UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Computer Vision-assisted Battery-free RFID Systems for Object Recognition, Localization and Orientation

by Zhongqin Wang

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Min Xu

University of Technology Sydney
Faculty of Engineering and Information Technology

November 2020

Certificate of Original Authorship

I, Zhongqin Wang declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution.

Also, I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Nanjing University of Posts and Telecommunications, China.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 18/11/2020

ABSTRACT

Computer Vision-assisted Battery-free RFID Systems for Object Recognition, Localization and Orientation

by
Zhongqin Wang
Supervisor: Dr. Min Xu

Battery-free radio frequency identification (RFID) is a promising technique in Internet of Things (IoT) applications that use wireless signals to identify a physical object from its attached RFID tag. Compared to the existing barcode identification systems, RFID can still work in the non-line-of-sight (NLOS) scenarios that some obstructions block the identifier. Recently, many researchers start regarding each RFID tag as a battery-free sensor, whose indicator is the backscatter signal fingerprint reported by an RFID reader. Since the sensor could sense the change in the position and orientation of an RFID tag relative to a reader antenna as well as surroundings, a variety of battery-free RFID sensing systems are proposed for object localization, direction tracking, material recognition, human breathing/heartbeat rate assessment, liquid leakage detection, etc. However, some technical challenges still remain to be addressed in these purely RFID-based systems. This thesis introduces computer vision (CV) techniques into RFID systems to minimize the impact of RF phase periodicity and multipath interference. In the thesis, three categories of CV-assisted battery-free RFID systems for object recognition, localization and orientation are designed, and the main contributions include:

1) This thesis presents RF-Focus, a CV-assisted system that recognizes moving RFID-tagged objects within the region of interest and tracks their trajectories in multipath environments. To achieve RF-Focus, novel RSSI/RF phase-distance models with additional multipath terms compared to traditional models are proposed to characterize the impact of multipath interference, and thereby a dual-reader-antenna solution is designed to deal with it. Moreover, the multipath terms in RSSI and RF phase can be leveraged to clean the phase shift caused by frequency-dependent RFID hardware characteristics in RF phase. After that, an innovative fusion algorithm is designed to match position proposals outputted by a 2D camera and the cleaned RF phase for object recognition. In the experiments, RF-Focus achieves 91.67% ROI object recognition in multipath environments when simultaneously tracking five moving objects.

2) This thesis proposes RF-MVO, a CV-assisted system that locates stationary RFID tags in 3D space without driving a platform carrying reader antennas along a predefined trajectory or pre-deployed track. To achieve RF-MVO, a 2D camera is affixed to reader antennas. A fusion model is designed to fuse camera trajectory in the camera view with depth-enabled RF phase to achieve real-world trajectory transformation and tag DOA estimation. On this basis, a novel 3D localization is

proposed, which could avoid consuming huge computations to search for all possible regions. In addition, a joint optimization algorithm is designed to accelerate RF-MVO and improve its estimation accuracy. Finally, this thesis introduces horizontal dilution of precision widely used in satellite positioning systems to find out the optimal localization result. The experiments show that RF-MVO achieves 6.23cm localization accuracy in 3D space.

3) This thesis proposes RF-Orien3D, a CV-assisted system that leverages the variation of each tag radiation pattern in a two-RFID-tag array to estimate a labeled object's spatial directions (i.e., azimuth and elevation) in multipath environments. To achieve RF-Orien3D, this work proposes novel RSSI/RF phase-distance models when tag mutual coupling and multipath interference both occur. In the models, one variable to be estimated is tag radiation pattern, which is simulated by building a two-tag array from a 2D image; another is modulation factor, which is estimated using RFID fingerprints in non-coupling and coupling in free space. On this basis, a convolutional neural network (CNN)-based method is proposed by simulating all multipath impacts on RFID fingerprints based on the proposed fingerprint models to pre-train a CNN and then collecting measured data to fine-tune the CNN for 3D orientation. In the experiments, RF-Orien3D achieves median angle errors of 29° and 11° in azimuth and elevation.

Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my Ph.D. degree!

First of all, I must thank my supervisor, Min Xu, for her support and guidance. She allows me to study at UTS and leads me into a new research field. Here I could have a chance to exchange new ideas with excellent research teams. I would also like to thank her for my daily help. Thanks to my co-supervisor, Xiaoying Kong, for being my co-supervisor. She also gives me a lot of guidance in my research.

I would like to thank my family for all the support through my research. They work hard to provide me a stable living environment, and their support strengthens my belief in pursuing my life's ideals. I look forward to my future success in my career to repay parental grace.

My studying life at UTS would be such enjoyable with my friends and colleagues: Haimin Zhang, Lingxiang Wu, Zhiyuan Shi, Madhumita, Lei Sang, Ruiheng Zhang, Xiaoxu Li, Wanneng Wu, Yunkun Yang, Shuo Yang, Haodong Chang, and so on. I would like to thank them for helping me solve many research problems.

Thanks again to everyone who helped me!

List of Publications

Conference Papers

C-1. **Zhongqin Wang**, **Min Xu**, Ning Ye, Ruchuan Wang, Haiping Huang and Fu Xiao, "RF-Mirror: Mitigating Mutual Coupling Interference in Two-Tag Array Labeled RFID Systems", *Proc. of IEEE International Conference on Sensing, Communication and Networking(SECON)*, pp. 1-9, 2020. **(CORE B, Acceptance Rate: 27.9% (36/129))**

C-2. **Zhongqin Wang**, **Min Xu**, Ning Ye, Ruchuan Wang and Haiping Huang,"RF-Focus: Computer Vision-assisted Region-of-interest RFID Tag Recognition and Localization in Multipath-prevalent Environments", *Proc. ACM Interactive, Mobile,Wearable and Ubiquitous Technologies*, Vol. 3, No. 1, Article 29, 2019. **(UbiComp 2019, CORE A*)**

C-3. **Zhongqin Wang**, **Min Xu**, Ning Ye, Ruchuan Wang and Haiping Huang, "RF-MVO: Simultaneous 3D Object Localization and Camera Trajectory Recovery Using RFID Devices and a 2D Monocular Camera", *Proc. IEEE International Conference on Distributed Computing Systems (ICDCS)*, pp. 534-544, 2018. **(CORE A, Acceptance Rate: 20% (78/378))**

Journal Papers

J-1. **Zhongqin Wang**, **Min Xu**, Ning Ye, Ruchuan Wang, Haiping Huang and Fu Xiao, "Computer Vision-assisted 3D Object Localization via COTS RFID Devices and a Monocular Camera," *IEEE Transactions on Mobile Computing (TMC)*, 2019, doi: 10.1109/TMC.2019.2954830. **(CORE A*)**

Contents

Certificate ii

Abstract iii

Acknowledgments v

List of Publications vi

List of Figures xi

Abbreviation xiv

1 Introduction 1

1.1 Battery-free RFID in Internet of Things 1

1.2 Battery-free Sensing with RFID 2

1.3 Main Challenges 3

1.4 Research Contents 4

1.4.1 CV-assisted Region-of-interest Moving Object Recognition . . 4

1.4.2 CV-assisted Stationary Object 3D Localization 5

1.4.3 CV-assisted Two-RFID-Tag Labeled Object 3D Orientation . 5

1.5 Outline 6

2 Literature Review 7

2.1 Single-RFID-Tag Localization 7

2.2 Multi-RFID-Tag Localization and Orientation 8

2.3 RFID-CV Fusion Recognition and Localization 8

2.4 RFID-based Device-free Sensing 9

2.5 RFID Phase Shift Calibration 9

2.6 RFID Tag Radiation Pattern Modeling 10

3 CV-assisted ROI Moving Object Recognition 11

3.1 Introduction 11

3.2 Dealing with Multipath Interference 15

3.2.1 Multipath Signal Propagation in LOS 15

3.2.2 RSSI and RF Phase-distance Models in Multipath LOS 17
3.2.3 RSSI and RF Phase-distance Models in NLOS 18
3.2.4 The Shortcomings of Prior RSSI and RF Phase-distance Models 19
3.2.5 Multipath Interference in Two Tightly-spaced Antennas . . . 19
3.3 Dealing with Frequency-dependent Hardware Characteristics 21
3.3.1 Preliminary 21
3.3.2 Multipath Interference-powered Phase-shift Estimation 22
3.3.3 Deriving Multipath RSSI and RF Phase Measurements 23
3.3.4 Calculating Dominant Reflected Path Distance 24
3.3.5 Calculating Tag-to-antenna Distance 25
3.3.6 Refining Hardware-related Phase Shifts 26
3.4 Fusing RFID and CV for Region-of-interest RFID Tag Recognition and Localization 26
3.4.1 Candidate Region Extraction from 2D Image 27
3.4.2 2D Image to 3D World Coordinate Transformation 27
3.4.3 Reader Antenna Localization in Camera-centered Coordinate System 28
3.4.4 RFID and CV Sampling Synchronization 28
3.4.5 RFID-CV Fusion for ROI RFID Tag Recognition and Localization 29
3.5 Evaluation 30
3.5.1 Implementation 31
3.5.2 Effectiveness of Multipath RSSI and RF Phase Modelling . . . 32
3.5.3 Hardware-related Phase Shift Estimation Performance 33
3.5.4 RFID and CV Fusion for ROI RFID Tag Recognition and Localization 35
3.6 Conclusion 40

4 CV-assisted Stationary Object 3D Localization 41

4.1 Introduction 41
4.2 Visual Odometry Background 44
4.3 DOA and Scale Factor Searching for 3D RFID Localization 45
4.3.1 DOA and Scale Factor Searching 45
4.3.2 RFID Tag Localization in 3D Space 49

4.4 Coarse-to-Fine Tag Position and Scale Factor Optimization 51
4.5 Optimal Tag Position and Scale Factor Selection from Multiple Antenna Arrays 52
4.6 Implementation 55
4.7 Evaluation 57
4.7.1 Experiment setup 57
4.7.2 HDOP Performance 58
4.7.3 Tag localization performance 59
4.7.4 Trajectory recovery performance 61
4.7.5 Microbenchmarks 61
4.7.6 Case Study for Mis-shelved Book Detection 65
4.8 Conclusion 65

5 CV-assisted Two-RFID-Tag Labeled Object 3D Orientation 66

5.1 Introduction 66
5.2 Formulating Mutual Coupling Between Two RFID Tags in Multipath Environment 69
5.3 Tracking Object Orientation in 3D Space 71
5.3.1 Orientation Estimation in Low Multipath 71
5.3.2 Convolutional Neural Network-assisted Orientation Estimation in Rich Multipath 73
5.4 Simulating Radiation Pattern of Each Element in Two-RFID-Tag Array 74
5.5 Calculating Scaled Modulation Factor 77
5.5.1 Tag Phase Shift Estimation 77
5.5.2 Scaled Modulation Factor Estimation 79
5.6 Generating Spatial Orientation Labels for CNN 81
5.7 Implementation & Evaluation 82
5.7.1 Implementation 82
5.7.2 Evaluation 83
5.8 Results 84
5.8.1 Orientation Tracking Performance 84
5.8.2 Performance Comparison 86
5.8.3 Microbenchmarks 87
5.9 Conclusion 89

6 Conclusion **90**

6.1 Contributions 90

6.2 Future Work 92

Bibliography **94**

List of Figures

1.1 Battery-free RFID system workflow 1

3.1 A conveyor belt-based application for automatic ROI RFID tagged-object recognition. 12
3.2 Multipath signal propagation in LOS. 16
3.3 Angle of arrival of two tightly-spaced antennas. 20
3.4 The impact of frequency-hopping on RSSI and RF phase. 21
3.5 Experiment setup for tag-to-antenna estimation. 23
3.6 The sum of reflected signals in the case with a dominant multipath interference. 23
3.7 Extract object proposals of moving toy trains. 27
3.8 Experiment setup. 31
3.9 Measured and simulated results in LOS and NLOS. 31
3.10 Impact of Magnetic Materials for RFID Tag Antenna Isolation. . . . 33
3.11 Impact of three reflector materials. 33
3.12 Impact of tag-to-antenna distance at $d_0 = 100cm$. 34
3.13 Impact of tag-to-antenna distance at $d_0 = 200cm$. 34
3.14 Impact of operating frequency. 35
3.15 Measurement accuracy comparison. 35
3.16 The CDF of average power error. 36
3.17 RF-Focus Experiment Setup. 37
3.18 ROI RFID tag recognition performance. 38
3.19 ROI RFID tag localization performance. 38
3.20 Impact of ROI RFID tag number. 39
3.21 Impact of fusion sample number. 39
3.22 Impact of false-positive reading RFID tag number. 39

4.1 DOA estimation model. 46

4.2 Peak searching. 48
4.3 Power spectrum. 48
4.4 DOA averaging. 49
4.5 RFID tag localization. 51
4.6 Uncertainty region of RFID tag localization. 54
4.7 Experiment setup. 57
4.8 HDOP performance 59
4.9 3D tag localization performance. 60
4.10 Tag position error comparison. 61
4.11 Tag position error in x-axis, y-axis and z-axis. 61
4.12 The CDF of scale-factor error. 61
4.13 A example of camera trajectory recovery. 61
4.14 Array element size. 62
4.15 Angular spacing. 62
4.16 Scale-factor spacing. 63
4.17 RFID tag type. 63
4.18 Multipath interference. 64
4.19 Tag mutual coupling. 64
4.20 Experiment setup. 64
4.21 Ordering error. 64

5.1 Illustration of RF-Orien3D design. Due to the impact of tag mutual coupling in a two-RFID-tag array, each tag's original radiation pattern is modified, and their backscatter signal fingerprints carry orientation-dependent features, which can be captured by RF-Orien3D for spatial orientation estimation. 67
5.2 Azimuth and elevation angles in tag-array-centric coordinate system. 71
5.3 CNN architecture with transfer learning. 74
5.4 Building two-RFID-tag array for radiation pattern analysis from a 2D photo. 75
5.5 Tag phase shift over different tag separations. 79
5.6 Experiment setup for orientation labelling. 81
5.7 Actual RFID fingerprint collection in different scenarios. 83
5.8 Orientation tracking performance in different scenarios. 85
5.9 2D orientation error. 86

5.10 3D orientation error. 86
5.11 X-Y plane. 88
5.12 X-Z plane. 88
5.13 RSSI vs. tag-antenna distance. 88
5.14 Impact of tag separation. 88
5.15 Impact of tag-antenna distance. 89

Abbreviation

RFID-Radio Frequency Identification
IOT-Internet of Things
RF-Radio Frequency
EPC- Electronic Product Code
LOS-Line of Sight
RSSI-Received Signal Strength Indication
mmWave-millimetre Wave
CV-Computer Vision
ROI-Region of Interest
MVO-Monocular Visual Odometry
DOA-Direction of Arrival
HDOP-Horizontal Dilution of Precision
CTFs-Channel Transfer Functions
CNN-Convolutional Neural Network
COTS-Commercial Off-the-Shelf
NLOS-Non-Line-of-Sight
CW-Continuous Wave
ASK-Amplitude Shift Keying
PSK-Phase Shift Keying
PR-ASK-Phase-reversal Amplitude Shift Keying
DSB-ASK-Double Sideband-Amplitude Shift Keying
AOA-Angle of Arrival
GMMs-Gaussian Mixture Models
NTP-Network Time Protocol
2D-2 Dimensional
3D-3 Dimensional
VO-Visual Odometry
GPS-Global Positioning System
FPS-Frames Per Second
SVD-Singular Value Decomposition
LTK-Low-Level-Reader-Protocol Toolkit
SLAM-Simultaneous Localization and Mapping

Chapter 1

Introduction

1.1 Battery-free RFID in Internet of Things

Battery-free Radio Frequency Identification (RFID) provides feasible connectivity between the physical world and the internet, which has been becoming a fundamental technology in Internet of Things (IoT). It is a form of wireless communication technique for object identification that uses radio waves to identify an object embedded or pre-attached with a battery-free RFID tag, which has been widely used in many fields from industry to retail, healthcare, manufacturing, entertainment, hospitality, etc. A typical battery-free RFID system consists of an RFID reader, reader antennas and battery-free RFID tags. Its workflow is described in Fig. 1.1: *1)* the reader transmits a radio frequency (RF) signal to space through an antenna transmitter; *2)* an RFID tag receives the signal, producing a backscatter signal back to the antenna receiver; *3)* a host computer analyzes the tag data, including its identifier (i.e., Electronic Product Code, EPC) and other information stored in the tag memory, on upper applications. A series of RFID standards have been proposed for commercial product development and use. A global RFID standard ISO/IEC 18000-63[47] regulates how RFID systems work, what operating frequencies they operate at, how RFID data is transferred, and how communication works between an RFID reader and a passive RFID tag.

Since passive tags do not require an additional power supply to power up, they are much cheaper, smaller and easier to be implemented in practice than other battery-powered identification devices. In traditional barcode scanning systems, we need to manually rotate an object to a certain direction to see the barcode for reliable scanning and scan objects one by one. Fortunately, an RFID system can achieve a

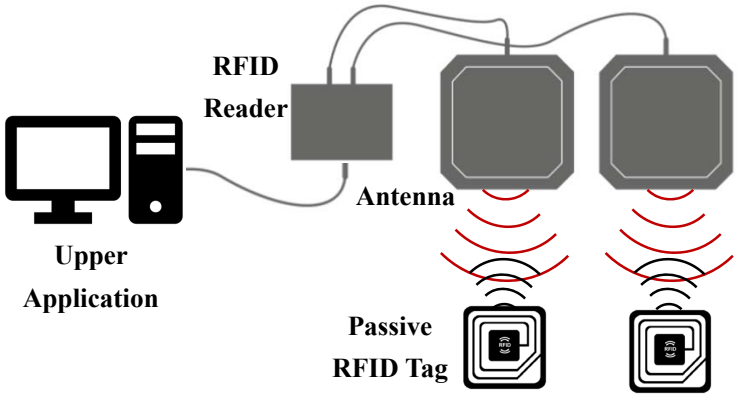


Figure 1.1 : Battery-free RFID system workflow

typical reading range of as long as 6∼10 meters and enable to read many RFID tags from any orientation simultaneously, so RFID could reduce labor cost and accelerate inventory process. More importantly, passive tags do not require line of sight (LOS) to communicate with an RFID reader than other identity carriers like one- and two-dimensional (2D) barcodes. And RFID can protect products against counterfeiting by directly adding a globally unique identification number into each product during a manufacturing process. The RFID-tagged product can be individually tracked through the supply chain in order to identify whether it is authentic or to access additional information about its source for item-level visibility. In total, accurate, automated and real-time information from RFID enables true IoT applications.

