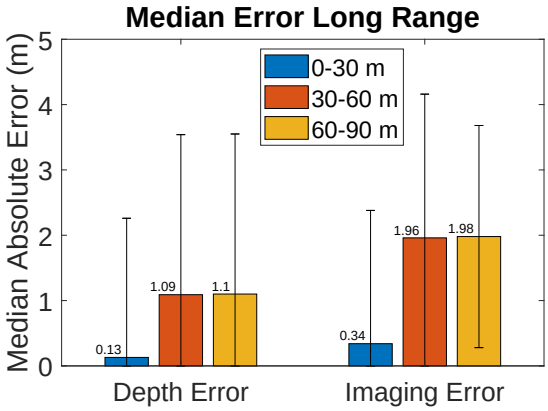

Figure 4.20: This shows median errors for Metamoran depth estimation and imaging performance up to 90m.

error in depth across range for Metamoran and the two baseline systems. Metamoran outperforms both baselines across all categories. We note that in these baselines, monocular depth estimation outperforms naive fusion unlike in 4.9.2. This is because Monocular depth estimation benefits from our metric due to its large azimuth span of many points that are thus more likely to be close to a point in the LIDAR baseline, versus the fewer, and clustered profiles given by naive fusion.

Fig. 4.18 shows CDF of the median error in depth for Metamoran and the two baseline systems. Metamoran has a median error of **0.8m** across all collected data, compared to 3.4m for monocular depth estimation and 5.04m for naive radar and camera fusion. These correspond to mean values of 1.82m, 6.59m, and 8.27m respectively due to long tail effects.

Improving segmentation in PLOS: A point to note that improves our accuracy in partial line-of-sight in Fig. 4.16 is the ability to detect objects that image segmentation misses or offers low confidence on due to occlusions due to obstructions. Fig. 4.19 shows one representative example of this effect for a partial line-of-sight image where an object that was occluded and low-confidence in the camera image was clearly detected based on radar processing.

4.9.4 Range Extension

Method: In addition to the data collected for Sec.4.9.2, we further collect 17 scenes at 2 sites for a large reflector (car) with an additional resolution/range bucket: 17.8 cm at 60-90m. At these extended ranges, car depth is no longer measurable with our baselines, and the sign and person are no longer detectable even with the assistance of Metamoran. We do not collect distances above 90m: since we already observed at 90m that the entire car appears as a single pixel on our radar, distances above this become unreliable.

Results: We show the results for depth resolution and imaging of Metamoran compared to the lidar

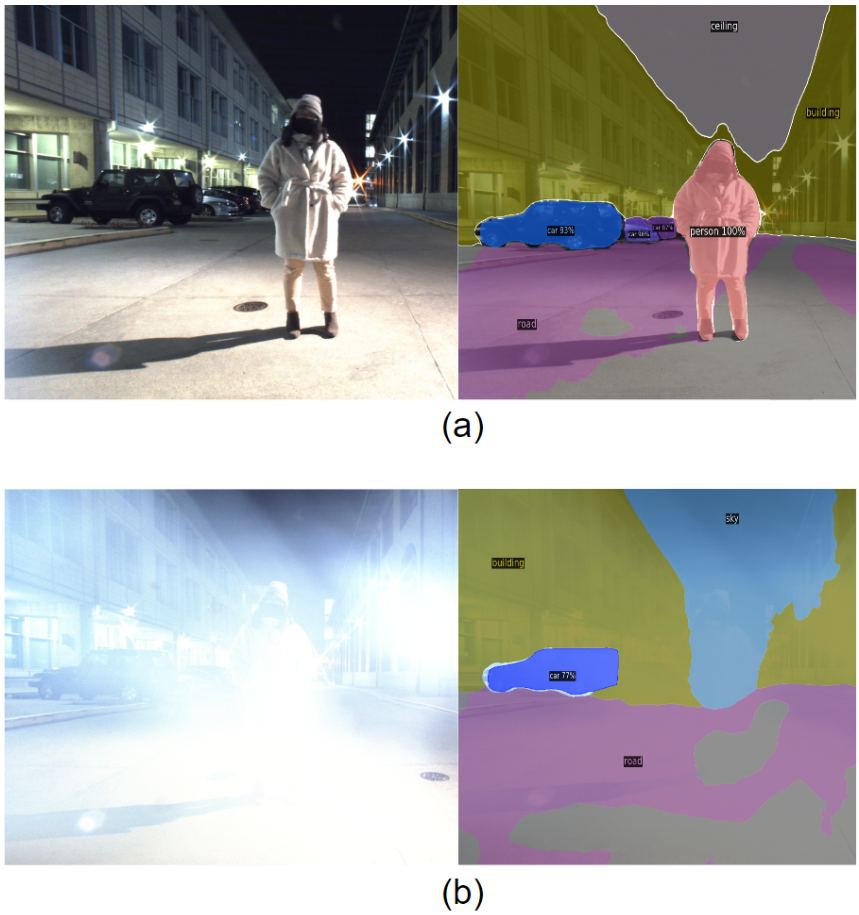

Figure 4.21: **Limitations of Metamoran:** Metamoran can struggle when vision algorithms fail significantly such as complete occlusions (e.g. fog), such as above.

ground truth in Fig. 4.20. We see slight degradation with the increased distance, although it is minimal. We note that the performance degradation in practice is that the reflector is detected less often, particularly in the presence of clutter. At 90m, our 1.4° of azimuth resolution is spaced at 2.2m, and imaging relies very heavily on the successful reception of single pixels.