1.2 Battery-free Sensing with RFID

Initially, RFID is designed for automatic object identification. Some recent research works have given RFID a whole new sense. They regard an RFID tag as a battery-free sensor attached to an object. The sensor indicator is the backscatter signal fingerprint, including received signal strength indicator (RSSI) and RF phase. Since it varies with environmental change and the distance of a tag to a reader antenna, many battery-free sensing systems have been proposed to achieve RFID localization [110, 65, 64, 106, 108], mechanical vibration period sensing [111, 57], liquid leakage detection [37], material recognition and profile imaging [93], through-wall motion tracking [102, 113], human breathing and heart rate estimation [100, 117, 90, 114], human activity recognition [99, 30, 29, 92] and so on. Compared to other wireless RF signals, e.g., millimeter Wave (mmWave) and Wifi, RFID is a flexible and low-cost solution and can provide ID-dependent object tracking and recognition, making its use appeal to both the industry and the research community. Many IoT applications will benefit from battery-free RFID sensing works. For example,

1) Object Position Tracking. At present, most airline companies are still using barcode systems to identify and sort passengers' luggage, which inevitably requires consuming a lot of time and effort to scan each baggage to different flights. Accurate baggage identification and localization can help sorters improve their sorting efficiency and prevent baggage from being transferred to other wrong sorting conveyors. In industrial fields, quality assurance systems require workers to check the number of products inside packages and quickly locate defective products on the assembly line from these packages with the same appearance. In a library application scenario, when each of the books on shelve is attached with an RFID tag, librarians can easily obtain the order of the books on shelves and find out those misshelved books by obtaining their position information. Developing automatic pick-and-place robots is becoming a promising way to help care for the elderly. It has been of great interest in the robotic community and industry to enable the robot to search for an object, pick it up, and then deliver it to the elderly. The first fundamental task is to recognize this object of interest and thereby locate it in space.

2) Object Authentication for Unmanned Supermarkets. In recent years, unmanned supermarkets with new shopping experience have attracted more and

more attention. According to different approaches used, unmanned supermarkets can be divided into four categories: 1) Amazon Go with pure computer vision (CV); 2) Bingo Box and Easy Go with RFID; 3) TaoCoffee with RFID and CV fusion; 4) F5 future stores with QR code. Among them, RFID-based systems exist a crucial challenge: customers may tear off an RFID tag from a cheap item and then attach it to an expensive item, or someone may use fake RFID tags to cheat this RFID system. One feasible solution is to fuse RFID and CV for object authentication. When checking out, a CV system could depend on the existing deep learning-based methods to recognize each item and then obtain their positions in space. At the same time, an RFID system could also identify each item and locate their spatial positions. By matching ID and position information from RFID and CV, it can effectively avoid the cheating problem.

1.3 Main Challenges

At present, state-of-the-art works have undoubtedly made essential progress in battery-free sensing. However, they still face the following problems in practical object recognition, localization and orientation applications.

1) Region-of-Interest Moving Object Recognition. In some conveyorized application scenarios, RFID systems may only concern with those RFID tags that exist on the conveyor belt. Due to unpredictable RFID signal propagation, however, some RFID tags outside the region may be read, making the systems falsely consider these unwanted tags are located in the region of interest (ROI). An effective solution to deal with this issue is that each RFID tag can be accurately pinpointed in space. In realistic environments, however, the transmitted RF signal from a reader antenna will be reflected off some reflectors (such as tagged objects, walls, and furniture) to an RFID tag, in addition to the signal traveling along the direct tag-to-antenna path. This phenomenon is called the multipath effect. In this case, the distorted signal fingerprint cannot accurately indicate the tag-to-antenna distance for tag localization. Hence, how to minimize the multipath interference for fine-grained ROI moving object recognition is a challenge.

2) Stationary Object 3D Localization. Since RF phase repeats from 0 to 2π radians every half a wavelength, several possible candidate positions may exist in a surveillance region. This phenomenon is called RF phase periodicity. Some existing solutions in stationary RFID localization fix reader antennas to a robotic platform and then move it along a pre-deploying track at a constant speed. And some deploy an antenna on a drone and use a very complicated camera-based system to capture its trajectory in space. Besides, most of the existing solutions need to pre-define a surveillance region where the tracked RFID tags may exist and divide the region into centimeter-level or even smaller grids for position estimation. However, the range of this pre-defined region is usually unknown in advance. And if a large region is given, the vast computation for searching will seriously affect the real-time performance. Therefore, how to achieve low-latency and high-precision RFID spatial positioning when a reader antenna moves along an unknown trajectory is a challenge.

3) Multi-Tag Labeled Object 3D Orientation. Multi-tag based sensing is

becoming a hotspot in recent research. More RFID tags on a tracked object can not only provide more signal fingerprints for sensing optimization but also eliminate RF phase periodicity. More importantly, the multi-tag systems could offer an opportunity to estimate the direction of the RFID-tagged object in space. However, when these RFID tags are deployed very close to each other, the electromagnetic fields between them will interfere with each other, thereby causing the serious tag mutual coupling effect. This effect will result in unexpected changes in the reported RSSI and RF phase of each tag compared to the tag located in space by itself. Therefore, how to suppress the impact of tag mutual coupling in a multi-tag system is a challenge.

1.4 Research Contents

This thesis explores how to combine RFID and CV for object recognition, localization and orientation in practical scenarios. This RFID-CV fusion could provide many potential benefits to improve the sensing reliability and accuracy compared to purely RFID-based solutions.

1.4.1 CV-assisted Region-of-interest Moving Object Recognition

To minimize the impact of multipath interference, we propose RF-Focus, a CV-assisted RFID system that recognizes and tracks moving RFID-tagged objects within the ROI in multipath-prevalent environments. RF-Focus cleans reported RF phase by removing additional phase shifts caused by multipath interference and frequency-dependent hardware characteristics. Then it pairs the cleaned RF phase with image candidate regions that are likely to contain the RFID-tagged objects to be tracked for recognition and localization. However, we need to address three challenges:

1) Multipath Interference Modeling and Suppression. We build a multipath propagation model to characterize the impact of multipath interference on RFID signal fingerprints. On this basis, we further propose a dual-antenna solution to combat multipath interference by utilizing two tightly-spaced antennas.

2) Frequency-dependent Hardware Characteristic Removal. Different operating frequencies cause different phase shifts in RF phase. To remove its impact in RF-Focus, we compute the tag-to-antenna distance only using one reader antenna and several RFID tags by utilizing the changes in RSSI and RF phase due to multipath interference. In this case, the constant hardware-related phase shifts can be estimated and then removed at all available frequencies. Before RFID and CV fusion, we need to remove the phase shift from raw RF phase.

3) RFID-CV Fusion for ROI Moving Object Recognition. We fuse CV image information with RFID to minimize tag position uncertainty caused by RF phase periodicity. To realize it, we firstly conduct a series of preliminary works, including image candidate region extraction, 2D image to 3D world coordinate transformation, reader antenna localization in the camera-centered world coordinate system and sampling synchronization. Then we design an RFID and CV matching score by associating the cleaned RF phase and image data to recognize each ROI RFID tag and obtain its corresponding trajectory with real-time performance.

1.4.2 CV-assisted Stationary Object 3D Localization

To deal with the limitation of acquiring a reader antenna's position at each sampling time and improve real-time performance in 3D localization, we propose RF-MVO, a CV-assisted RFID system for stationary 3D localization. In RF-MVO, a 2D monocular camera is mounted on one or more reader antennas. As the antennas move across target RFID tags, an RFID reader collects RFID data, including EPC, RF phase, and reading timestamp. At the same time, the camera captures a sequence of 2D images. RF-MVO does not require the LOS between the camera and RFID tags and only relies on the existing monocular visual odometry (MVO) technique to estimate antenna/camera trajectory in the camera view. By fusing depth-enabled RF phase, RF-MVO estimates the trajectory in the real-world view and then pinpoints RFID tag positions in 3D space. To achieve such a system, we need to cope with three challenges:

1) DOA and Scale Factor Searching for 3D Localization. We emulate a virtual antenna array following antenna mobility and then build a direction of arrival(DOA)-based spatial power spectrum to search the spatial directions of the RFID tag relative to this antenna array, as well as a scale factor for real-world camera/antenna trajectory recovery. On this basis, we propose a novel algorithm to calculate the interaction of the spatial lines passing through the tag and antenna elements in each array as the final tag position in 3D space. Since directly searching a high-resolution DOA and scale factor in the spatial power spectrum is a time-consuming process, we only obtain the coarse DOA and fine scale factor as initial values for the following optimization algorithm that can accelerate our system.

2) Coarse-to-Fine Tag Position and Scale Factor Optimization. An optimization algorithm is proposed to balance our system's estimation accuracy and real-time performance. The algorithm consists of three parts. DOA refinement is to zoom in the DOA with a higher searching resolution. 3D RFID localization is to locate the tag position with the refined DOA and scale factor. Nonlinear optimization is to perform the optimization of the scale factor and tag position, which then recursively performs scale-factor-dependent DOA refinement and tag localization.

3) Optimal Tag Position and Scale Factor Selection. We rely on Horizontal Dilution of Precision (HDOP), widely used in satellite positioning systems to measure the confidence level of tag position and scale factor results due to the change in tag-antenna geometry. It can help our system output the optimal tag position and scale factor for trajectory recovery.

1.4.3 CV-assisted Two-RFID-Tag Labeled Object 3D Orientation

To deal with the distortion in RFID signal fingerprint due to tag mutual coupling when multiple tags are simultaneously attached on an object for orientation tracking, we propose RF-Orien3D, a CV-assisted RFID system that enables tracking 3D object directions (i.e., azimuth and elevation) in multipath-rich scenarios, where the tracked object is attached with a two-RFID-tag array, and a reader antenna is fixed. At a high level, when two tags are located closely, the tag mutual coupling effect

will be strong and thereby result in the change in the radiation pattern (i.e., tag gains in all spatial directions) of each tag. RF-Orien3D operates by capturing the variation for orientation tracking. To achieve RF-Orien3D, we need to solve three challenges:

1) Mutual Coupling Modeling in Multipath Environment. In a scenario with two closely-spaced RFID tags, we model the RFID backscatter signal of a responding tag in channel transfer functions (CTFs) by taking into account a mutual signal emitted from its adjacent tag and some reflected signals from surroundings. Then we transform the reported RSSI and RF phase to represent the signal's amplitude and phase shift. In this case, we propose the improved RSSI and RF phase-distance models with additional coupling and multipath terms compared to most of the existing models.

2) CNN-Powered 3D Orientation Tracking. We design a convolutional neural network (CNN)-based method for orientation tracking, which consists of two components: the first is to simulate a large number of RFID fingerprints based on the extended models in all possible multipath environments to pre-train a CNN; the second is to collect actual RFID fingerprints to fine-tune this pre-trained CNN. Once fine-tuning is over, the updated CNN can be applied to predict the RFID-tagged object direction in space.

3) Radiation Pattern Simulation and Scaled Modulation Factor Estimation. In the RFID fingerprint models, two unknown parameters are required to be estimated in advance: one is the modified tag radiation pattern, which is simulated by building a two-tag array model from a 2D image; another is scaled impedance-dependent modulation factor, which is determined using RFID fingerprints in the non-coupling and coupling cases.

1.5 Outline

The rest of this report is organized as follows. Chapter 2 reviews some related works. Chapter 3 describes how to achieve RF-Focus that cleans additional phase shifts in the reported RF phase caused by multipath interference and hopping frequency, and combines the cleaned RF phase with image proposals to track moving RFIDs. Chapter 4 describes how to achieve RF-MVO that fuses camera/antenna trajectory in the camera view outputted by MVO with depth-enabled RF phase to locate stationary RFIDs. Chapter 5 describes how to achieve RF-Orien3D that captures the changes in radiation pattern caused by tag mutual coupling for object orientation tracking in 3D space. The above three chapters are summarized from my published or submitted papers. Finally, we make a conclusion in Chapter 6.

Chapter 2

Literature Review

2.1 Single-RFID-Tag Localization

Many previous research works use RSSI to achieve RFID positioning following the relationship between signal power loss and propagation distance in free space. In a 2D plane, a tag position is determined by calculating the intersection of the two circles whose centers are the locations of two antennas and radiuses are the estimated tag-to-antenna distances. In 3D space, three antennas are required to achieve spatial positioning. However, RSSI is sensitive to multipath interference and the unknown radiation pattern of a reader antenna in different directions, which may not be a reliable metric to indicate the antenna-to-tag distance accurately. To improve RSSI-based localization accuracy, some works [71, 118, 15, 95] pre-deploy many reference RFID tags in a surveillance region at known positions and then compare their RSSI values with that of the tracked tag for localization. The final position accuracy is subject to the number of reference tags, and not all scenarios can allow people to conduct this preparation.

Recently, some commercial off-the-shelf (COTS) RFID readers enable outputting RF phase with higher position resolution than RSSI, which has been paid much more attention in RFID sensing applications. RF-IDraw [96] and BackPos [61] need to deploy many reader antennas at given positions to eliminate phase periodicity. However, they can only locate a tag in a 2D plane. Some motion-based methods like MobiTagbot [79], RF-Scanner [60], AdaRF [108] and RF-3DScan [17] require deploying reader antennas on a robot moving along an already-known trajectory at a predefined speed. They leverage the antenna movement to emulate a set of virtual reader antennas and build a holography to locate RFID tags. However, some magnetic tracks the robot moves along are required to be deployed in advance. Similarly, RFly [64] deploys an antenna relay on a drone and relies on a complicated camera-based system to capture its fly trajectory in space. Tagoram [110] and TrackT [104] can track moving RFID tags at cm-level accuracy. They first specify a surveillance region and then divide it into mm-level grids. On this basis, they depend on inverse synthetic aperture radar technique to estimate the likelihood of being an RFID tag at each grid. However, they require moving an RFID tag along a given trajectory with certain regulations and constraints, which may make these approaches still far from meeting the requirements of real-life applications. And as the surveillance region increases in 3D space, huge computation will jeopardize real-time localization performance. PolarDraw [80] and Pantomime [82] can only achieve relative position tracking for moving RFID tags by exploiting two linearly reader antennas polarization mismatch to a tag. RFind [65] powers narrowband

RFID tags to work over a large virtual bandwidth, without any RFID hardware modification. By computing the time-of-flight from a reader antenna to an RFID tag, RFind estimates its absolute tag-to-antenna distance. Then two or more antennas are exploited for 2D or 3D tag localization, respectively. However, RFind requires customized devices to produce additional ultra-wideband RF signals. And it also needs to take a few seconds to hop over specified frequencies for localization. The latency challenges the accuracy in moving object tracking.

2.2 Multi-RFID-Tag Localization and Orientation

Many recent works start attaching an RFID tag array consisting of two or more RFID tags on an object for localization and orientation tracking. In particular, the tag-to-tag spacing is limited within a quarter of wavelength to combat phase periodicity. One-more-tag [106] uses RF phase measurements to calculate the distance difference of each reader antenna to two RFID tags. Since the geometry relationship among these tags is known, a pair of RFID tags can produce a hyperbola. Given multiple reader antennas, all hyperbolas may intersect at a point that a tracked tag may be present in theory. RF-Dial [18] requires moving the tagged object and depends on the difference in tag-to-antenna distances to estimate its translation and rotation matrices. Spin-Antenna [91] leverages the polarization matching/mismatching between RFID tags and a linearly polarized antenna to track tagged object movement. However, the sensing accuracy of these systems is subject to tag mutual coupling interference. Tagyro [105] is the first system to deal with the coupling effect for orientation tracking. It requires estimating the distorted geometry among RFID tags in space by rotating the tag array, and then transforms the phase differences of arrival between RFID tags into an orientation spectrum that can characterize the likelihood of the object orientation. However, since the impact of tag mutual coupling on measured RSSI and RF phase is dependent on different tag-to-antenna orientations, its pre-estimated tag geometry will keep changing. In addition, some works [22, 109, 38] instead leverage tag mutual coupling for localization. They deploy many RFID tags in the region in advance. Once a tracked tab is close to one of a reference tag, tag mutual coupling may result in the obvious change in RSSI or RF phase of the reference tag. Based on this phenomenon, these works could achieve coarse-resolution position estimation. However, many reference tags require to be pre-deployed in the surveillance region.

2.3 RFID-CV Fusion Recognition and Localization

Recognizing an object and capturing its location in an image is a general task in CV fields. Since a series of deep neural network models are proposed by feeding a large-scale annotated dataset, recent works [34, 77, 41, 78, 76] have achieved high-precision recognition accuracy with real-time performance. When inputting an image into the detector, the systems enable to determine whether a target object exists in the image and if so where it occurs in the image. Compared to RFID systems, however, the image-based solutions exhibit three limitations in some scenarios. First, it is not practical in ad-hoc scenarios due to lack of labels for training. Second, it is

difficult to distinguish two objects with the similar appearance in the camera view, e.g., express parcels with different object inside. Third, such image-based solution only work well in LOS scenarios, and its estimation accuracy is much sensitive to the light change. RFID-based systems can read the information of attached RFID tags, providing accurate object identification, and perform in non-line-of-sight (NLOS) scenarios as long as the backscatter signals could penetrate through obstacles.

Recently, there is a growing interest in RFID-CV fusion for object recognition and tracking. RF Vision [62] designs a mobile robot equipped with a stereo camera, RFID devices and a laser range finder scan to infer relative positions among RFID-tagged objects. RF-ISee [84] and ID-Match [56] find the optimal matching between RFID tags and detected moving blobs. These works mainly rely on RSSI measurements, which have been proven a fairly unreliable parameter. TagVision [25] uses a 2D camera to capture the trajectories of multiple mobile objects and then differentiates from them according to the correlations between RF phase and the distance of the camera to motion blobs. However, when multiple objects with the same appearance move close to each other, CV-based detections may not be accurately assigned to each track over time. In "Tell me what I see" [107], a Kinect V2 depth camera and two reader antennas are fixed on a rotation-enabled platform. By simultaneously rotating them to scan RFID-tagged objects, the system can associate the depth of camera field with the tag-to-antenna distance calculated by RF phase for target matching. However, since the system needs to take some time to continuously sample, identifying moving tagged objects remains challenging.

2.4 RFID-based Device-free Sensing

Device-free sensing has recently received considerable attention. Each object to be tracked does not carry any sensor devices or RFID tags. Compared to other existing RF techniques like FMCW [7, 8, 6, 9] and WiFi [74, 97, 98, 101], RFID tags are regarded as battery-free signal received antennas. These battery-less RFID tags have a short-range sensing zone, which can effectively avoid unwanted interference from non-sensing targets. However, the capability of these tags to capture the reflections off the target will be weaker than FMCW and WiFi. To deal with this issue, many existing works build a tag array consisting of many tightly-spaced RFID tags for sensing. Tadar[113] is the first device-free RFID work for localization through wall by extracting the reflection off the human body. RF-HMS [102] can determine the forward or backward motion direction in NLOS scenarios. RF-IPad [23], GRfid [121] and RF-finger [89] track finger movements based on the changes in RSSI and RF phase for gesture recognition. TagScan [93] is a system that can identify the material type and image the target horizontal cut. Compared to RFID-tagged works, device-free sensing cannot achieve fine-grained distinction among tracked objects.

2.5 RFID Phase Shift Calibration

D-watch [94] calibrates the hardware-related phase shifts by manually measuring the tag-to-antenna distance. However, this solution cannot provide enough accurate calibration result. Tagspin [26] locates multiple RFID reader antennas using more

than two RFID-tagged rotating disks with a constant moving speed, which could effectively save antenna deployment time. Tagyro [105] can calibrate RF phase over different operating frequencies to a reference frequency such that the output RF phase looks like coming from a common frequency. However, it is challenging to estimate RF phase ambiguity difference by measuring the tag-to-antenna distance difference between the tag position and the initial position at different times. The works [10, 65] perform with a well-isolated bistatic antenna configuration to minimize the effect of the hardware-related phase shifts. However, most COTS RFID readers can only support the monostatic antenna configuration. The RFID systems [58, 80] only work at a specific operating frequency to combat the frequency-dependent phase shifts due to hardware characteristics. In this thesis, our work will introduce how to accurately estimate the hardware-related phase shift in advance by accurately measuring the tag-to-antenna distance.

2.6 RFID Tag Radiation Pattern Modeling

Pattern multiplication theorem [14] is used to model the radiation pattern of a tag array. However, it requires that each antenna is uncoupled, which cannot be applied in multi-RFID-tag lalebled systems. Some works [12, 66] indicate that tag mutual coupling may affect the reading performance of stacked RFID tags. However, they need specialized devices to measure mutual impedances between tags. Feng et al. [63] demonstrate that tag mutual coupling can enhance the tag gains in some tag-to-antenna directions while weakening the gains in others. However, the analysis model ignores the variation in the radiation pattern of each tag element due to the mutual coupling effect. Stefano et al. [20] designs a two-tag array system for building crack monitoring based on the fact that mutual impedance is sensitive to the change in the tag-to-tag spacing. However, the work is unable to calculate the tag radiation pattern. In the thesis, we generate an RFID tag model from an image and construct a tag array to simulate each tag's radiation pattern using antenna analysis components in Matlab.

Chapter 3

CV-assisted ROI Moving Object Recognition

Capturing RFID tags in the ROI is challenging. Many issues, such as multipath interference, frequency-dependent hardware characteristics and phase periodicity, make RF phase difficult to accurately indicate the tag-to-antenna distance for RFID tag localization. In this work, we propose a comprehensive solution, called RF-Focus, which fuses RFID and CV techniques to recognize and locate moving RFID-tagged objects within ROI. Firstly, we build a multipath propagation model and propose a dual-antenna solution to minimize the impact of multipath interference on RF phase. Secondly, by extending the multipath model, we estimate phase shifts due to hardware characteristics at different operating frequencies. Thirdly, to minimize the tag position uncertainty due to RF phase periodicity, we leverage CV to extract image regions of being likely to contain ROI RFID-tagged objects and then associate them with the processed RF phase after the removal of the phase shifts due to multipath interference and hardware characteristics for recognition and localization. Our experiments demonstrate the effectiveness of multipath modeling and hardware-related phase shift estimation. When five RFID-tagged objects are moving in the ROI, RF-Focus achieves the average recognition accuracy of 91.67% and localization accuracy of 94.26% given a false positive rate of 10%.

3.1 Introduction

Passive radio frequency identification is a promising object identification technique that uses wireless RF signals to read the globally unique identity (i.e., EPC) of an object from its affixed battery-free RFID tag. Compared to other identity carriers like one or two-dimensional barcode, passive RFID tags do not require LOS to communicate with an RFID reader, which has been widely used in automated inventory management, product anti-counterfeit and real-time retail loss prevention. In typical RFID applications, a COTS RFID reader is equipped with one or more directional RFID reader antennas to monitor RFID-tagged objects passing through a door, gateway or other specific regions. Let us consider an RFID-based belt conveyor system shown in Fig. 3.1. Reader antennas are mounted on a gantry to identify which RFID-tagged objects are moving in the ROI on the conveyor. Unfortunately, since a commonly-used antenna has a relatively large reading range, some RFID tags carried by people, on other nearby belt conveyors or in package staging areas outside the ROI may be unintentionally read. This phenomenon is called *false positive reading*. To avoid the system erroneously consider these false-positive reading tags as ROI ones, accurately locating ROI RFID tag positions is of importance for conveyorized applications. For example, in the RFID-assisted airport sorting systems, accurate ROI RFID tag identification and localization can

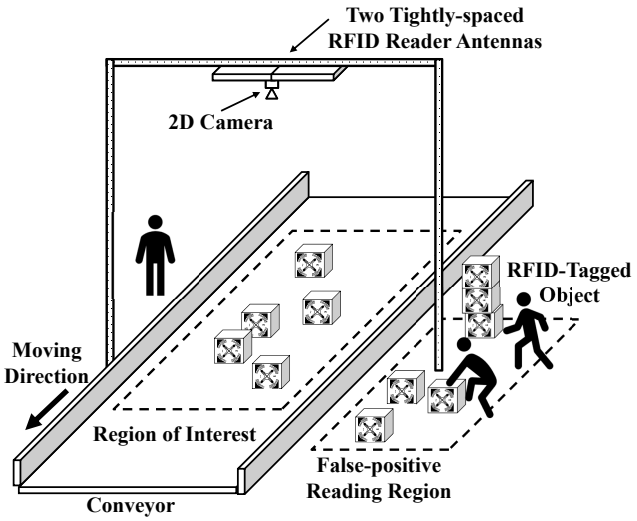


Figure 3.1 : A conveyor belt-based application for automatic ROI RFID tagged-object recognition.

help sorters improve their sorting efficiency and prevent RFID-tagged baggage from being transferred to wrong sorting conveyors. In the RFID-assisted quality assurance system, it can help workers more quickly and easily identify defective products on the assembly line compared to manual checks or barcode systems.

Many previous RFID solutions [24, 54] to address false-positive readings are achieved by reducing the signal transmission power of an RFID reader in order to decrease the size of the reading region. However, the lower power may result in increasing the number of false-negative readings, i.e., the ROI RFID tags are unable to be identified. Accurately locating RFID tags can be a feasible solution to distinguish ROI or false-positive reading tags. However, since the reported RF phase for the tag-to-antenna indicator is repeated every half a wavelength the signal travels within the range of 0 to 2π (i.e., RF phase periodicity), most of the existing motion-based RFID localization approaches [110, 79, 81] require moving a tag or reader antennas along a given trajectory with certain regulations and constraints, which may make these approaches still far from meeting the requirements of real-life applications. A recent RFID localization technique, RFind [65], computes the time-of-flight from a reader antenna to an RFID tag for absolute tag-to-antenna distance measurement. In [65], two or more antennas are exploited for 2D or 3D tag localization, respectively. However, RFind needs to take a few seconds to hop over specified frequencies for localization. The latency makes tracking a moving RFID-tagged object in real-time very challenging.

Furthermore, in a real environment when an antenna transmitter sends radio waves to space, the received signal at the antenna receiver is likely to be interfered with the signal reflections off floor, wall, conveyor components and even RFID-tagged objects, in addition to the expected reflection off an RFID tag. This phenomenon is called multipath interference. Due to the impact of multipath interference, the RF signal fingerprints reported by a COTS RFID reader, such as RSSI and RF phase, may not accurately indicate the tag-to-antenna distance. More im-

portantly, RFID localization accuracy can be significantly affected by frequency hopping, which has not been considered by previous works. In some regions [5], an RFID reader is required to hop the operating frequency from one to the next at a predefined rate within a given region's frequency band, which is used to combat RFID co-channel or adjacent-channel interferences in the case that multiple RFID readers operate simultaneously at neighboring positions. Since the delay in propagating through RFID hardware components can not be negligible, different operating frequencies induce different hardware-related phase shifts, especially when a monostatic RFID reader antenna is used for both signal transmission and reception. To minimize the phase shift, many RFID systems [65, 21, 11, 10] apply a bistatic antenna configuration with transmitter-receiver isolation. However, the use of two antennas involves additional complexity and expense. Moreover, most of COTS RFID readers only support the monostatic configuration, which makes most of the bistatic antenna-based solution infeasible for real applications. Therefore, the localization accuracy of most of the existing RFID solutions is subject to multipath interference and frequency hopping, which may not be very reliable when dealing with false-positive reading, especially in multipath-prevalent environments.

As CV-based methods have been proved to be efficient for object detection, some researchers [107, 25, 56] propose CV-assisted RFID systems to improve localization reliability. Although CV-based algorithms [83, 40, 50, 42] can obtain entire trajectories of moving objects by analyzing camera sensor data without being affected by multipath interference, it is difficult (or for some cases, even impossible) to distinguish multiple trajectories and associate them with individual objects, especially when multiple objects having the exact same appearance.

In this work, we propose an RFID and CV hybrid system with COTS RFID devices and a 2D monocular camera, called *RF-Focus*, which can recognize and locate moving RFID-tagged objects within an ROI in multipath-prevalent environments in aid of CV techniques. Fig. 3.1 shows a typical system deployment in such conveyorized systems, which consists of a pair of tightly-spaced antennas and a 2D monocular camera mounted on a gantry. Our major goals are to 1) remove additional phase shifts caused by multipath interference and frequency-dependent hardware characteristics from the reported RF phase to obtain the correct tag-to-antenna indicator and 2) analyze camera sensor data to reduce tag position uncertainty caused by RF phase periodicity.

Realizing such a hybrid system for ROI RFID tag recognition and localization has three key challenges discussed as follows:

Challenge 1. The multipath interference is of large importance in determining the total received signal at the antenna receiver, which will significantly affect the reported RF phase to correctly indicate the tag-to-antenna distance. However, previous RSSI and RF phase models [27, 96, 100] expressed by a function of the tag-to-antenna distance either have not contained any terms to indicate the impact of multipath interference. Tadar [113], the first RFID work for device-free object tracking, models RFID signal propagation in line-of-sight to analyze the multipath interference. However, this work does not provide the estimate of the multipath

interference in RSSI or RF phase as a function of some possible impacts, making it much challenging to understand the multipath interference.

Challenge 2. The frequency-dependent hardware characteristics induce propagation delays in RFID hardware at different operating frequencies, which can result in additional phase shifts in the reported RF phase. In an ideal environment without any multipath interference, if the tag-to-antenna distance is accurately measured, we can rely on a standard RF phase-distance model to estimate the hardware-related phase shifts at different frequencies. However, since a real reader antenna's position is located within the antenna casing, the manually-measured distance between an RFID tag and the center of the antenna casing surface may more or less deviate from the ground truth. Accurately locating a stationary RFID tag in space using existing RFID localization systems is still challenging. Since more than one reader antenna is required in these systems, the final localization accuracy suffers from tag-antenna geometry [61, 106].

Challenge 3. Due to RF phase periodicity, there may exist multiple candidate positions of an RFID tag to be located within the ROI. To minimize the tag position uncertainty, a feasible solution is to rely on CV techniques to exhaustively search all possible trajectories consisting of a sequence of image candidate regions that are likely to contain the RFID-tagged object over video frames, and then associate them with RFID data for object recognition and localization. However, time complexity will exponentially increase with the number of video frames, which will challenge the system's real-time performance.

RF-Focus introduces three innovations to deal with the above challenges:

1) To deal with the first challenge, we start by modelling a channel transfer function in multipath-prevalent environments to characterize the impact of multipath interference on RSSI and RF phase, and then derive novel RSSI and RF phase-distance models with additional multipath interference terms compared to prior models. On this basis, we demonstrate that the changes in RSSI and RF phase due to multipath interference for two tightly-spaced antennas can be approximately equal, which can be used to combat multipath interference in the following RFID and CV fusion procedure.

2) To address the second challenge, we firstly conduct an experiment to show the impact of frequency hopping on RSSI and RF phase. Then a novel multipath-powered algorithm is designed to measure the tag-to-antenna distance only using one antenna and several RFID tags, which can provide higher accuracy than manual measurement. After that, the phase shift caused by hardware characteristics at each operating frequency is estimated by solving an optimization problem. In the RFID and CV fusion, we remove the hardware-related phase shift from raw RF phase.

3) To deal with the third challenge, we define a matching score by associating CV image candidate regions with the processed RF phase after removing the phase shifts due to multipath interference and hardware characteristics only based on three successive fusion samples. Then the hybrid system can achieve real-time ROI RFID tag recognition and localization by finding the maximum of the matching score.

RF-Focus, to our knowledge, is the first CV-assisted RFID system for ROI RFID tag recognition and localization without pre-capturing the entire trajectory of each moving object in CV. The technical contributions are summarized as follows:

1) We formulate the multipath interference in RSSI and RF phase as a function of signal propagation paths and object reflection characteristics, which has not been presented by previous works. And a dual-antenna solution is proposed to minimize the impact of multipath interference on RF phase.

2) We proposed a multipath interference-powered method to estimate the phase shifts caused by frequency-dependent hardware characteristics. Our study reveals that RSSI is still an available parameter to sense environmental changes with high resolution and sensitivity like commonly-used RF phase, which is generally ignored by prior device-free RFID systems.

3) By removing phase shifts due to multipath interference and hardware characteristics from raw RF phase, we fuse CV image data into the RFID system to minimize tag position uncertainty due to RF phase periodicity. To guarantee real-time performance, we propose a matching score over three successive fusion samples to estimate the likelihood of being an RFID tag in each image region.

We implement and evaluate our hybrid system with COTS RFID devices and a 2D monocular camera. The experiments demonstrate the effectiveness of multipath interference modeling and the hardware-related phase shift estimation. When simultaneously differentiating 5 moving RFID-tagged objects within the ROI, RF-Focus achieves average recognition accuracy of 91.67% and position matching accuracy of 94.26% given a false positive rate of 10%.

3.2 Dealing with Multipath Interference

In the section, we introduce how to characterize the changes in RSSI and RF phase due to multipath interference. Then we propose a dual-antenna solution to combat multipath interference, which is a fundamental procedure to clean the raw RF phase for RFID and CV fusion. The experiments to demonstrate the effectiveness of the proposed RSSI and RF phase-distance models in multipath LOS and NLOS are described in Section 3.5.2.

3.2.1 Multipath Signal Propagation in LOS

As shown in Fig. 3.2, when the transmitter A_T of a monostatic reader antenna A emits a continuous wave (CW) signal to a multipath-prevalent environment, there are generally three categories of reflected signals back to the antenna receiver A_R:

Backscatter signal. The transmitted signal that travels along the direct LOS path from A_T to an RFID tag T (i.e., $A_T \rightarrow T$) combines with other reflected signals off each nearby reflector W_i to T (i.e., $A_T \rightarrow W_i \rightarrow T$) at the tag end. Once T captures the sufficient power to produce a backscatter signal, the signal emitted from T can be directly delivered back to A_R (i.e., $T \rightarrow A_R$) and again reflected off W_i to A_R (i.e., $T \rightarrow W_i \rightarrow A_R$), respectively. In particular, the signal reflection can significantly induce the loss in power. We ignore each signal reflected off W_i two

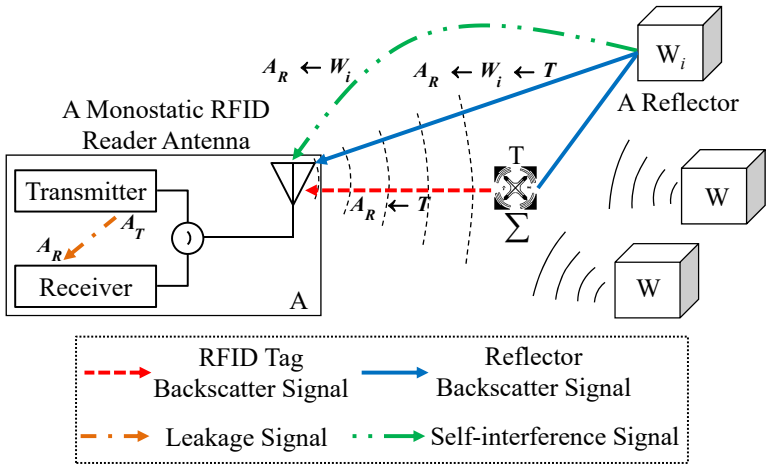


Figure 3.2 : Multipath signal propagation in LOS.

times (i.e., $A_T \rightarrow W_i \rightarrow T \rightarrow W_i \rightarrow A_R$) because it only contains much lower power than other received signals at A_R.