4.10 Limitations

An important limitation of our system is that its reliance on a camera makes vulnerable to excessive darkness and fully occlusive environmental conditions (e.g. very thick fog). Fig. 4.21 shows one such instance where our system mis-identifies an object (a person) due to heavy fog. We note, in these circumstances, the mmWave RADAR continues to operate and can continue to provide range information for objects in the environment, albeit with attenuated range and with poor angular resolution. For instance, despite the object type in Fig. 4.21 being labeled incorrectly, the depth value reported from mmWave radar is

approximately correct.

Further improvements to calibration could further refine our system and improve results – in particular, an ideal calibration device would be only a pixel large on our camera and also a very strong reflector in mmWave. In practice, this balance is difficult to strike, and we leave further experimentation of calibration materials to future work.

4.11 Summary

This chapter develops Metamoran, a hybrid mmWave and camera based system that achieves high-resolution depth images for objects at extended distances. Metamoran's secret sauce is a novel specular radar processing system that identifies the spatial bounds in azimuth and elevation of objects-of-interest using image segmentation on camera data to improve radar processing along the depth dimension. The resulting system is evaluated on real-world data sets that will be made openly available to obtain depth images of objects-of-interest including pedestrians and cars at distances of up to 60 m. We believe there is rich scope for future work in extending fused mmWave and camera-based depth imaging to broader classes of objects and ensuring resilience to severe occlusions.

Chapter 5

Discussion

5.1 Next Steps

For these projects specifically, the goal for both of these projects were to evaluate whether or not each approach was technically feasible and possibly useful. As such, we did not test our systems exhaustively in every possible environment, obstacle type, obstacle orientation, and so on – in fact, to do so would likely be impossible. However, there could be interesting next steps considering these sensing modalities for specific use cases: disaster relief, autonomous navigation, surveillance, etc. Specific applications could limit the number of specific object types that are of interest, thus requiring less data to set up the system. Furthermore, specific applications would contribute interesting challenges for what might be most of interest, such as non-line-of-sight sensing, mobility, and so on. This could be a rich and interesting space to explore, and each specific set of materials and deployment priorities could provide additional information about the sensing capabilities of these modalities.

5.2 Barriers to Adoption

Beyond technical challenges, we note that despite the ubiquity of WiFi, cameras, and mmWave radar in our society, there exist quite a few barriers to adoption for our proposed system architectures. Though most of these are likely resolvable with some effort, this was not an area we extensively developed. In this section, we discuss some potential pitfalls for widespread use of our systems.

5.2.1 IntuWition

For IntuWition, a few challenges exist as barriers to adoption. For one, the overall setup is a bit bulky for a UAV to carry around for a long time: it is a stack of four Intel Galileo boards attached to four fairly large

antennas, kept in place with foam and duct tape. Volumetrically, of course, the system can be reduced with the use of non-prototype hardware. We also note that with three receiver ports, a single Intel 5300 card *should* be perfectly able of receiving all antenna polarization orientations with a single card, which would halve the number of Intel Galileo boards necessary for this system. In practice, however, we found that the second and third antenna ports of these cards were scaled by some indeterminate quantity – perhaps a hardware artifact, coupling, or some sort of independent automatic gain control. Thus, our results were much more consistent using the first antenna for three individual Intel 5300 cards. We also note that Intel's 9000 series of WiFi cards use M.2 rather than PCIe, which could further miniaturize the footprint (though its channel state information is not openly accessible for the public). For mass deployment, of course, a dedicated board would still provide the smallest form factor. Form factor could also be further reduced by using dedicated 5.8GHz antennas rather than the dual band 2.54/5.8GHz antennas which we found more widely available.

Another challenge for adoption is the threat of interference from environmental objects. Indeed, the 2.54 and 5.8GHz bands are already famously crowded with a variety of connected devices and technologies. While we did not run into any problems with collisions for our experiments, which took place on an urban college campus, this is likely a problem that would be exacerbated as the number of connected devices continues to increase, particularly in urban areas. Furthermore, since IntuWition jumps across many frequencies to improve its performance, this further increases the system's reliance on relatively available channels. We note that the packets used for IntuWition are very short (50 Bytes), which can reduce load on a channel or can be sent multiple times to increase reliability in congested scenarios. Since IntuWition relies on the metadata of wireless packets to operate, it can also function during the transmission of useful data, for example, a video stream from a drone over WiFi. However, the need for channel hopping to function well can complicate this.