Leakage signal. In a monostatic antenna, a reader transmits a CW signal and listens for an RFID tag response simultaneously in full-duplex mode. The transmitted signal travels through the reader circuits and the antenna cable to the antenna and then back again. The additional signal (i.e., leakage signal) along the propagation path from A_T to A_R (i.e., $A_T \rightarrow A_R$) will add up to each received signal.

Self-interference signal. The transmitted signal arrives at W_i and then is directly reflected to A_R (i.e., $A_T \rightarrow W_i \rightarrow A_R$). Since the self-interference signal is the same as the original signal, its impact on received signals can be negligible. Let $s(t)$ be a CW signal emitted from A_T at time t, then the total received signal $r(t)$ at the receiver A_R is characterized as follows:

$$
\begin{aligned}
r(t) &= \underbrace{h_{A_T \rightarrow A_R}}_{A_T \rightarrow A_R} \left(\underbrace{h_{A_T \rightarrow T}}_{A_T \rightarrow T} + \underbrace{\sum_{i \geq 1} h_{W_i} h_{A_T \rightarrow W_i \rightarrow T}}_{A_T \rightarrow W \rightarrow T} \right) \left(\underbrace{h_T h_{T \rightarrow A_R}}_{T \rightarrow A_R} + \underbrace{h_T \sum_{i \geq 1} h_{W_i} h_{T \rightarrow W_i \rightarrow A_R}}_{T \rightarrow W \rightarrow A_R} \right) s(t) \\
&\approx \underbrace{h_{A_T \rightarrow A_R}}_{A_T \rightarrow A_R} \left[\underbrace{h_{A_T \rightarrow T} h_T h_{T \rightarrow A_R}}_{A_T \rightarrow T \rightarrow A_R} + \underbrace{\left(\sum_{i \geq 1} h_{W_i} h_{A_T \rightarrow W_i \rightarrow T} \right) h_T h_{T \rightarrow A_R}}_{A_T \rightarrow W \rightarrow T \rightarrow A_R} + \underbrace{h_{A_T \rightarrow T} h_T \left(\sum_{i \geq 1} h_{W_i} h_{T \rightarrow W_i \rightarrow A_R} \right)}_{A_T \rightarrow T \rightarrow W \rightarrow A_R} \right] s(t) \\
&= \underbrace{\mu_l e^{J\phi_l}}_{A_T \rightarrow A_R} \left[\underbrace{\mu_0 e^{J\phi_0} \mu_T e^{J\phi_T} \mu_0 e^{J\phi_0}}_{A_T \rightarrow T \rightarrow A_R} + 2 \underbrace{\left(\sum_{i \geq 1} \mu_{W_i} e^{J\phi_{W_i}} \mu_i e^{J\phi_i} \right) \mu_T e^{J\phi_T} \mu_0 e^{J\phi_0}}_{A_T \rightarrow W \rightarrow T \rightarrow A_R \;\&\; A_T \rightarrow T \rightarrow W \rightarrow A_R} \right] s(t) \\
&= \underbrace{\mu_l e^{J\phi_l}}_{A_T \rightarrow A_R} \left[\underbrace{\mu_T {\mu_0}^2 e^{J(2\phi_0 + \phi_T)}}_{A_T \rightarrow T \rightarrow A_R} + \underbrace{2 \mu_T \mu_0 e^{J(\phi_0 + \phi_T)} \sum_{i \geq 1} \mu_{W_i} \mu_i e^{J\left(\phi_i + \phi_{W_i}\right)}}_{A_T \rightarrow W \rightarrow T \rightarrow A_R \;\&\; A_T \rightarrow T \rightarrow W \rightarrow A_R} \right] s(t)
\end{aligned}
\tag{3.1}
$$

where $h = \mu e^{J\phi}$ represents the channel transfer function of each propagation path in a complex number, i.e., $h_{A_T \to T} = h_{T \to A_R} = \mu_0 e^{J\phi_0}$, $h_{A_T \to A_R} = \mu_l e^{J\phi_l}$, $h_T = \mu_T e^{J\phi_T}$ and $h_{W_i} = \mu_{W_i} e^{J\phi_{W_i}}$. The amplitude and phase parameters $\langle \mu_l, \phi_l \rangle$, $\langle \mu_{W_i}, \phi_{W_i} \rangle$ and $\langle \mu_T, \phi_T \rangle$ are determined by the characteristics of reader circuits, reflector material and RFID tag circuits, respectively. Here μ_l, μ_T and μ_{W_i} are also called the transmission gains to measure the loss in transmitted signal power. Specifically, for an RFID tag with a dipole antenna, the tag oriented along the electric field from the reader antenna may receive the maximum RF signal while the tag oriented perpendicular to the field may receive no RF signal at all, which can not be read. Hence, $\langle \mu_T, \phi_T \rangle$ are the angle-dependent parameters, varying with the orientation relative to reader antennas. Fortunately, a dual-dipole RFID tag (e.g., Impinj H47 RFID tag) can efficiently minimize the tag-to-antenna orientation sensitivity.

3.2.2 RSSI and RF Phase-distance Models in Multipath LOS

Further, let d_0 and $d_i (i \geq 1)$ be the distances of the direct path along $A_T \to T$ and the i-th reflected path along $A_T \to W_i \to T$ (or $T \to W_i \to A_R$), where $d_i \geq d_0$, and g_A be the antenna gain, so the amplitudes μ_0 and $\mu_i (i \geq 1)$ are represented by $\mu_0 = \frac{g_A}{d_0}$ and $\mu_i = \frac{g_A}{d_i}$ [72]. In addition, let c and f be the speed of light and the carrier wave frequency, so the phase shifts ϕ_0 and $\phi_i (i \geq 1)$ are calculated by $\phi_0 = \frac{2\pi f}{c} d_0$ and $\phi_i = \frac{2\pi f}{c} d_i$. According to Eq. 3.1, the total channel transfer function H_{LOS} of the received signal $r(t)$ in LOS can be written as

$$H_{LOS} = \mu_l \mu_T \left(\frac{g_A}{d_0} \right)^2 e^{J\left(\frac{4\pi f}{c} d_0 + \phi_l + \phi_T\right)} \left[1 + 2 \sum_{i \geq 1} \mu_{W_i} \frac{d_0}{d_i} e^{J\left(\frac{2\pi f}{c}(d_i - d_0) + \phi_{W_i}\right)} \right] \tag{3.2}$$

RSSI. The modulus of H_{LOS} (i.e., received signal amplitude (or voltage) V_{total} in the unit of Volts) is

$$V_{total} = \|H_{LOS}\| = \mu_l \mu_T \left(\frac{g_A}{d_0} \right)^2 \gamma_m \tag{3.3}$$

where $\gamma_m = \left\| 1 + 2 \sum_{i \geq 1} \mu_{W_i} \frac{d_0}{d_i} e^{J\left[\frac{2\pi f}{c}(d_i - d_0) + \phi_{W_i}\right]} \right\|$ is the multipath term of RSSI. For simplicity, γ_m is called multipath RSSI in the rest of the work. The range of μ_l,μ_T and μ_{W_i} is $[0, 1]$.

The received signal power P_{total} in the unit of Watts is proportional to the square of the voltage, which can be represented by $P_{total} = \eta \cdot (V_{total})^2$, where η is a constant. Since the RFID reader can report RSSI $rssi$ to indicate backscatter signal power in dBm rather than in Watts, we make the unit conversion of dBm to Watts using $Watts = 10^{\left(\frac{dBm}{10} - 3\right)}$. Thus, the RSSI-distance model in multipath LOS environments is

$$\mu_l \mu_T \left(\frac{g_A}{d_0} \right)^2 \gamma_m = \frac{1}{\sqrt{\eta}} 10^{\left(\frac{rssi}{20} - 1.5\right)} \tag{3.4}$$

RF Phase. The argument of H_{LOS} (i.e., phase shift ϕ_{total} in unit of radians) is

$$\begin{aligned}\phi_{total} &= Arg\left(H_{LOS}\right) + 2\pi k_{total} \\ &= Arg\left(\frac{4\pi f}{c}d_0 + \phi_l + \phi_T\right) + \phi_m + 2\pi k_{total}\end{aligned} \tag{3.5}$$

where $\phi_m = Arg\left(1 + 2\sum_{i\geq 1}\mu_{W_i}\frac{d_0}{d_i}e^{J\left[\frac{2\pi f}{c}(d_i - d_0) + \phi_{W_i}\right]}\right)$ is the multipath term of RF phase, i.e., multipath RF phase. The range of the function $Arg\left(\cdot\right)$ is in the closed-open interval $(-\pi, \pi]$, so we have $\phi_m \in (-\pi, \pi]$. k_{total} is an integer to make $Arg\left(H_{LOS}\right) \in (-\pi, \pi]$. Suppose that τ_l and τ_T are the signal propagation delays in the RFID reader and RFID tag, then the phase shifts ϕ_l and ϕ_T are denoted by $\phi_l = 2\pi f\tau_l$ and $\phi_T = 2\pi f\tau_T$, respectively.

Many COTS RFID readers (e.g., Impinj R420 reader used in our experiment) can report RF phase ϕ_{report}, which rotates by 2π for every one-half carrier wavelength. In the reader-to-tag forward link, the RFID reader generally adopts phase-reversal amplitude shift keying (PR-ASK) technique to modulate the RF signal in dense interrogator mode, so the reported RF phase of the received signal is opposite to the true RF phase. In addition, there are two tag backscatter modulation techniques in the tag-to-reader reverse link, i.e., amplitude shift keying (ASK) and phase shift keying (PSK). *1) ASK.* Most of the previous RFID systems work at 'MaxThroughput' reader mode with the highest data rate for RFID reading. However, this mode is significantly susceptible to noise interference and can only be applied in the scenario with a few RFID tags. *2) PSK.* According to [45], the 'AutosetDenseReader' reader mode is recommended to be configured in most scenarios, which can automatically select a reader mode to adapt for the application scenario. However, the output RF phase may have π radians of phase ambiguity due to phase reversals in PSK, meaning that the reported RF phase may be the true RF phase or the true RF phase plus π radians [43]. In general, this automatic reader mode is generally implemented. To deal with phase reversal in the received signal and the π jump, the RF phase ϕ for use is formulated as

$$\phi = \mathrm{mod}\left(-\phi_{report}, \pi\right) \tag{3.6}$$

In the following, the reported RF phase measurement for use is the version after this transformation by default. Thus, the RF phase-distance model in multipath LOS environments is

$$\phi + \pi k = \frac{4\pi f}{c}d_0 + \phi_l + \phi_T + \phi_m \tag{3.7}$$

where the unknown parameter πk, called integer phase ambiguity, is an integer multiple of π.

3.2.3 RSSI and RF Phase-distance Models in NLOS

In NLOS cases that the straight-line path between A and T is obstructed by an obstacle W_X but RF signals can still pass through W_X to read T, the channel transfer function of each backscatter signal path needs to additionally multiply

$h_{W_X} = \mu_{W_X} e^{J\phi_{W_X}}$. Thus, the total channel transfer function in NLOS is expressed as $H_{NLOS} = \mu_{W_X}^2 e^{J\left(2\phi_{W_X}\right)} H_{LOS}$. The RSSI and RF phase-distance models in NLOS are

$$\begin{cases} \dfrac{1}{\sqrt{\eta}} 10^{\left(\frac{rssi}{20} - 1.5\right)} = \mu_l \mu_T \mu_{W_X}^2 \left(\dfrac{g_A}{d_0}\right)^2 \gamma_m \\ \phi + \pi k = \dfrac{4\pi f}{c} d_0 + \phi_l + \phi_T + 2\phi_{W_X} + \phi_m \end{cases} \tag{3.8}$$

In particular, we need to indicate that these NLOS models can be used for device-free sensing systems. For example, in contrast to relative positioning in Tadar [113], it provides an opportunity to achieve absolute device-free target localization, which we will leave for our future work.

3.2.4 The Shortcomings of Prior RSSI and RF Phase-distance Models

Here we present the commonly-used RSSI and RF phase-distance models in most of existing works [110, 27, 79, 105], i.e.,

$$\begin{cases} \dfrac{1}{\sqrt{\eta}} 10^{\left(\frac{rssi}{20} - 1.5\right)} = \mu_l \mu_T \left(\dfrac{g_A}{d_0}\right)^2 \\ \phi_{report} + 2\pi k = \dfrac{4\pi f}{c} d_0 + \phi_l + \phi_T \end{cases} \tag{3.9}$$

The shortcomings are listed as follows:

1) The changes in RSSI and RF phase due to multipath interference are ignored.

2) The reported RF phase ϕ_{report} can not be directly applied to estimate the tag-to-antenna distance. Prior position-based RFID sensing works can perform well because they only capture the relative change in RF phase to indicate the changes in tag positions.

3) The reader can only work at "MaxThroughput" reader mode in an application scenario with few RFID tags. And in this mode, the reported RSSI and RF phase measurements are more susceptible to multipath interference than other reader modes.

4) The phase shifts ϕ_l and ϕ_T are frequency-dependent parameters, which can not be easily removed under frequency hopping.

3.2.5 Multipath Interference in Two Tightly-spaced Antennas

Dual-antenna Hypothesis: Given two tightly-spaced reader antennas relatively far away from an RFID tag, multipath interference may induce similar changes in multipath RSSI and RF phase corresponding to each antenna.

As shown in Fig. 3.3, given two tightly-spaced reader antennas A_1 and A_2 with the spacing of D, $\theta_0(A)$ and $\theta_i(A)$, $i \geq 1$, represent the angle of arrival (AOA) of the direct path along $A \to T$ and the i-th reflected path along $A \to W_i \to T$, respectively. Since the reflectors besides the RFID tag T may produce dominant

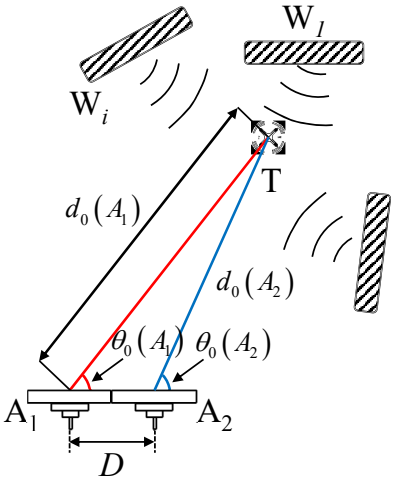


Figure 3.3 : Angle of arrival of two tightly-spaced antennas.

multipath interference relative to those far away from T, we can assume that $\theta_i(A) \approx \theta_0(A)$ for these nearby reflectors. And the path distance difference between A_1 and A_2 is calculated by $d_i(A_1) - d_i(A_2) \approx D\cos\theta_i(A)$, $i \geq 0$. Thus, we have

$$\begin{aligned} d_i(A_1) - d_0(A_1) &= d_i(A_2) - d_0(A_2) + D\cos\theta_i(A) - D\cos\theta_0(A) \\ &\approx d_i(A_2) - d_0(A_2) \end{aligned} \tag{3.10}$$

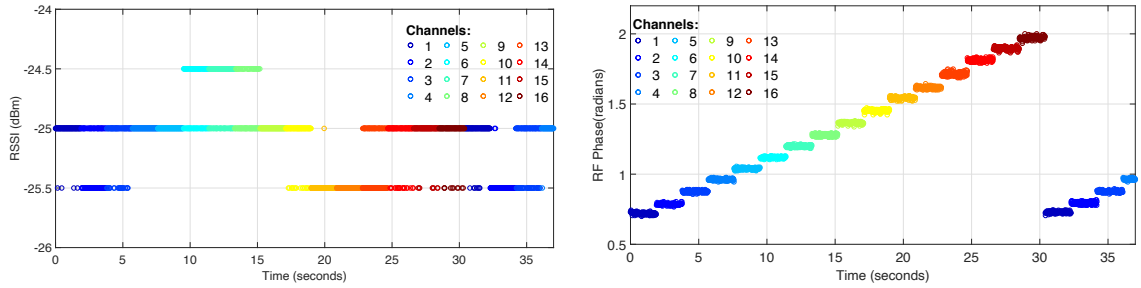
And together with $d_i(A_2) \gg D\cos\theta_i(A)$, the ratio of $d_0(A_1)$ to $d_i(A_1)$ is approximated as

$$\begin{aligned} \frac{d_0(A_1)}{d_i(A_1)} &\approx \frac{d_0(A_2) + D\cos\theta_0(A)}{d_i(A_2) + D\cos\theta_i(A)} \\ &\approx \frac{d_0(A_2)}{d_i(A_2)} \end{aligned} \tag{3.11}$$

According to Eq. 3.2 in multipath LOS, we get

$$\sum_{i\geq 1} \mu_{W_i} \frac{d_0(A_1)}{d_i(A_1)} e^{J\left(\frac{2\pi f}{c}[d_i(A_1) - d_0(A_1)] + \phi_{W_i}\right)} \approx \sum_{i\geq 1} \mu_{W_i} \frac{d_0(A_2)}{d_i(A_2)} e^{J\left(\frac{2\pi f}{c}[d_i(A_2) - d_0(A_2)] + \phi_{W_i}\right)} \tag{3.12}$$

Therefore, we can infer that $\gamma_m(A_1) \approx \gamma_m(A_2)$ and $\phi_m(A_1) \approx \phi_m(A_2)$. This case can be better complied as the tag-to-antenna distance increases and the two antennas are more tightly spaced. In the processing of RFID and CV fusion, we depend on this dual-antenna hypothesis to minimize the effect of multipath interference. Note that in NLOS scenarios, since an obstacle with uncontrolled materials between an RFID tag and reader antennas can attenuate the received signal power at the tag end, the reported RSSI and RF phase is more susceptible to multipath interference compared to LOS cases, which may challenge the proposed hypothesis. In this work, we mainly focus on the LOS scenarios. The solutions to deal with NLOS cases are discussed in Section 3.6.