5.2.2 Metamoran

Metamoran also faces a few potential barriers to adoption. For one, though mmWave is widely deployed today, the angular resolution of most systems (wireless docking stations, driver-assistance technologies) is quite coarse in practice. For most practical mmWave systems that exist today, it can be argued that angular resolution is not critical for functionality: a car that is nearby while parallel parking, for example, requires essentially the same next steps regardless of whether it is at 12 degrees or 16 degrees. This might limit the large-scale production of mmWave radars available at a low price point, if no killer applications for higher-resolution mmWave radars emerge. Though market factors are of course difficult to predict, because TI has

already developed the MMWCAS-RF-EVM module that we used which offers 1.4° azimuthal elevation, we are optimisitic that industrial research will be able to push forward new applications with the technology now available to them.

Another major challenge to adoption for Metamoran is the difficulties that mobility would introduce, which we did not evaluate. First, of course, there is pretty extensive computation that needs to be simplified to achieve real time: semantic segmentation and depth completion, which both require depth maps, as well as our template simulation approach to removing noise from a radar capture, which has not been optimized at all by runtime. Currently, we run it using MATLAB with 12 threads on a 3.7GHz processor, with 256GB of RAM, and each capture takes a minute or two to process. This could of course be further optimized by improving parallelization of our code, using GPUs, moving it off MATLAB to a more efficient language, etc. These were out of scope for our project, however.

Chapter 6

Conclusions and Future Work

As the world develops to become more connected, it is interesting to consider possibilities that exist for using existing and widely deployed wireless infrastructure. In this thesis, we have explored two such possibilities. The first, IntuWition, explores how WiFi can sense the location and object material of obstacles in the environment. Metamoran takes this a step further, by building on the known angular location and object type of obstacles to be able to determine their depth and imaging profile at high resolution.

There remains a large breadth of future work related to my vision, and for sensing using commodity wireless as a whole. In the context of my vision, there is a glaring piece of future work that is yet to be explored: the intersection of Metamoran and Intuwition. How can WiFi, camera, and mmWave all be explored to provide the most effective imaging of a scene? What opportunities in hybridization exist between WiFi and mmWave, and WiFi and cameras? This is an interesting area of work since WiFi brings much more resilience in NLOS environments than is available to either cameras or mmWave, due to its longer wavelength. Furthermore, this area of research has not been well explored, since there are few sensing environments presently that combine mmwave radar and cameras (both common on autonomous vehicles, for example), with WiFi. One limitation as to why this might be the case is that WiFi as a communication modality naturally suits indoor environments, with a limited range of a few tens of meters in the 5GHz band (what IntuWition uses exclusively). To hybridize WiFi with mmWave and cameras, range extension of WiFi-based sensing would likely be a critical part of problem. Range can be extended a bit by using both the 2.54GHz and the 5GHz ranges, and might also be extended further by correlating received packets with an expected packet. This could be an interesting avenue to explore for future work.

More broadly, I think fusion is likely the way forward for commodity wireless as a sensing modality

to be adopted. Not only is wireless concerning and confusing to the general public – recent 5G skepticism is evidence of that – but it is also foreign to many electrical and computer engineers as well. To be able to present wireless hybridized with familiar sensing systems used by roboticists and computer vision researchers – who are largely the ones developing the sensing technology for self-driving cars, etc. – might encourage the adoption of wireless as a sensing modality in more practical scenarios. This would need to occur in parallel with transparent code and functionality, with specific examples of use cases, to support usability. The adoption of commodity wireless for sensing is further complicated by the different properties present in each frequency band, distinct from advances in other areas of research (camera resolution for example) in that an entirely new set of advantages and challenges occur with each new standard and each new frequency.

Regarding future directions in using commodity wireless as a sensor, I think it is difficult to predict what communication technologies will ultimately be ubiquitous enough to take advantage of for sensing. There is already controversy, for example, that LTE-U might overpower WiFi and other technologies to emerge as the dominant player in unlicensed spectrum. We can look at trends – for example, increasing interest in networks of satellites – as potential new areas to explore using commodity wireless as a sensor. However, a fundamental challenge exists in researching commodity wireless as a sensing modality in academic settings: the reliance on commercially available devices to already be available. Thus, for long-term relevance of this space, I think it is important for future work to be framed around contributions in fundamental sensing approaches that can be applied to other technologies, as well as close collaboration with industrial research.

Bibliography

[1] Artoolkit v5.x. (Accessed on 08/06/2018). 34

[2] Mesa-x-evu evaluation unit datasheet. 16

[3] Millitary radar technology. https://sputniknews.com/news/201504071020549211/. (Accessed on 06/23/2018). 16

[4] *SETI Observations*, 2021. 2

[5] Adib, F., Kabelac, Z., Katabi, D., and Miller, R. C. 3d tracking via body radio reflections. In *NSDI* (2014), vol. 14, pp. 317–329. 11, 16, 30

[6] Adib, F., and Katabi, D. *See through walls with WiFi!*, vol. 43. ACM, 2013. 11, 30

[7] Alahi, A., Haque, A., and Fei-Fei, L. RGB-W: When vision meets wireless. In *Proceedings of the IEEE International Conference on Computer Vision* (2015), pp. 3289–3297. 13

[8] Alessandretti, G., Broggi, A., and Cerri, P. Vehicle and guard rail detection using radar and vision data fusion. *IEEE transactions on intelligent transportation systems 8*, 1 (2007), 95–105. 14

[9] Ali, K., Liu, A. X., Wang, W., and Shahzad, M. Keystroke recognition using wifi signals. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (2015), ACM, pp. 90–102. 11

[10] AudienceProject. Average number of connected devices in u.s. households 2020, 2021. 1

[11] Bahl, P., and Padmanabhan, V. Radar: an in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)* (2000), vol. 2, pp. 775–784 vol.2. 10

[12] Bansal, K., Rungta, K., Zhu, S., and Bharadia, D. Pointillism: Accurate 3d bounding box estimation with multi-radars. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems* (New York, NY, USA, 2020), SenSys '20, Association for Computing Machinery, p. 340–353. 55

[13] Bartlett, M. S. Smoothing periodograms from time-series with continuous spectra. *Nature 161* (1948), 686–687. 10

[14] Bengio, Y., LeCun, Y., et al. Scaling learning algorithms towards ai. 34

[15] Bhat, S. F., Alhashim, I., and Wonka, P. Adabins: Depth estimation using adaptive bins. *arXiv preprint arXiv:2011.14141* (2020). 13, 41, 46, 54

[16] Bian, X., Shao, Y., Tian, W., Wang, S., Zhang, C., Wang, X., and Zhang, Z. Underwater topography detection in coastal areas using fully polarimetric sar data. *Remote Sensing 9*, 6 (2017). 13

[17] Bijelic, M., Gruber, T., and Ritter, W. A benchmark for lidar sensors in fog: Is detection breaking down? In *2018 IEEE Intelligent Vehicles Symposium (IV)* (2018), IEEE, pp. 760–767. 13

[18] Björklund, S., Johansson, T., and Petersson, H. Evaluation of a micro-doppler classification method on mm-wave data. In *2012 IEEE Radar Conference* (2012), IEEE, pp. 0934–0939. 41, 43

[19] Cesar Iovescu, S. R. *The fundamentals of millimeter wave radar sensors*, 2020. 12

[20] Chang, S., Zhang, Y., Zhang, F., Zhao, X., Huang, S., Feng, Z., and Wei, Z. Spatial attention fusion for obstacle detection using mmwave radar and vision sensor. *Sensors 20*, 4 (2020), 956. 13, 41

[21] Chen, C.-C., Higgins, M. B., O'Neill, K., and Detsch, R. Ultrawide-bandwidth fully-polarimetric ground penetrating radar classification of subsurface unexploded ordnance. *IEEE Transactions on Geoscience and Remote Sensing 39*, 6 (June 2001), 1221–1230. 26

[22] Cho, H., Seo, Y., Kumar, B. V. K. V., and Rajkumar, R. R. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (2014), pp. 1836–1843. 14

[23] Clark, B. An efficient implementation of the algorithm'clean'. *Astronomy and Astrophysics 89* (1980), 377. 52

[24] Dandu, K., Samala, S., Bhatia, K., Moallem, M., Subburaj, K., Ahmad, Z., Breen, D., Jang, S., Davis, T., Singh, M., et al. High-performance and small form-factor mm-wave cmos radars for automotive and industrial sensing in 76-to-81ghz and 57-to-64ghz bands. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)* (2021), vol. 64, IEEE, pp. 39–41. 41

[25] Depatla, S., and Mostofi, Y. Crowd counting through walls using wifi. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)(PERCOM)* (March 2018), vol. 00, pp. 1–10. 11

[26] Depatla, S., Muralidharan, A., and Mostofi, Y. Occupancy estimation using only wifi power measurements. *IEEE Journal on Selected Areas in Communications 33*, 7 (July 2015), 1381–1393. 11

[27] Dhekne, A., Gowda, M., Zhao, Y., Hassanieh, H., and Choudhury, R. R. Liquid: A wireless liquid identifier. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (New York, NY, USA, 2018), MobiSys '18, ACM, pp. 442–454. 6, 12, 22

[28] Emmons, G. A., and Alexander, P. M. Polarization scattering matrices for polarimetric radar, Mar 1893. 20

[29] Fang, S., Islam, T., Munir, S., and Nirjon, S. Eyefi: Fast human identification through vision and wifi-based trajectory matching. In *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)* (2020), IEEE, pp. 59–68. 13