(a) RSSI. (b) RF Phase.

Figure 3.4 : The impact of frequency-hopping on RSSI and RF phase.

3.3 Dealing with Frequency-dependent Hardware Characteristics

In this section, we firstly conduct an experiment to illustrate the impact of frequency hopping on RSSI and RF phase. Then we extend the multipath models in Section 3.2.1 to the case with a dominant multipath signal for hardware-related phase shift estimation, which is another fundamental procedure for RFID and CV fusion. The algorithm performs with only one reader antenna and several RFID tags. The experiments to demonstrate the phase-shift estimation performance are described in Section 3.5.3.

3.3.1 Preliminary

Here we take an experiment to show the impact of different operating frequencies on RSSI and RF phase measurements. Without multipath interference, the RSSI and RF phase over distance are shown in Eq. 3.9. An RFID reader operates in a fixed transmitted power and in a sequential sequence over 16 frequency channels (i.e., 920.625∼924.375 MHz with 250kHz frequency spacing in China) to read a stationary RFID tag in free space with very low multipath noise. The tag-to-antenna manually-measured distance is 35cm between the antenna casing surface center and the RFID tag center. As shown in Fig. 3.4, each color of samples corresponds to a frequency channel and the dwell time per channel is about 2 seconds. Note that in United States or some other countries, an RFID reader will randomly hop to an available channel every about 0.2 seconds [44]. We also find that RSSI samples change a little while there exist many piecewise RF phase samples. The main reasons are explained as follows:

RSSI. The work [65] indicates that the tag reflection capability (i.e., corresponding to the parameter μ_T in our work) is related to the operating frequency of the signal for tag modulation. However, since most COTS RFID readers work within a very narrow frequency band, the parameter μ_T will almost keep unchanged. In addition, μ_l and g_A are frequency-independent parameters. Therefore, RSSI measurements do not change with different frequencies.

RF Phase. Due to unchanged prorogation delays τ_l and τ_T, we take the difference between two RF phase measurements $\phi(f_i)$ and $\phi(f_j)$ at operating frequencies

f_i and f_j, i.e.,

$$\phi(f_i) - \phi(f_j) + \Delta k\pi = \frac{4\pi d_0}{c}(f_i - f_j) + 2\pi(\tau_l + \tau_T)(f_i - f_j) \tag{3.13}$$

Ignoring the phase ambiguity $\Delta k\pi$ (i.e., $\Delta k\pi = 0$) and substituting $d_0 \approx 35cm$ into Equation (3.13), the hardware-related phase-shift difference between Channel 1 and Channel 16, i.e., the value of $2\pi(\tau_l + \tau_T)(f_{16} - f_1)$, is about 1.2 radians. In the FCC regions with the band of 902.75∼927.25MHz, the phase difference will become larger. Unfortunately, the manually-measured distance d_0 is inaccurate because the antenna phase center may be within the plastic casing plane. With even 1cm measurement error, the estimate of the hardware-related phase shift at Channel 1 will introduce additional 0.385 radians phase-shift error. In the following, we will introduce how to leverage the proposed multipath RSSI and RF phase models to accurately estimate the tag-to-antenna distance.

3.3.2 Multipath Interference-powered Phase-shift Estimation

To intuitively introduce our method, we deploy an experiment setup in multipath LOS shown in Fig. 3.5. Two RFID tags T_1 and T_2 are deployed very close to a reader antenna A in 3D space. To combat the coupling effect [105] (i.e., the change in RF phase and RSSI caused by neighboring RFID tags), the distance between any two RFID tags requires greater than half a wavelength (i.e., about 16 cm). A reflector W_1 with a strong reflection coefficient is placed in front of these tags at different positions X each time. Other environmental multipath sources such as the floor, a distant wall and furniture are relatively far from the tag. In this case, the backscatter signals reflected off these sources with long propagation distances and random propagation delays may cancel out each other at the receiver (see Fig. 3.6) such that the reflected signal off W_1 can dominate the multipath signal, i.e.,

$$\sum_{i\geq 1} \mu_{W_i} \frac{g_A}{d_i} e^{J\left(\frac{2\pi f}{c} d_i + \phi_{W_i}\right)} \approx \mu_{W_1} \frac{g_A}{d_1} e^{J\left(\frac{2\pi f}{c} d_1 + \phi_{W_1}\right)} \tag{3.14}$$

Then the total channel transfer function H_{LOS} of the received signal $r(t)$ can be rewritten as

$$H_{LOS} \approx \mu_l \mu_T \left(\frac{g_A}{d_0}\right)^2 e^{J\left(\frac{4\pi f}{c} d_0 + \phi_l + \phi_T\right)} \left[1 + 2\mu_{W_1} \frac{d_0}{d_1} e^{J\left(\frac{2\pi f}{c}(d_1 - d_0) + \phi_{W_1}\right)}\right] \tag{3.15}$$

Accordingly, the multipath RSSI and RF phase are formulated as

$$\begin{cases} \gamma_m = \sqrt{1 + 2x_m z_m + (z_m)^2} \\ \phi_m = 2\arctan \dfrac{y_m z_m}{\sqrt{(1 + x_m z_m)^2 + (y_m z_m)^2} + 1 + x_m z_m} \end{cases} \tag{3.16}$$

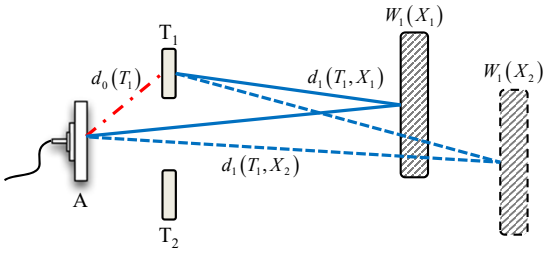


Figure 3.5 : Experiment setup for tag-to-antenna estimation.

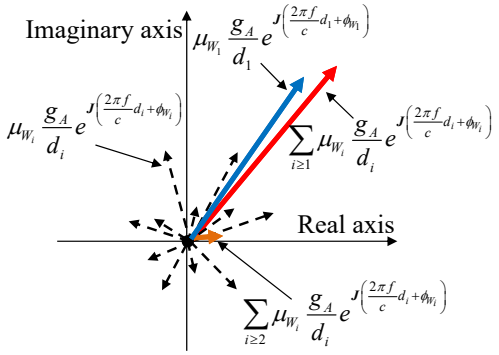


Figure 3.6 : The sum of reflected signals in the case with a dominant multipath interference.

where

$$\begin{cases} x_m = \cos\left[\dfrac{2\pi f}{c}\left(d_1 - d_0\right) + \phi_{W_1}\right] \\ y_m = \sin\left[\dfrac{2\pi f}{c}\left(d_1 - d_0\right) + \phi_{W_1}\right] \\ z_m = 2\mu_{W_1}\dfrac{d_0}{d_1} \end{cases} \tag{3.17}$$

Further, if the tag-to-antenna distance d_0 be far less than the dominant reflected path length d_1 along $A \to W_1 \to T$ (or $A \to T \to W_1$), i.e., $d_0 \ll d_1$, we have $2\mu_{W_1}\frac{d_0}{d_1} \ll 1$ due to the reflection coefficient $\mu_{W_1} \in (0,1)$. In this case, two approximate expressions $(1+x)^{\frac{1}{2}} \approx 1 + \frac{x}{2}$ (if $|x| \ll 1$) and $\arctan x \approx x$ (if $x \to 0$) can be used to rewrite Eq. 3.16, i.e.,

$$\begin{cases} \gamma_m = 1 + 2\mu_{W_1}\dfrac{d_0}{d_1}\cos\left[\dfrac{2\pi f}{c}\left(d_1 - d_0\right) + \phi_{W_1}\right] \\ \phi_m = 2\mu_{W_1}\dfrac{d_0}{d_1}\sin\left[\dfrac{2\pi f}{c}\left(d_1 - d_0\right) + \phi_{W_1}\right] \end{cases} \tag{3.18}$$

3.3.3 Deriving Multipath RSSI and RF Phase Measurements

In absence and presence of W_1, the RSSI and RF phase measurements at the same frequency are $\langle rssi, \phi \rangle$ and $\langle rssi\,(X)\,, \phi\,(X) \rangle$, respectively. Canceling unknown parameters (i.e., η, ϕ_l, ϕ_T, μ_l, μ_T and g_A) based on Eq. 3.4 and Eq. 3.7, the multipath RF phase and RSSI can be derived as follows:

$$\begin{cases} \gamma_m = 10^{\frac{rssi(X) - rssi}{20}} \\ \phi_m = \phi\,(X) - \phi + \pi k_m \end{cases} \tag{3.19}$$

where k_m is an integer to make ϕ_m within $(-\pi, \pi]$.

3.3.4 Calculating Dominant Reflected Path Distance

For the reflector W_1 at the positions X_1 and X_2, the multipath measurements of an RFID tag T_1 are denoted as $\langle \phi_m(T_1, X_1), \gamma_m(T_1, X_1) \rangle$ and $\langle \phi_m(T_1, X_2), \gamma_m(T_1, X_2) \rangle$, respectively. The parameters μ_{W_1} and ϕ_{W_1} are independent to the reflector positions. Substituting them into Eq. 3.18, we can obtain

$$\begin{cases} 2\mu_{W_1} \dfrac{d_0(T_1)}{d_1(T_1, X_1)} = \sqrt{[\gamma_m(T_1, X_1) - 1]^2 + [\phi_m(T_1, X_1)]^2} \\ 2\mu_{W_1} \dfrac{d_0(T_1)}{d_1(T_1, X_2)} = \sqrt{[\gamma_m(T_1, X_2) - 1]^2 + [\phi_m(T_1, X_2)]^2} \end{cases} \tag{3.20}$$

and

$$\begin{cases} \dfrac{2\pi f}{c} [d_1(T_1, X_1) - d_0(T_1)] + \phi_{W_1} + \pi k(T_1, X_1) = \arctan\left[\dfrac{\phi_m(T_1, X_1)}{\gamma_m(T_1, X_1) - 1}\right] \\ \dfrac{2\pi f}{c} [d_1(T_1, X_2) - d_0(T_1)] + \phi_{W_1} + \pi k(T_1, X_2) = \arctan\left[\dfrac{\phi_m(T_1, X_2)}{\gamma_m(T_1, X_2) - 1}\right] \end{cases} \tag{3.21}$$

where $k(T_1, X_1)$ and $k(T_1, X_2)$ are unknown integers. Combing with Eq. 3.20 and Eq. 3.21, we can calculate the dominant reflected path distances $d_1(T_1, X_1)$ and $d_1(T_1, X_2)$ using the following equations,

$$\begin{cases} \dfrac{d_1(T_1, X_2)}{d_1(T_1, X_1)} = \alpha(T_1, X_1, X_2) \\ \dfrac{2\pi f}{c} [d_1(T_1, X_1) - d_1(T_1, X_2)] + \pi \Delta k(T_1, X_1, X_2) = \beta(T_1, X_1, X_2) \end{cases} \tag{3.22}$$

where

$$\begin{cases} \alpha(T_1, X_1, X_2) = \dfrac{\sqrt{[\gamma_m(T_1, X_1) - 1]^2 + [\phi_m(T_1, X_1)]^2}}{\sqrt{[\gamma_m(T_1, X_2) - 1]^2 + [\phi_m(T_1, X_2)]^2}} \\ \beta(T_1, X_1, X_2) = \arctan\left[\dfrac{\phi_m(T_1, X_1)}{\gamma_m(T_1, X_1) - 1}\right] - \arctan\left[\dfrac{\phi_m(T_1, X_2)}{\gamma_m(T_1, X_2) - 1}\right] \\ \Delta k(T_1, X_1, X_2) = k(T_1, X_1) - k(T_1, X_2) \end{cases} \tag{3.23}$$

In the following, we give details on how to determine the value of $\Delta k(T_1, X_1, X_2)$. Suppose that $d_{A \to W_1}(X)$ is the distance between A and W_1, and $d_{W_1 \to T_1}(X)$ denote the distance between W_1 and T_1. The distances $d_1(T_1, X_1)$ and $d_1(T_1, X_2)$ can be expressed as

$$\begin{cases} d_1(T_1, X_1) = d_{A \to W_1}(X_1) + d_{W_1 \to T_1}(X_1) \\ d_1(T_1, X_2) = d_{A \to W_1}(X_2) + d_{W_1 \to T_1}(X_2) \end{cases} \tag{3.24}$$

Let $D(X_1, X_2)$ be the manually-measured distance between the reflector positions X_1 and X_2. Based on the triangle inequality theorem $|a + b| \leq |a| + |b|$ and

the triangle rule that the length difference of two sides is smaller than the length of the third side, we can obtain the following inequation,

$$\begin{aligned}|\Delta d_1(T_1,X_1,X_2)| &= |d_1(T_1,X_1)-d_1(T_1,X_2)| \\ &= |d_{A\to W_1}(X_1)-d_{A\to W_1}(X_2)+d_{W_1\to T_1}(X_1)-d_{W_1\to T_1}(X_2)| \\ &\le |d_{A\to W_1}(X_1)-d_{A\to W_1}(X_2)|+|d_{W_1\to T_1}(X_1)-d_{W_1\to T_1}(X_2)| \\ &< 2D(X_1,X_2)\end{aligned} \tag{3.25}$$

According to Eq. 3.18, we have

$$|\beta(T_1,X_1,X_2)-\pi\Delta k(T_1,X_1,X_2)| < \frac{4\pi f}{c}D(X_1,X_2) \tag{3.26}$$

Thus, $\Delta k(T_1,X_1,X_2)$ can be determined as follows:

$$\Delta k(T_1,X_1,X_2)=\begin{cases}-\kappa, & \beta(T_1,X_1,X_2)\in[0,\pi),\Delta d_1(T_1,X_1,X_2)\ge 0\\ -(\kappa+1), & \beta(T_1,X_1,X_2)\in[-\pi,0),\Delta d_1(T_1,X_1,X_2)\ge 0\\ \kappa+1, & \beta(T_1,X_1,X_2)\in(0,\pi],\Delta d_1(T_1,X_1,X_2)\le 0\\ \kappa, & \beta(T_1,X_1,X_2)\in(-\pi,0],\Delta d_1(T_1,X_1,X_2)\le 0\end{cases} \tag{3.27}$$

where κ is an integer, which is calculated by

$$\forall\kappa\in\left[0,\frac{4f}{c}D(X_1,X_2)\right):\ \ \frac{4\pi f}{c}|\Delta d_1(T_1,X_1,X_2)|\in[\kappa\pi,(\kappa+1)\pi] \tag{3.28}$$

Similarly, we can also depend on above steps to calculate dominant reflected path distances of T_2, i.e., $d_1(T_2,X_1)$ and $d_1(T_2,X_2)$.