[30] FCC. The economic value generated by current and future allocations of unlicensed spectrum, 2009. 1

[31] Gerla, M., Lee, E.-K., Pau, G., and Lee, U. Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds. In *Internet of Things (WF-IoT), 2014 IEEE World Forum on* (2014), IEEE, pp. 241–246. 13

[32] Göhring, D., Wang, M., Schnürmacher, M., and Ganjineh, T. Radar/lidar sensor fusion for car-following on highways. In *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on* (2011), IEEE, pp. 407–412. 22

[33] Group, . W. W. L. W. *IEEE 802.11ac-2013 - IEEE Standard for Information technology–Telecommunications and information exchange between systems—Local and metropolitan area networks–Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications–Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz.*, 2013. 10

[34] Guan, J., Madani, S., Jog, S., Gupta, S., and Hassanieh, H. Through fog high-resolution imaging using millimeter wave radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 11464–11473. 12, 53

[35] Haider, M. K., Ghasempour, Y., and Knightly, E. W. Search light: Tracking device mobility using indoor luminaries to adapt 60 ghz beams. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing* (New York, NY, USA, 2018), Mobihoc '18, Association for Computing Machinery, p. 181–190. 14

[36] Halperin, D., Hu, W., Sheth, A., and Wetherall, D. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review 41*, 1 (2011), 53–53. 18, 31

[37] Han, S., Wang, X., Xu, L., Sun, H., and Zheng, N. Frontal object perception for intelligent vehicles based on radar and camera fusion. In *2016 35th Chinese Control Conference (CCC)* (2016), pp. 4003–4008. 13

[38] Han, S., Wang, X., Xu, L., Sun, H., and Zheng, N. Frontal object perception for intelligent vehicles based on radar and camera fusion. In *2016 35th Chinese Control Conference (CCC)* (2016), IEEE, pp. 4003–4008. 14

[39] Hao, J., Yuan, Y., Ran, L., Jiang, T., Kong, J. A., Chan, C. T., and Zhou, L. Manipulating electromagnetic wave polarizations by anisotropic metamaterials. *Phys. Rev. Lett. 99* (Aug 2007), 063908. 19

[40] Hou, K., Zhou, Z., So, A. M.-C., and Luo, Z.-Q. On the linear convergence of the proximal gradient method for trace norm regularization. In *Advances in Neural Information Processing Systems* (2013), pp. 710–718. 29

[41] Huang, A. S., Bachrach, A., Henry, P., Krainin, M., Maturana, D., Fox, D., and Roy, N. Visual odometry and mapping for autonomous flight using an rgb-d camera. In *Robotics Research*. Springer, 2017, pp. 235–252. 13

[42] Huang, D., Nandakumar, R., and Gollakota, S. Feasibility and limits of wi-fi imaging. *Proceedings of the ACM SenSys* (2014), 266–279. 11

[43] Incorporated, T. I. User's guide: mmwave studio cascade, 2018. 55

[44] Ji, Z., and Prokhorov, D. Radar-vision fusion for object classification. In *2008 11th International Conference on Information Fusion* (2008), IEEE, pp. 1–7. 14

[45] John, V., Nithilan, M., Mita, S., Tehrani, H., Sudheesh, R., and Lalu, P. So-net: Joint semantic segmentation and obstacle detection using deep fusion of monocular camera and radar. In *Pacific-Rim Symposium on Image and Video Technology* (2019), Springer, pp. 138–148. 14

[46] Jose, E., and Adams, M. D. An Augmented State SLAM formulation for Multiple Line-of-Sight Features with Millimetre Wave RADAR. 12

[47] Karanam, C. R., Korany, B., and Mostofi, Y. Magnitude-Based Angle-of-Arrival Estimation, Localization, and Target Tracking. In *Proceedings of ACM/IEEE International Conference on Information Processing in Sensor Networks* (April 2018), pp. 254–265. 11

[48] Karanam, C. R., and Mostofi, Y. 3d through-wall imaging with unmanned aerial vehicles using wifi. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks* (2017), ACM, pp. 131–142. 6, 11

[49] Kato, T., Ninomiya, Y., and Masaki, I. An obstacle detection method by fusion of radar and motion stereo. *IEEE Transactions on Intelligent Transportation Systems 3*, 3 (2002), 182–188. 14

[50] Kianoush, S., Savazzi, S., Vicentini, F., Rampa, V., and Giussani, M. Device-free rf human body fall detection and localization in industrial workplaces. *IEEE Internet of Things Journal 4*, 2 (2016), 351–362. 13

[51] Kim, G., Eom, J., Park, S., and Park, Y. Occurrence and characteristics of mutual interference between lidar scanners. In *Photon Counting Applications 2015* (2015), vol. 9504, International Society for Optics and Photonics, p. 95040K. 13

[52] Kolyadov, D. V., Ligthart, L. P., and Kozlov, A. I. Amplitude-phase method allowing the determination of the complex dielectric permittivity of underlying surfaces using polarimetric radar remote sensing. In *2006 European Radar Conference* (Sep. 2006), pp. 142–145. 19, 28

[53] Kotaru, M., Joshi, K., Bharadia, D., and Katti, S. Spotfi: Decimeter level localization using wifi. *SIGCOMM Comput. Commun. Rev. 45*, 4 (Aug. 2015), 269–282. 10, 11

[54] Kumar, S., Gil, S., Katabi, D., and Rus, D. Accurate indoor localization with zero start-up cost. *Proceedings of the 20th annual international conference on Mobile computing and networking - MobiCom '14* (2014), 483–494. 1, 10, 29

[55] Kumar, S., Shi, L., Ahmed, N., Gil, S., Katabi, D., and Rus, D. Carspeak: a content-centric network for autonomous driving. *ACM SIGCOMM Computer Communication Review 42*, 4 (2012), 259–270. 13

[56] Lee, W. S. L., Fumeaux, C., and Withayachumnankul, W. Metasurfaces for terahertz polarimetry. In *2018 IEEE Asia-Pacific Conference on Antennas and Propagation (APCAP)* (Aug 2018), pp. 514–516. 13

[57] Lekic, V., and Babic, Z. Automotive radar and camera fusion using generative adversarial networks. *Computer Vision and Image Understanding 184* (2019), 1–8. 14

[58] Li, H., Shrestha, A., Fioranelli, F., Le Kernec, J., Heidari, H., Pepa, M., Cippitelli, E., Gambi, E., and Spinsante, S. Multisensor data fusion for human activities classification and fall detection. In *2017 IEEE SENSORS* (2017), pp. 1–3. 14

[59] Li, Z., Lao, M., Phang, S. K., Hamid, M. R. A., Tang, K. Z., and Lin, F. Development and design methodology of an anti-vibration system on micro-uavs. 39

[60] Liu, H., Wang, Y., Zhou, A., He, H., Wang, W., Wang, K., Pan, P., Lu, Y., Liu, L., and Ma, H. Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4*, 4 (2020), 1–28. 43

[61] Liu, X. Airborne lidar for dem generation: some critical issues. *Progress in physical geography 32*, 1 (2008), 31–49. 44

[62] Liu, Z., Cai, Y., Wang, H., Chen, L., Gao, H., Jia, Y., and Li, Y. Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions. *IEEE Transactions on Intelligent Transportation Systems* (2021). 41

[63] Long, N., Wang, K., Cheng, R., Yang, K., and Bai, J. Fusion of millimeter wave radar and rgb-depth sensors for assisted navigation of the visually impaired. In *Millimetre Wave and Terahertz Sensors and Technology XI* (2018), vol. 10800, International Society for Optics and Photonics, p. 1080006. 13

[64] Mazgula, J., Sapis, J., Hashmi, U. S., and Viswanathan, H. Ultra reliable low latency communications in mmwave for factory floor automation. *Journal of the Indian Institute of Science 100*, 2 (2020), 303–314. 43

[65] Mejias, L., McNamara, S., Lai, J., and Ford, J. Vision-based detection and tracking of aerial targets for uav collision avoidance. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on* (2010), IEEE, pp. 87–92. 13

[66] Mitchell, D., Blanche, J., and Flynn, D. An evaluation of millimeter-wave radar sensing for civil infrastructure. In *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (2020), IEEE, pp. 0216–0222. 12

[67] Mullen, L. J., and Contarino, V. M. Hybrid lidar-radar: seeing through the scatter. *IEEE Microwave magazine 1*, 3 (2000), 42–48. 44

[68] Napier, A., Corke, P., and Newman, P. Cross-calibration of push-broom 2d lidars and cameras in natural scenes. In *2013 IEEE International Conference on Robotics and Automation* (2013), IEEE, pp. 3679–3684. 44, 45

[69] Nguyen, P., Truong, H., Ravindranathan, M., Nguyen, A., Han, R., and Vu, T. Matthan: Drone presence detection by identifying physical signatures in the drone's rf communication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services* (New York, NY, USA, 2017), MobiSys '17, ACM, pp. 211–224. 11

[70] Nishio, T., and Ashok, A. High-speed mobile networking through hybrid mmwave-camera communications. In *Proceedings of the 3rd Workshop on Visible Light Communication Systems* (New York, NY, USA, 2016), VLCS '16, Association for Computing Machinery, p. 37–42. 14

[71] Nobis, F., Geisslinger, M., Weber, M., Betz, J., and Lienkamp, M. A deep learning-based radar and camera sensor fusion architecture for object detection. In *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)* (2019), IEEE, pp. 1–7. 53

[72] Olivier, A., Bielsa, G., Tejado, I., Zorzi, M., Widmer, J., and Casari, P. Lightweight indoor localization for 60-ghz millimeter wave systems. In *2016 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (2016), pp. 1–9. 12