3.3.5 Calculating Tag-to-antenna Distance

When the reflector W_1 locates at X_1, the multipath measurements of T_2 are denoted as $\langle\phi_m(T_2,X_1),\gamma_m(T_2,X_1)\rangle$. Then we can calculate the tag-to-antenna distances $d_0(T_1)$ and $d_0(T_2)$ by

$$\begin{cases}\dfrac{d_0(T_1)}{d_0(T_2)}=\dfrac{d_1(T_1,X_1)}{d_1(T_2,X_1)}\alpha(T_1,T_2,X_1)\\ \dfrac{2\pi f}{c}[d_0(T_2)-d_0(T_1)+\Delta d_1(T_1,T_2,X_1)]+\pi\Delta k(T_1,T_2,X_1)=\beta(T_1,T_2,X_1)\end{cases} \tag{3.29}$$

where

$$\begin{cases}\alpha(T_1,T_2,X_1)=\dfrac{\sqrt{[\gamma_m(T_1,X_1)-1]^2+[\phi_m(T_1,X_1)]^2}}{\sqrt{[\gamma_m(T_2,X_1)-1]^2+[\phi_m(T_2,X_1)]^2}}\\ \beta(T_1,T_2,X_1)=\arctan\left[\dfrac{\phi_m(T_1,X_1)}{\gamma_m(T_1,X_1)-1}\right]-\arctan\left[\dfrac{\phi_m(T_2,X_1)}{\gamma_m(T_2,X_1)-1}\right]\\ \Delta d_1(T_1,T_2,X_1)=d_1(T_1,X_1)-d_1(T_2,X_1)\\ \Delta k(T_1,T_2,X_1)=k(T_1,X_1)-k(T_2,X_1)\end{cases} \tag{3.30}$$

Here we introduce how to determine the value of $\Delta k\left(T_1, T_2, X_1\right)$. Let $D\left(T_1, T_2\right)$ be the manually-measured distance between the tags T_1 and T_2, then we can construct the following inequation to determine $\Delta k\left(T_1, T_2, X_1\right)$,

$$\left|\beta\left(T_1, T_2, X_1\right) - \frac{2\pi f}{c}\Delta d_1\left(T_1, T_2, X_1\right) - \pi\Delta k\left(T_1, T_2, X_1\right)\right| < \frac{2\pi f}{c}D\left(T_1, T_2\right) \quad (3.31)$$

And $\Delta k\left(T_1, T_2, X_1\right)$ can be determined as follows:

$$\Delta k\left(T_1, T_2, X_1\right) = \begin{cases} -\kappa, & \beta\left(T_1, T_2, X_1\right) - \frac{2\pi f}{c}\Delta d_1\left(T_1, T_2, X_1\right) \in [0, \pi), \Delta d_0\left(T_1, T_2\right) \geq 0 \\ -(\kappa+1), & \beta\left(T_1, T_2, X_1\right) - \frac{2\pi f}{c}\Delta d_1\left(T_1, T_2, X_1\right) \in [-\pi, 0), \Delta d_0\left(T_1, T_2\right) \geq 0 \\ \kappa+1, & \beta\left(T_1, T_2, X_1\right) - \frac{2\pi f}{c}\Delta d_1\left(T_1, T_2, X_1\right) \in (0, \pi], \Delta d_0\left(T_1, T_2\right) \leq 0 \\ \kappa, & \beta\left(T_1, T_2, X_1\right) - \frac{2\pi f}{c}\Delta d_1\left(T_1, T_2, X_1\right) \in (-\pi, 0], \Delta d_0\left(T_1, T_2\right) \leq 0 \end{cases} \quad (3.32)$$

where $\Delta d_0\left(T_1, T_2\right) = d_0\left(T_1\right) - d_0\left(T_2\right)$. κ is an integer, which is calculated by

$$\forall\kappa \in \left[0, \frac{2f}{c}D\left(T_1, T_2\right)\right) : \quad \frac{2\pi f}{c}\left|\Delta d_0\left(T_1, T_2\right)\right| \in [\kappa\pi, (\kappa+1)\pi] \quad (3.33)$$

3.3.6 Refining Hardware-related Phase Shifts

When we collect RSSI and RF phase measurements over all available channels at different reflector positions, each RFID tag will have multiple tag-to-antenna distance estimates. Suppose that $d_0\left(T_i, f_j, X_k\right)$ denotes the tag-to-antenna distance estimate of the tag T_i at the frequency f_j and the reflector position X_k, and $\phi\left(T_i, f_j\right)$ represents the RF phase measurement at f_j in absence of the reflector. The optimal hardware-related phase shift at f_j can be estimated by solving the following maximization problem,

$$\phi_h{}^*\left(f_j\right) = \underset{\phi_h(f_j)\in[0,\pi]}{\arg\max} \sum_{i,j,k} \left|\cos\Delta\phi\left(T_i, f_j, X_k\right)\right| \quad (3.34)$$

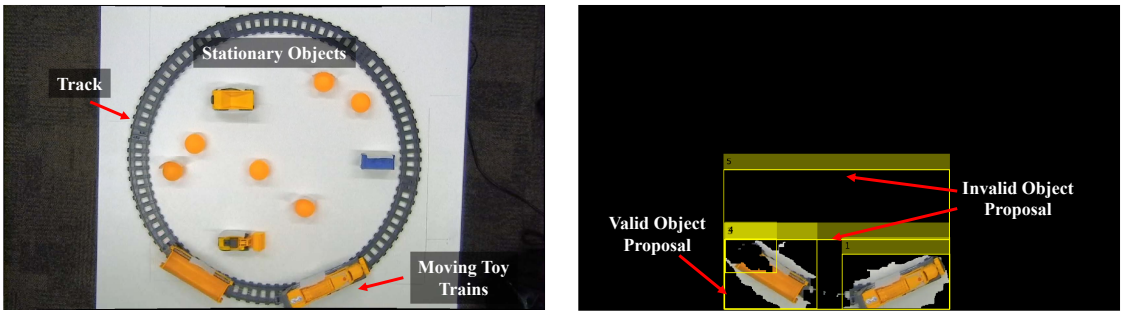
where

$$\begin{cases} \phi_h\left(f_j\right) = \mathrm{mod}\left[\phi_l\left(f_j\right) + \phi_T\left(f_j\right), \pi\right] \\ \Delta\phi\left(T_i, f_j, X_k\right) = \phi\left(T_i, f_j\right) - \frac{4\pi f_j}{c}d_0\left(T_i, f_j, X_k\right) - \phi_h\left(f_j\right) \end{cases} \quad (3.35)$$

And the function $\left|\cos\left(\cdot\right)\right|$ is to remove RF phase periodicity, so the phase-shift estimate is the true phase-shift modulo π. Finally, the hardware-related phase shifts at all available channels can be estimated. Note that since the ground-truth antenna position (i.e., antenna phase center) is typically a point on the centerline of the antenna [35], we deploy an RFID tag well-aligned with the antenna casing surface center to estimate this point by substituting the estimated hardware-related phase shift.

3.4 Fusing RFID and CV for Region-of-interest RFID Tag Recognition and Localization

So far, we have introduced how to deal with multipath interference and frequency-dependent hardware characteristics. In this section, we give details on RFID and CV fusion for ROI RFID tag recognition and localization by cleaning raw RF phase measurements.



(a) A video frame. (b) Moving objects.

Figure 3.7 : Extract object proposals of moving toy trains.

3.4.1 Candidate Region Extraction from 2D Image

Recall that the RF phase for use is a periodic function with the period π radians, which repeats at every quarter of the wavelength of the tag-to-antenna distance. In this case, RFID localization accuracy is subject to RFID tag position uncertainty coming from RF phase periodicity. Instead of recognizing and locating stationary RFID-tagged objects in the ROI, we mainly focus on moving objects in our conveyorized scenario of this work.

Here we extract the candidate region of each moving object in the field of camera view to minimize the RFID tag position ambiguity. At first, a Gaussian Mixture Models (GMMs)-based foreground detector [83] is exploited to detect the absence or presence of moving objects and segment the foreground mask of the moving objects from a video frame. Then we map the foreground mask into the video frame to ignore the background image information. Finally, we adopt purely image-based Edged Boxes [119] to extract the candidate regions highlighted by enclosing a series of bounding boxes (called object proposals), without pre-training any object detector like CNN-based algorithms [70]. Apart from providing faster computation (about 0.2 seconds per frame) and higher detection accuracy than other image-based methods like Selective Search [88], Edged Boxes also offers an additional objectness score for each object proposal to indicate the likelihood of an object being present. Hence, we choose those object proposals with high scores to reduce the effect of invalid object proposals containing more than one object. For example, Fig. 3.7a shows that a video frame in which many stationary objects are deployed within the track and two moving toy trains are moving on the track. Fig. 3.7b shows that Edge Boxes algorithm can effectively extract multiple candidate regions (i.e., object proposals) from the image that only contains the moving foreground.

3.4.2 2D Image to 3D World Coordinate Transformation

To associate image information with RFID reports, a fundamental procedure is to perform the perspective transformation from a 2D image coordinate system to a 3D world coordinate system, which can be achieved by Zhang's camera calibration algorithm [116]. The main steps are summarized as follows:

Prepare calibration images. A typical camera calibration pattern is an asym-

metric checkerboard, where one side contains an even number of black and white squares, and the other contains an odd number of squares. The checkerboard is affixed on a flat surface. Our fixed 2D monocular camera captures the checkerboard images from different orientations and distances.

Estimate camera parameters. The checkerboard corners can be automatically detected in each image to estimate intrinsic and extrinsic camera parameters and lens distortion coefficients.

Transform 2D image coordinates to 3D world coordinates. In the camera filed, we place a checkerboard in the same plane as RFID-tagged objects to be tracked. Then we capture the checkerboard image to compute the rotation matrix and the translation matrix of the camera relative to the checkerboard. On this basis, each pixel in the 2D ROI image can be transformed into the 3D world coordinate in the camera-centered coordinate system. In particular, we specify the centroid of each object proposal as its location in the 2D ROI image, so the corresponding 3D world coordinate can also be obtained.

3.4.3 Reader Antenna Localization in Camera-centered Coordinate System

Due to an unknown camera view angle relative to each reader antenna, it is unable to directly measure the antenna position relative to the camera lens center in 3D space as its 3D world point in the camera-centered coordinate system. As mentioned above, the coordinate of each corner in the checkerboard with the same plane as tracked objects can be calculated. And the distance of each corner to the casing surface center of each reader antenna can be manually measured with a tapeline or laser range-finder. Let $(x_{C_i}, y_{C_i}, z_{C_i})$ be the known 3D world coordinate of the i-th corner C_i in the camera-centered coordinate system, $d(C_i, A_j)$ be the manually-measured distance between C_i and A_j, and $\left(x_{A_j}, y_{A_j}, z_{A_j}\right)$ be the 3D world coordinate of the antenna A_j to be calculated. Thus, we build a sphere equation with the center at C_i and the radius $d(C_i, A_j)$ as follows:

$$\left(x_{C_i} - x_{A_j}\right)^2 + \left(y_{C_i} - y_{A_j}\right)^2 + \left(z_{C_i} - z_{A_j}\right)^2 = \left[d(C_i, A_j)\right]^2 \tag{3.36}$$

Note that since $d(C_i, A_j)$ is much larger than the distance difference between the ground-truth and the casing surface center, the impact of antenna position measurement error can be ignored. In this case, the closet point to all spheres corresponding to different corner coordinates is regarded as the position of A_j, which can be solved by least-squares minimization [28].

3.4.4 RFID and CV Sampling Synchronization

In our experiment, an RFID reader and a 2D camera both connect to a router with Network Time Protocol (NTP) server, which can provide a common time baseline for RFID and CV systems, respectively. The camera frame rate is fixed over time. However, since an RFID reader employs a slotted-aloha inventory scheme, the sampling time between successive inventories of the same RFID tag is unpredictable

[43] for each antenna, which is mainly determined by tag population and environmental change. To achieve RFID and CV synchronization, the CV sampling timestamp is regarded as the baseline. Given a CV sample $\left\langle I^{\ell}, t_c^{\ell}\right\rangle$ (i.e., 2D image and timestamp) and a sequence of RFID samples $\langle EPC\left(A_k\right), \phi\left(A_k\right), f\left(A_k\right), t^r\left(A_k\right)\rangle$ (i.e., EPC, RF phase, operating frequency and reading timestamp) corresponding to the reader antennas A_1 and A_2, the RF phase measurements of an RFID tag at the sampling synchronization time t_c^{ℓ} can be calculated by

$$\begin{cases} \phi^{\ell}\left(A_1\right) = \underset{t_r^i(A_1)}{\arg\min} \left|t_r^i\left(A_1\right) - t_c^{\ell}\right| \\ \phi^{\ell}\left(A_2\right) = \underset{t_r^i(A_2)}{\arg\min} \left|t_r^i\left(A_2\right) - t_c^{\ell}\right| \end{cases} s.t. \begin{cases} EPC\left(A_1\right) = EPC\left(A_2\right) \\ f\left(A_1\right) = f\left(A_2\right) \end{cases} \tag{3.37}$$

3.4.5 RFID-CV Fusion for ROI RFID Tag Recognition and Localization

1) RFID and CV Matching Scoring. For an RFID tag to be identified, let $F^{\ell} = \left\{\phi^{\ell}, P^{\ell}\right\}$ denote a pair of RFID-CV fusion data at the sampling synchronization time t_c^{ℓ}, which consists of RF phase measurements $\phi^{\ell} = \left\{\phi\left(A_1^{\ell}\right), \phi\left(A_2^{\ell}\right)\right\}$ of two RFID reader antennas and object proposals $P^{\ell} = \left\{P_1^{\ell}, P_2^{\ell}, ...\right\}$ of an image. Here we define an RFID and CV matching score $\Gamma\left(P_i^{\ell}\right)$ to estimate the likelihood of containing the target RFID tag in the object proposal P_i^{ℓ} as follows:

$$\begin{cases} \Gamma\left(P_i^{\ell}\right) = \underset{\phi_m\left(A^{\ell}\right)\in[-\pi,\pi]}{\max} \dfrac{1}{2}\sum_{k=1}^{2}\left|\cos\Delta\vartheta\left(A_k^{\ell}, P_i^{\ell}\right)\right| \\ \Delta\vartheta\left(A_k^{\ell}, P_i^{\ell}\right) = \phi\left(A_k^{\ell}\right) - \dfrac{4\pi f}{c} d_0\left(A_k^{\ell}, P_i^{\ell}\right) - \phi_m\left(A^{\ell}\right) - {\phi_h}^*\left(A_k, f\right) \end{cases} \tag{3.38}$$

where ${\phi_h}^*\left(A_k, f\right)$ is the known hardware-related phase shift of the antenna A_k at the frequency f. $d_0\left(A_k^{\ell}, P_i^{\ell}\right)$ is the distance of the antenna A_k^{ℓ} to the object proposal P_i^{ℓ}.

Analysis: 1) Matching Principle. According to the dual-antenna hypothesis, the multipath RF phase relative to the antennas A_1 and A_2 may be approximately equal to each other, i.e., $\phi_m\left(A_1^{\ell}, P_i^{\ell}\right) \approx \phi_m\left(A_2^{\ell}, P_i^{\ell}\right)$, so we denote the multipath RF phase as $\phi_m\left(A^{\ell}\right)$. Due to $\left|\cos\Delta\vartheta\left(A_k^{\ell}, P_i^{\ell}\right)\right| \in [0,1]$, the range of the matching score $\Gamma\left(P_i^{\ell}\right)$ lies between 0 to 1. If the RFID tag locates at the object proposal P_i^{ℓ}, $\Gamma\left(P_i^{\ell}\right)$ will have the maximum close to 1; otherwise, the value of $\Gamma\left(P_i^{\ell}\right)$ may be close to 0. Thus, we can search an optimal multipath RF phase $\phi_m\left(A^{\ell}\right)$ within the range of $(-\pi, \pi]$ to solve the maximization problem. *2) Antenna Spacing.* There is a tradeoff between increasing the matching resolution and reducing the multipath RF phase difference. As the antenna spacing increases, the difference between $\left|\cos\Delta\vartheta\left(A_1^{\ell}, P_i^{\ell}\right)\right|$ and $\left|\cos\Delta\vartheta\left(A_2^{\ell}, P_i^{\ell}\right)\right|$ accordingly increases. However, it will combat our tightly-spaced antenna hypothesis mentioned above. Therefore, in practice, when the tracked RFID tags are close to the antenna pair, the small antenna spacing is required. Otherwise, the spacing can be appropriately increased. Due to the 8dBi gain antenna size of $26cm \times 26cm$ in our experiment, the minimum spacing between their casing surface centers is 26cm.

2) Recognition and Localization Based on Matching Score. Due to RF phase periodicity, there may exist one or more object proposals with approximate matching scores. To balance between minimizing the position ambiguity and improving the real-time performance, three successive fusion data $F^{\ell-1}$, F^{ℓ} and $F^{\ell+1}$ are employed to calculate the optimal object proposal P^{ℓ} of the tag being present in F^{ℓ} by finding the maximum of the average matching scores as follows:

$$\begin{aligned} P^{\ell} = & \underset{\substack{P_i^{\ell-1}\in P^{\ell-1} \\ P_j^{\ell}\in P^{\ell} \\ P_k^{\ell+1}\in P^{\ell+1}}}{\arg\max} \frac{1}{3}\sum_{i,j,k}\left[\Gamma\left(P_i^{\ell-1}\right)+\Gamma\left(P_j^{\ell}\right)+\Gamma\left(P_k^{\ell+1}\right)\right] \\ s.t. & \begin{cases} \left\|X\left(P^{\ell-1}\right)-X\left(P^{\ell}\right)\right\| \leq \delta_P \\ \left\|X\left(P^{\ell}\right)-X\left(P^{\ell+1}\right)\right\| \leq \delta_P \end{cases} \end{aligned} \tag{3.39}$$

where $\left\|X\left(P^{\ell-1}\right)-X\left(P^{\ell}\right)\right\|$ is the distance between the centroid coordinates of object proposals $P^{\ell-1}$ and P^{ℓ} in the world coordinate system. δ_P is the maximum reachability distance. Here we set δ_P as $\lambda/4$ (about 8cm) by default to eliminate the position ambiguity due to RF phase periodicity. In our experiment, the sampling rate of the RFID system (i.e., the reading times of an RFID tag per second) is about 0.08 seconds/reading when 40 RFID tags exist in the reading zone. The setting of δ_P can be applied in the scenarios that the tag moves at the maximum speed of 1 m/s, which can meet most the practical applications.

To ensure the system recognition accuracy, RF-Focus performs over an RFID-CV fusion data sequence with the number of N fusion samples. The possible tag trajectory can be represented by an object proposal sequence,

$$Trajectory\left(\mathbb{F}^{\ell}\right) = \left\{P^{\ell+1}, P^{\ell+1}, ..., P^{\ell+N-2}\right\} \tag{3.40}$$

And its average matching score is computed as follows,

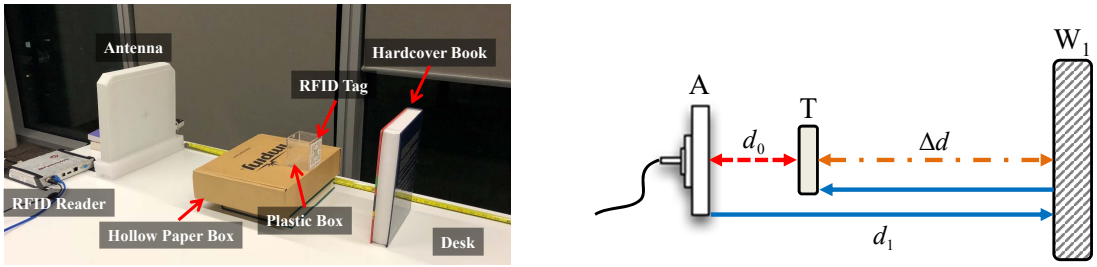
$$\overline{\Gamma}\left(\mathbb{F}^{\ell}\right) = \frac{1}{N-2}\sum_{k=1}^{N-2}\Gamma\left(P^{\ell+k}\right) \tag{3.41}$$

If the value of $\overline{\Gamma}\left(\mathbb{F}^{\ell}\right)$ is larger than a predefined threshold δ_{ROI}, we consider that the tracked RFID tag locates within the ROI and its trajectory is $Trajectory\left(\mathbb{F}^{\ell}\right)$; otherwise, it is a false-positive reading RFID tag outside the ROI.