[73] Palffy, A., Kooij, J. F., and Gavrila, D. M. Occlusion aware sensor fusion for early crossing pedestrian detection. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (2019), IEEE, pp. 1768–1774. 41

[74] Patel, P., and Holtzman, J. Analysis of a simple successive interference cancellation scheme in a ds/cdma system. *IEEE Journal on Selected Areas in Communications 12*, 5 (1994), 796–807. 52

[75] pdal.io. Point data abstraction library. 59

[76] Poulton, C. V., Byrd, M. J., Russo, P., Timurdogan, E., Khandaker, M., Vermeulen, D., and Watts, M. R. Long-range lidar and free-space data communication with high-performance optical phased arrays. *IEEE Journal of Selected Topics in Quantum Electronics 25*, 5 (2019), 1–8. 13

[77] Prabhakara, A., Singh, V., Kumar, S., and Rowe, A. Osprey: A mmwave approach to tire wear sensing. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services* (2020), pp. 28–41. 51

[78] Praks, J., Koeniguer, E. C., and Hallikainen, M. T. Alternatives to target entropy and alpha angle in SAR polarimetry. *IEEE Transactions on Geoscience and Remote Sensing 47*, 7 (2009), 2262–2274. 13, 17

[79] Praks, J., Koeniguer, E. C., and Hallikainen, M. T. Alternatives to target entropy and alpha angle in sar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing 47*, 7 (July 2009), 2262–2274. 20

[80] Pu, Q., Gupta, S., Gollakota, S., and Patel, S. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking* (2013), ACM, pp. 27–38. 11, 16, 30

[81] Qiu, H., Ahmad, F., Bai, F., Gruteser, M., and Govindan, R. Avr: Augmented vehicular reality. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (New York, NY, USA, 2018), MobiSys '18, ACM, pp. 81–95. 13

[82] Ramasamy, S., Sabatini, R., Gardi, A., and Liu, J. Lidar obstacle warning and avoidance system for unmanned aerial vehicle sense-and-avoid. *Aerospace Science and Technology 55* (2016), 344–358. 13

[83] Reza, M. A., Kosecka, J., and David, P. Farsight: Long-range depth estimation from outdoor images. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2018), pp. 4751–4757. 49

[84] S. Depatla, C. R. K., and Mostofi, Y. Robotic through-wall imaging. *IEEE Antenna and Propagation Magazine, special issue on Electromagnetic Inverse Problems for Sensing and Imaging* (April 2017). 11

[85] Sarbolandi, H., Lefloch, D., and Kolb, A. Kinect range sensing: Structured-light versus time-of-flight kinect. *Computer vision and image understanding 139* (2015), 1–20. 13

[86] Savazzi, S., Rampa, V., Vicentini, F., and Giussani, M. Device-free human sensing and localization in collaborative human–robot workspaces: A case study. *IEEE Sensors Journal 16*, 5 (2015), 1253–1264. 13

[87] Saxena, A., Chung, S. H., Ng, A. Y., et al. Learning depth from single monocular images. In *NIPS* (2005), vol. 18, pp. 1–8. 49

[88] Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation 34*, 3 (1986), 276–280. 10

[89] Schuler, D. L., Lee, J. S., Ainsworth, T. L., and Grunes, M. R. Terrain topography measurement using multipass polarimetric synthetic aperture radar data. *Radio Science 35*, 3 (May 2000), 813–832. 13

[90] Schuler, D. L., Lee, J.-S., and Grandi, G. D. Measurement of topography using polarimetric sar images. *IEEE Transactions on Geoscience and Remote Sensing 34*, 5 (Sep 1996), 1266–1277. 13

[91] Sen, S., Roy Choudhury, R., and Nelakuditi, S. Csma/cn: Carrier sense multiple access with collision notification. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2010), MobiCom '10, ACM, pp. 25–36. 30

[92] Sengupta, A., Jin, F., and Cao, S. A dnn-lstm based target tracking approach using mmwave radar and camera sensor fusion. In *2019 IEEE National Aerospace and Electronics Conference (NAECON)* (2019), IEEE, pp. 688–693. 13

[93] Shi, J., Wang, J., Hsu, A. Y., O'Neill, P. E., and Engman, E. T. Estimation of bare surface soil moisture and surface roughness parameter using l-band sar image data. *IEEE Transactions on Geoscience and Remote Sensing 35*, 5 (1997), 1254–1266. 13, 17

[94] Smolyanskiy, N., Kamenev, A., and Birchfield, S. On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2018), pp. 1007–1015. 13, 41

[95] Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., and Babu, R. V. A taxonomy of deep convolutional neural nets for computer vision. *Frontiers in Robotics and AI 2* (2016), 36. 25

[96] Stanko, S., Notel, D., Wahlen, A., Huck, J., Kloppel, F., Sommer, R., Hagelen, M., and Essen, H. Active and passive mm-wave imaging for concealed weapon detection and surveillance. In *2008 33rd International Conference on Infrared, Millimeter and Terahertz Waves* (2008), IEEE, pp. 1–2. 12

[97] Steinhauer, M., Ruoß, H.-O., Irion, H., and Menzel, W. Millimeter-wave-radar sensor based on a transceiver array for automotive applications. *IEEE transactions on microwave theory and techniques 56*, 2 (2008), 261–269. 47

[98] StereoLabs, 2020. 13, 41

[99] Steux, B., Laurgeau, C., Salesse, L., and Wautier, D. Fade: A vehicle detection and tracking system featuring monocular color vision and radar data fusion. In *Intelligent Vehicle Symposium, 2002. IEEE* (2002), vol. 2, IEEE, pp. 632–639. 14

[100] Texas-Instruments. Ti mmwcas-rf-evm. `https://www.ti.com/tool/MMWCAS-RF-EVM`, 2021. 44

[101] The MathWorks, I. I. 55

[102] Tokoro, S. Automotive application systems of a millimeter-wave radar. In *Proceedings of Conference on Intelligent Vehicles* (1996), IEEE, pp. 260–265. 43

[103] Toriyama, A. Dragonball z: The zeta sword, May 1994. 41

[104] Tse, D., and Viswanath, P. *Fundamentals of wireless communication*. Cambridge university press, 2005. 48

[105] Vasisht, D., Kumar, S., and Katabi, D. Decimeter-Level Localization with a Single WiFi Access Point. *Nsdi 2016*, ii (2016), 165–178. 11, 22, 28, 29

[106] Wang, J., Xiong, J., Chen, X., Jiang, H., Balan, R. K., and Fang, D. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2017), MobiCom '17, ACM, pp. 288–300. 12, 22

[107] Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., and Liu, H. E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2014), MobiCom '14, ACM, pp. 617–628. 11

[108] Wei, T., and Zhang, X. Mtrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2015), MobiCom '15, Association for Computing Machinery, p. 117–129. 1, 12

[109] Wenger, J. Automotive mm-wave radar: Status and trends in system design and technology. 41

[110] Wu, K. Wi-metal: Detecting metal by using wireless networks. In *Communications (ICC), 2016 IEEE International Conference on* (2016), IEEE, pp. 1–6. 13, 22

[111] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019. 46

[112] Xie, Y., Li, Z., and Li, M. Precise power delay profiling with commodity wifi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2015), MobiCom '15, ACM, pp. 53–64. 11

[113] Xie, Y., Xiong, J., Li, M., and Jamieson, K. xd-track: Leveraging multi-dimensional information for passive wi-fi tracking. In *Proceedings of the 3rd Workshop on Hot Topics in Wireless* (New York, NY, USA, 2016), HotWireless '16, ACM, pp. 39–43. 11

[114] Xiong, J., and Jamieson, K. Arraytrack: A fine-grained indoor location system. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)* (Lombard, IL, Apr. 2013), USENIX Association, pp. 71–84. 10

[115] Xiong, J., Sundaresan, K., and Jamieson, K. Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (2015), ACM, pp. 537–549. 11, 30

[116] Xu, C., Gao, M., Firner, B., Zhang, Y., Howard, R., and Li, J. Towards robust device-free passive localization through automatic camera-assisted recalibration. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems* (2012), pp. 339–340. 13

[117] Yang, Z., Pathak, P. H., Zeng, Y., Liran, X., and Mohapatra, P. Monitoring vital signs using millimeter wave. In *MobiHoc* (2016), pp. 211–220. 1, 12

[118] Yeo, H.-S., Flamich, G., Schrempf, P., Harris-Birtill, D., and Quigley, A. Radarcat: Radar categorization for input & interaction. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (2016), ACM, pp. 833–841. 13

[119] Zakharov, A., Zakharova, L., Sorochinsky, M., and Chimitdorzhiev, T. Sar polarimetry in remote sensing of arctic region. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (July 2018), pp. 2394–2397. 13

[120] Zeiler, M. D. ADADELTA: an adaptive learning rate method. *CoRR abs/1212.5701* (2012). 25

[121] Zhang, R., and Cao, S. Extending reliability of mmwave radar tracking and detection via fusion with camera. *IEEE Access 7* (2019), 137065–137079. 13

[122] Zhao, C., Sun, Q., Zhang, C., Tang, Y., and Qian, F. Monocular depth estimation based on deep learning: An overview. *Science China Technological Sciences 63*, 9 (Jun 2020), 1612–1627. 13

[123] Zhengping Ji, and Prokhorov, D. Radar-vision fusion for object classification. In *2008 11th International Conference on Information Fusion* (2008), pp. 1–7. 13

[124] Zhu, Y., Zhu, Y., Zhao, B. Y., and Zheng, H. Reusing 60ghz radios for mobile radar imaging. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2015), MobiCom '15, ACM, pp. 103–116. 12