Analysis: Given M object proposals in each object proposal set, exhaustively searching the tag trajectory with the maximum average matching score has exponential time complexity $O\left(M^N\right)$, which will seriously jeopardize the system's real-time performance. Instead, since RF-Focus performs only using three successive fusion data, the time complexity drops to $O\left[M^3 \times (N-2)\right]$ and the localization error will not accumulate over time.

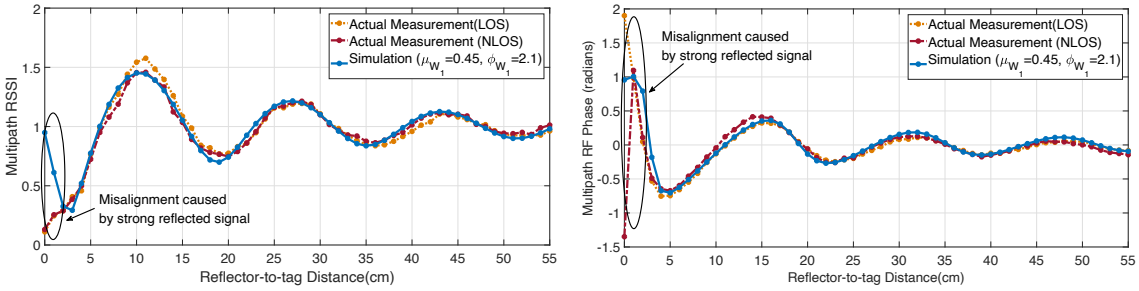
3.5 Evaluation

In this section, we introduce the implementation of RF-Focus. Then we conduct a series of RFID-related experiments to evaluate the performance of dealing with



(a) Experimental scenario. (b) Positional relationship.

Figure 3.8 : Experiment setup.



(a) Multipath RSSI. (b) Multipath RF Phase.

Figure 3.9 : Measured and simulated results in LOS and NLOS.

multipath interference and frequency-dependent hardware characteristics. Finally, we evaluate the performance of RFID and CV fusion for ROI RFID tag recognition and localization.

3.5.1 Implementation

Hardware Implementation. We build a prototype using an Impinj R420 RFID reader, two circularly polarized 8dBi-gain reader antennas, dual-dipole Impinj H47 RFID tags, anti-metal RFID tags and a 2D monocular camera. The 2D camera costs about 70 US dollars while the RFID system costs about 2000 US dollars. Thus, the increase in the RFID and CV fusion system cost compared to previous purely RFID systems can be ignored. The RFID reader is configured to work at 'AutoSet Dense' mode over all operating frequencies from 920.625 MHz to 924.375MHz. The frame rate of the camera is set to 30 fps. A reader antenna with a size of $26cm \times 26cm \times 4.5cm$ can achieve about 6∼10m reading range. The size of an Impinj H47 RFID tag is $4.4cm \times 4.4cm$. The packaged anti-metal RFID tag with the size of $10cm \times 3.3cm$ consists of the Alien Higgs 9630 RFID tag and 0.5cm thick foam material affixed on the back of the tag.

Backend Implementation. The proposed RFID and CV fusion approach is implemented in C# and Matlab 2016b. We exploit Impinj Software Developer's Kit and OpenCV library to record RFID reports (i.e., EPC, RF phase, RSSI, Timestamp) and 2D images. The programs for data collection and performance evaluation both run on a laptop with a 2.3 GHz CPU (Intel i5-6200U) and 4 GB memory.

3.5.2 Effectiveness of Multipath RSSI and RF Phase Modelling

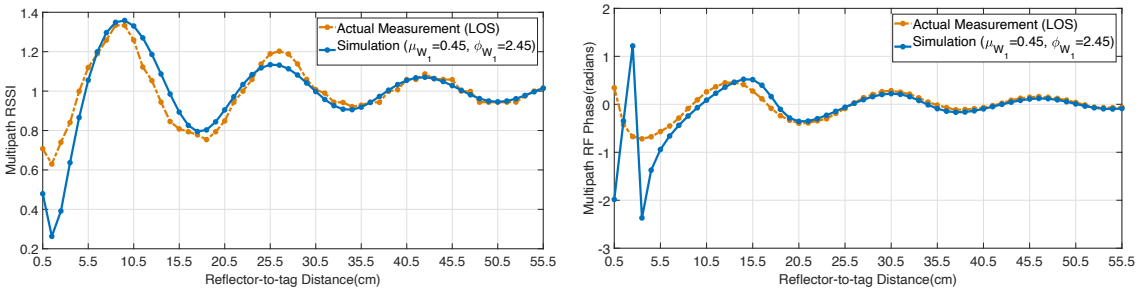
Default Experiment Setup. In Fig. 3.8, a reader antenna A, an Impinj H47 RFID tag T and a reflector W_1 (i.e., 5cm thick hardcover book) are sequentially deployed in a straight line in LOS between the RFID tag and the antenna. We align the position centers of T and W_1 with the center of A such that we have $d_1 = d_0 + 2\Delta d$. We vary the reflector-to-tag distance Δd from 0 to 55cm at 1cm step. RF phase and RSSI measurements are collected in the absence and presence of the reflector W_1 to calculate multipath terms using Eq. 3.19. To intuitively observe the changes in multipath RSSI and RF phase, the reader works at a fixed frequency (920.625MHz).

Metrics. The parameters μ_{W_1}, ϕ_{W_1}, d_0 and d_1 are substituted into Eq. 3.16 to simulate multipath RF phase and RSSI. We observe whether the actual measurements can comply with our simulation.

Simulation Fitting in LOS and NLOS. An RFID tag is placed at the distance of about 52cm to the antenna, i.e., $d_0 = 52cm$. In the NLOS scenario, we use a 3cm thick wood board to block the tag-to-antenna LOS path. Here we set $\mu_{W_1} = 0.45$ and $\phi_{W_1} = 2.1$. In Fig. 3.9, we can see that the actual measurements deviate far from the simulated values when the reflector-to-tag distance Δd is less than about 3cm. In this case, the actual multipath RSSI is very close to 0. The main reason is that the surface of the book W_1 can produce a strong reflected signal in the opposite direction to the incident signal. Since W_1 is very close to the tag T, the reflected signal is of the approximate amplitude as the incident signal, which can cancel the signal power at the tag end. The RFID-active power for tag demodulation falls rather rapidly to 0 and RF phase follows the abrupt change in received power. With the increase of Δd, the amplitude of the reflected signal off W_1 is accordingly reduced such that the multipath measurements in LOS and NLOS comply with the simulated results. Therefore, this experiment can validate the effectiveness of our multipath models.

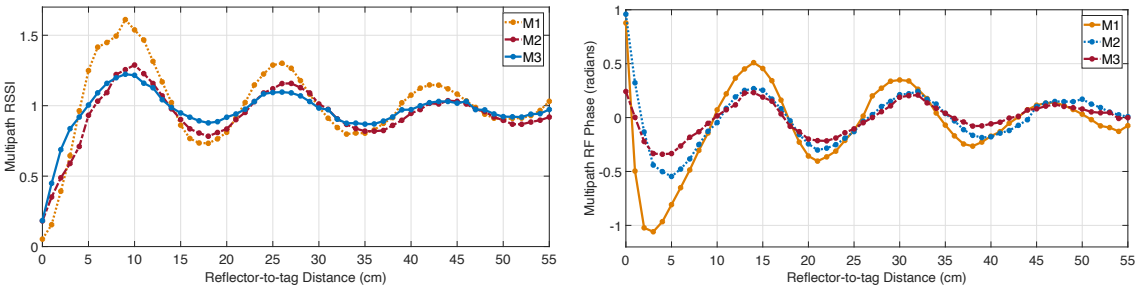
Dealing with Strong Reflected Signal. An anti-metal RFID tag with a magnetic isolator sheet can be used to combat the strong reflected signal [51] when an RFID tag approaches metal or non-metallic objects with high density materials (such as hardcover book, thick glass, wood). In our experiment, the reflector-to-tag distance starts from 0.5cm due to a 0.5cm thick spacer. The tag-to-antenna distance is about 35cm. Figure 3.10 shows that when the reflector is very close to the anti-metal tag, there are no abrupt changes in multipath RSSI and RF phase measurements because the magnetic material can effectively suppress the reflected signal strength. However, the actual measurements deviate far from the simulations given $\mu_{W_1} = 0.45$ and $\phi_{W_1} = 2.5$. As the reflector-to-tag distance increases, the actual measurements are well-matched to the simulations. The impact of magnetic material on backscatter signals is substantially complicated, which we leave in our future work. To guarantee good tag readability, a magnetic sheet is suggested to be inserted between an RFID tag and its tagged object in practice.

Impact of Reflector Material. Three different reflector materials are used, including 3cm thick wood board (M1), 6cm thick softcover book (M2) and 3cm



(a) Multipath RSSI. (b) Multipath RF Phase.

Figure 3.10 : Impact of Magnetic Materials for RFID Tag Antenna Isolation.



(a) Multipath RSSI. (b) Multipath RF Phase.

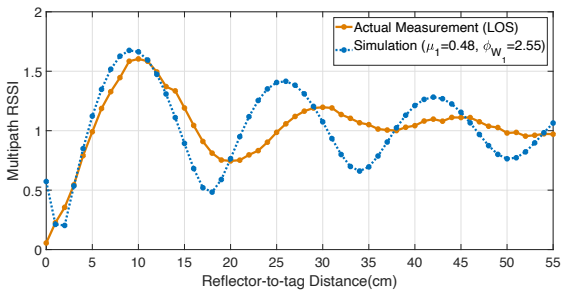
Figure 3.11 : Impact of three reflector materials.

thick softcover book (M3). The tag-to-antenna distance remains unchanged. The empirical ranking of reflection coefficients is $\mu_{W_1}(M_1) > \mu_{W_1}(M_2) > \mu_{W_1}(M_3)$. Figure 3.11 shows that the larger the reflection coefficients, the larger amplitudes the multipath RF phase and RSSI curves have. And the curve periodicities are relatively different from each other. The observations are in accord with the impact of μ_{W_1} and ϕ_{W_1} in our multipath models.

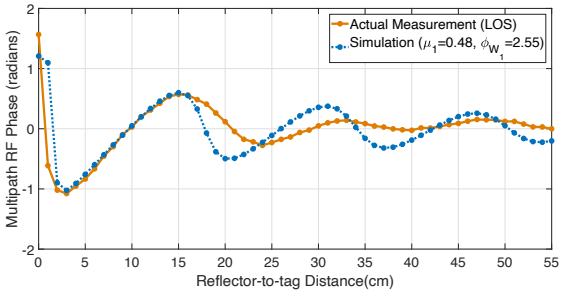
Impact of Tag-to-antenna Distance. Figure 3.12 and Figure 3.13 show the impact of tag-to-antenna distances at about 100cm and 200cm. We can see that the curve amplitude does increase with the larger tag-to-antenna distance, following the proposed multipath models. And when the reflector-to-tag distance ranges from 2cm to about 15cm in multipath RF phase (10cm in multipath RSSI), the measured results can accurately match with the simulation. With the increase of the reflector-to-tag distances, however, the actual measurements deviate far from the simulated results. Recall that we assume that the reflected signal off the reflector dominates multipath signals. However, the longer tag-to-antenna distance weakens the strength of the reflected signal, so the effect of other environmental interference sources (mainly from the desk in our experiment) will accordingly increase. To better comply with the dual-antenna solution to combat multipath interference, the tracked RFID tags should not be too far away from reader antennas.

3.5.3 Hardware-related Phase Shift Estimation Performance

Experiment Setup. Based on the above multipath model discussions, we obtain that 1) to improve the robustness to noise, a reflector with high reflection coefficient should be used; 2) to reduce the effect of other multipath sources, RFID tags

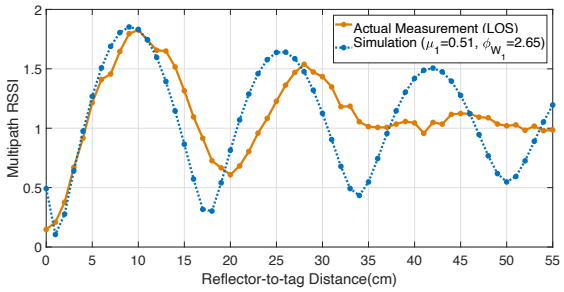


(a) Multipath RSSI ($d_0 = 100cm$).

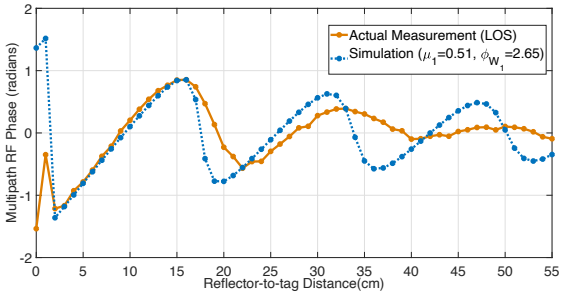


(b) Multipath RF Phase ($d_0 = 100cm$).

Figure 3.12 : Impact of tag-to-antenna distance at $d_0 = 100cm$.



(a) Multipath RSSI ($d_0 = 200cm$).



(b) Multipath RF Phase ($d_0 = 200cm$).

Figure 3.13 : Impact of tag-to-antenna distance at $d_0 = 200cm$.

should be placed as close as possible to the antenna, but these tags must locate outside the near-filed region to avoid inductive coupling effect, i.e., $d_0 > \frac{2D^2}{\lambda} \approx 42.25cm$, where D is the antenna plane diameter (about 26cm) and λ is the wavelength (about 32cm). Three Impinj H47 dual-dipole RFID tags are deployed in the main beam of the directional antenna. The distances of each tag to the antenna casing surface center are 45.8cm, 43cm and 45.8cm, respectively. The spacing between adjacent tags is 16cm, which is enough to remove the coupling effect. The reflector is a metal plane, which locates at a distance of 70cm to the antenna position at the initial. Since the reflector-to-tag distance is relatively large, the tags can capture enough radiated power for demodulation. Then we randomly move the reflector away from the antenna each time.

Baseline. We compare our method with two baselines, i.e., manual measurement (used in D-watch [94]) and automic reader antenna localization Tagspin [26]. A total of 10 persons are invited to measure the distance of each RFID tag to the antenna casing surface center with a tape. Tagspin depends on angle of arrival-based technique to locate RFID reader antennas using two or more RFID-tagged rotating disks moving at a constant speed.

Metrics. We open the antenna's plastic casing to measure the distance between a metal patch center (i.e., the mechanical antenna phase center) and the RFID tag center as the tag-to-antenna reference distance, which is very close to the ground truth. Note that not all reader antennas can be easily opened in practice. We adopt the error distance, i.e., the Euclidean distance between the estimate and the reference distance, as our basis metric. We remove the outliers that deviate far from the reference distance from our algorithm.

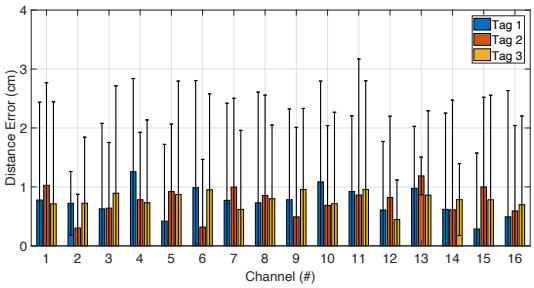


Figure 3.14 : Impact of operating frequency.

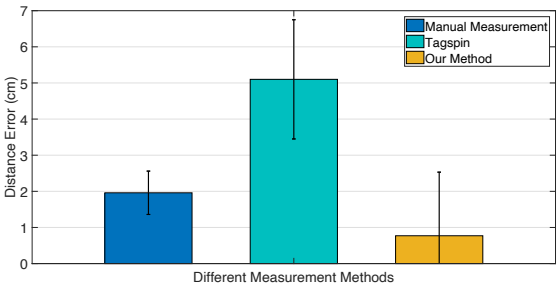


Figure 3.15 : Measurement accuracy comparison.

Accuracy Comparison. Figure 3.14 shows the average tag-to-antenna distance error of three RFID tags over 16 frequency channels. The experiment results are close to each other, which can demonstrate that the antenna phase center does not change with different operating frequencies. Figure 3.15 shows that for our method powered by the multipath model, the average tag-to-antenna distance error is 0.764cm with the standard deviation of 1.76cm, outperforming manual measurement and Tagspin by 2.54 times and 6.61 times, respectively. The localization accuracy of Tagspin is subject to tag-antenna geometry. In our experiment, the estimate of the hardware-related phase shift at Channel 1 is about 1.172 radians.

Effectiveness of Phase Shift Estimation Algorithm. To further verify the effectiveness of our algorithm, we collect 5-minute RF phase measurements of an RFID tag affixed on the rotating disk at the fixed-frequency and hopping-frequency modes, respectively. We firstly measure an initial tag-to-antenna distance at the beginning of rotating the disk and then use the AOA model in [26] to calculate the tag-to-antenna distance at different time when rotating the disk. After that, we can leverage the beamforming technique for AOA estimation [94] to calculate the average power of the aligned received signal. Note that the function $|\cos(\cdot)|$ is to eliminate π radians RF phase periodicity instead of a complex exponential function in our experiment. The maximum average power is 1. We repeat the experiment 30 times with different antenna positions. Figure 3.16 shows that under the fixed-frequency case, the hardware-related phase shifts can be fully removed and the average power error is 0.306. Under the hopping-frequency case, the average power error without any phase-shift compensation is 0.589. Then we use the estimate of hardware-related phase shifts to compensate the non-uniformity effect and the corresponding average power error drops to 0.316 close to the fixed-frequency case.

3.5.4 RFID and CV Fusion for ROI RFID Tag Recognition and Localization

Default Experiment Setup. *1) Reader Antenna and 2D Camera Deployment.* In Figure 3.17, a pair of reader antennas and a 2D camera are all fixed on the drop ceiling. The antenna spacing is set to 35cm. The height of the reader antennas to the desk is about 180 cm. The size of the ROI in the field of 2D camera view is about $63cm \times 63cm$. The system performance is free from the camera's view angle because the corresponding rotation and the translation matrices of the camera relative to

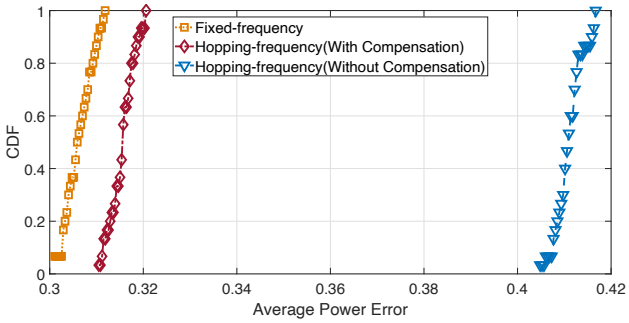


Figure 3.16 : The CDF of average power error.

the ROI plane can be determined in the process of camera calibration. The ROI is located within the main beams of the two directional antennas. *2) ROI RFID Tag.* An Impinj H47 RFID tag (with a 2-cm-thick foam spacer) and an anti-metal RFID tag are affixed on the top of a toy train, respectively. Our prior test demonstrates that the 2cm reflector-to-tag distance is enough to avoid the abrupt change in RF phase. The RFID-tagged toy train moves on the circular track at a speed of about 18cm/s to emulate a conveyorized scenario. The experiment is conducted in LOS between RFID tags and reader antennas. *3) False-positive Reading RFID Tag.* We randomly moves ten false-positive reading RFID tags outside the ROI but in the RFID reading zone. *4) The Number of Object Proposals.* In a video frame, top 10 scoring object proposals are input into our system. In practice, we should depend on the number of target objects in different application scenarios to adjust the number of object proposals.

Multiple ROI RFID Tag Data Simulation. Here we introduce how to simulate multiple RFID-tagged toy trains that move on the track simultaneously. In each experiment, we collect about 30 seconds of data during which the toy train moves along the track 3 laps anticlockwise. Then we divide RFID and CV fusion data of each tag (including the ROI tag on the toy train and ten false-positive reading tags) into the number of K non-overlapping successive segments, just like there are a total of K tagged objects simultaneously moving with different trajectories in the ROI. Each segment has a length of N samples. We set $K = 5$ and $N = 30$ by default. We repeat the experiment 30 times at different starting points of the toy train. Note that unless specifically mentioned, we depend on the above default parameters for system performance evaluation. **Metrics**. *1) Recognition performance.* To characterize RF-Focus's recognition performance, we mainly focus on Receiver Operating Characteristic (ROC) curve by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR), where TPR and FPR may change with different recognition thresholds. *2) Localization performance.* To characterize RF-Focus's localization performance, we leverage Kernelized Correlation Filters (KCF) algorithm [42] to obtain a ground-truth bounding box P_{GT} of the moving toy train in each video frame. Note that the algorithm needs to specify the object image template to be tracked in the first frame. Then we calculate the Intersection over Union (IoU) between the output object proposal P^* and the ground truth P_{GT}, i.e., $IoU = \frac{P^* \cap P_{GT}}{P^* \cup P_{GT}}$. If the value of IoU is higher than 0.5, we consider that the object

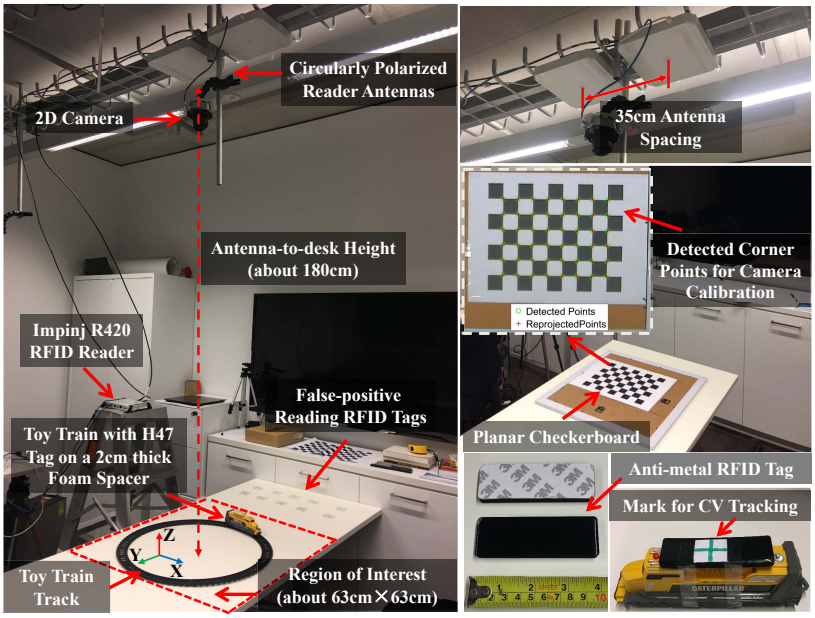


Figure 3.17 : RF-Focus Experiment Setup.

proposal P^* is a desired one correctly containing the ROI RFID tag. Thus, the localization error is defined as $Error_P = 1 - \frac{|\{IoU(P^*)\geq 0.5\}|}{N-2}$, where $|\{IoU\,(P^*) \geq 0.5\}|$ represents the number of correct matching object proposals.

Baseline. We compare our system RF-Focus with three methods, i.e., TagVision [25], "Tell me what I see" [107] and RSSI-differencing.

TagVision. TagVision calculates the matching score by subtracting the measured RF phase difference from the theoretical RF phase difference between the current and initial motion blobs, where each blob is detected by the camera.

Tell me what I see. "Tell me what I see" needs to simultaneously rotate two reader antennas binding with a Kinect V2 depth camera to remove position ambiguity coming from RF phase periodicity. To match our experiment setup with the fixed devices, we leverage its RF phase-based RFID tag localization component to calculate the matching score by subtracting the measured RF phase difference from the theoretical RF phase difference between two reader antennas.

RSSI-Differencing. The RSSI-differencing is to subtract RSSI measurements between two successive RFID readings, which can be formulated as follows by using our multipath RSSI-distance model in Equation (3.4):

$$\frac{\gamma_m\left(A_1^\ell, P_i^\ell\right)}{\gamma_m\left(A_2^\ell, P_i^\ell\right)} = 10^{\left[\frac{rssi\left(A_1^\ell, P_i^\ell\right) - rssi\left(A_2^\ell, P_i^\ell\right)}{20}\right]}\left[\frac{d_0\left(A_2^\ell, P_i^\ell\right)}{d_0\left(A_1^\ell, P_i^\ell\right)}\right]^2 \tag{3.42}$$

According to dual-antenna hypothesis, we have $\gamma_m\left(A_1^\ell, P_i^\ell\right) \approx \gamma_m\left(A_2^\ell, P_i^\ell\right)$. Thus, the matching score can be defined as $\Gamma_{rssi}\left(P_i^\ell\right) = \frac{\gamma_m\left(A_1^\ell, P_i^\ell\right)}{\gamma_m\left(A_2^\ell, P_i^\ell\right)}$. When the value of $\Gamma_{rssi}\left(P_i^\ell\right)$ is close to 1, the tracked RFID tag may exist in P_i^ℓ with the highest probability.

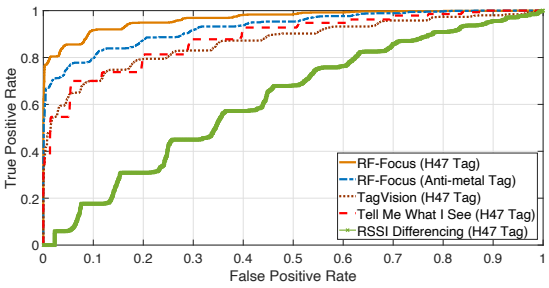


Figure 3.18 : ROI RFID tag recognition performance.

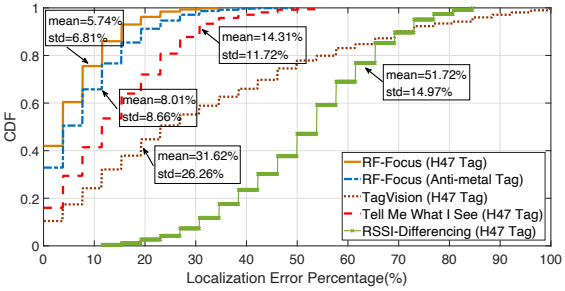
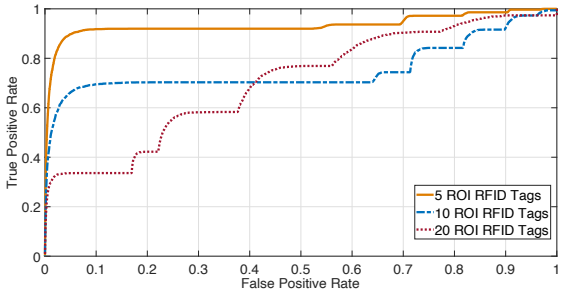


Figure 3.19 : ROI RFID tag localization performance.

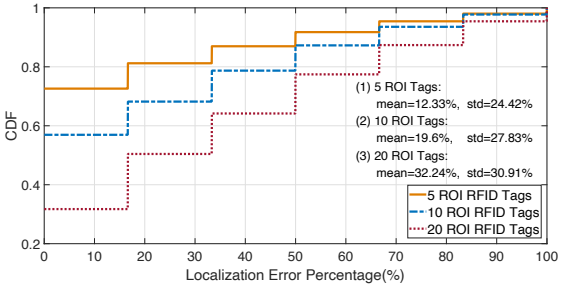
Here we instead use the function of $|\cos(\cdot)|$ to eliminate the RF phase ambiguity πk. Based on the corresponding matching scores of TagVision, "Tell me what I see" and RSSI-differencing, the possible ROI tag trajectory is recovered by our proposed method.

ROI RFID Tag Recognition and Localization Performance. In the default experiment setup, we first evaluate the accuracy of ROI RFID tag recognition over different methods. Figure 3.18 shows the ROC results of RF-Focus, TagVision, "Tell me what I see" and RSSI-differencing. Given a FPR of about 0.1, our system RF-Focus can achieve the TPRs of 0.9167 for H47 tag and 0.798 for anti-metal tag, while TagVision, "Tell me what I see" and RSSI-differencing achieve the TPRs of 0.7, 0.7 and 0.176. Therefore, an appropriate recognition threshold ε_{ROI} in RF-Focus can be chosen to effectively differentiate ROI RFID tags. Figure 3.19 shows the CDF (Cumulative Distribution Function) of localization errors, which can achieve the average localization errors of 5.74%, 8.01%, 31.62%, 14.31% and 51.72% in RF-Focus (H47 tag), RF-Focus (anti-metal tag), TagVision, "Tell me what I see" and RSSI-Differencing.

The recognition and localization performance of RF-Focus is much better than other methods. And the dual-dipole H47 tag can achieve higher localization accuracy than the anti-metal tag. The anti-metal tag is a single-port tag using a dipole antenna configuration, while the Impinj H47 RFID tag is an omni-directional tag, which can capture more stable RF signal energy and reduce tag-to-antenna orientation sensitivity. In the following experiments, the H47 RFID tag is used for evaluation. In TagVision, multipath interference at the initial position may be different from that at other positions, such that TagVision is subject to the changes in multipath interference. In "Tell me what I see", although it enables to combat multipath interference and frequency-dependent hardware characteristics, RF phase periodicity will result in more number of object proposals with the approximate matching scores than RF-Focus. In RSSI-differencing, RSSI is sensitive to the tag-to-antenna orientation, antenna polarization and multipath interference. A comprehensive experiment in [75] is conducted to measure RSSI of an Impinj H47 tag from XY, YZ and XZ planes over different tag-to-antenna distances. The result shows that the RFID tag has an RSSI deviation of 5.75 dB in the main beam because different tag-to-antenna orientations may induce different tag reflection coefficients. However, since the dual-dipole H47 tag can capture enough power for backscatter



(a) Recognition performance.



(b) Localization performance.

Figure 3.20 : Impact of ROI RFID tag number.

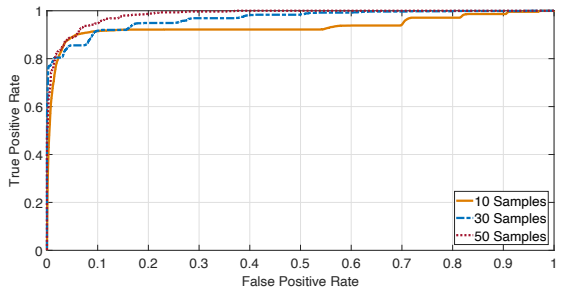


Figure 3.21 : Impact of fusion sample number.

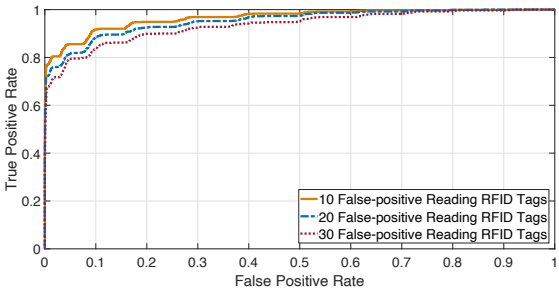


Figure 3.22 : Impact of false-positive reading RFID tag number.

modulation from different orientations, the impact of the tag-to-antenna orientation on RF phase can be ignored. Thus, the RSSI-based performance is worse than RF phase-based systems.

Impact of ROI RFID Tag Number. More RFID-tagged objects within the ROI weaken our system's recognition and localization performance. We run the experiment by varying the number of ROI RFID tags, i.e., 5, 10, 20. Accordingly, the number of object proposals is 50, 100 and 200 due to multiple tag simulation. The number of samples in each fusion data sequence is $N = 10$ such that the simulated object trajectories will be different from each other. Figure 3.20a shows that the TPRs drop from 0.918 for 5 ROI tags to 0.336 for 20 ROI tags given a FPR of 0.1. And Figure 3.20b shows that the average localization error increases from 12.33% for 5 ROI tags to 32.24% for 20 ROI tags. Due to the RF phase periodicity, there may exist more object proposals with approximate matching scores. To minimize the impact of positional ambiguity, a feasible solution is to calculate the average matching score using more successive fusion data. However, it may increase the computation delay.

Impact of RFID and CV Fusion Sample Number. More fusion data samples can improve our system's recognition performance. Here we mainly focus on the impact of sample number on recognition accuracy. Since we leverage three successive fusion samples for tag position estimation each time, the localization accuracy is independent of sample number. We vary the sample number in a sequence over 10, 30 and 50. Figure 3.21 shows that the TPR increases from 0.92 for 10 samples to 0.98 for 50 samples, given a FPR of 0.1. More samples can give more distinc-

tive difference in average matching scores. However, our system has to consume much more time to collect the fusion samples, which affects the real-time recognition performance. Therefore, we can rely on actual application scenarios to set an approximate sample number.

Impact of False-positive Reading RFID Tag Number. More false-positive reading RFID tags outside the ROI slightly decreases our system's recognition accuracy. We run the experiment by varying the number of false-positive reading tags, i.e., 10, 20 and 30. In each experiment, we randomly move part of these tags. The ROC curves are shown in Figure 3.22. Since the average sampling rate of each ROI RFID tag will decrease with more false-positive RFID tags, the time interval between successive inventories of the same tag for two RFID reader antennas will accordingly increase. Consequently, the multipath RF phase difference between the two antenna readings may be magnified. To deal with this issue, more fusion samples are required to guarantee the fine-grained recognition accuracy.

3.6 Conclusion

In this work, we present RF-Focus, which relies on COTS RFID devices and a 2D monocular camera for ROI RFID tag recognition and localization in multipath-prevalent environments. One key innovation is to formulate multipath interference in RSSI and RF phase to deal with phase shifts due to multipath interference and frequency-dependent hardware characteristics. Another is to associate image information with the cleaned RF phase after the unwanted phase shift removal to reduce tag position ambiguity due to RF phase periodicity. Experimental results demonstrate the effectiveness of multipath interference modeling, hardware-related phase shift estimation and RFID-CV fusion. We believe that RF-Focus can be applied in many practical scenarios, including but not limit to baggage sorting and quality assurance.

Chapter 4

CV-assisted Stationary Object 3D Localization

In most RFID localization systems, acquiring a reader antenna's position at each sampling time is challenging, especially for those antenna-carrying robot or drone systems with unpredictable trajectories. In this work, we present RF-MVO that fuses RFID and computer vision for stationary RFID localization in 3D space by attaching a light-weight 2D monocular camera to two reader antennas in parallel. Firstly, the existing monocular visual odometry only recovers a camera/antenna trajectory in the camera view from 2D images. By combining it with RF phase, we design a model to estimate a scale factor for real-world trajectory transformation, along with spatial directions of an RFID tag relative to a virtual antenna array due to the mobility of each antenna. Then we propose a novel RFID localization algorithm that does not require exhaustively searching all possible positions within the pre-specified region. Secondly, to speed up the searching process and improve localization accuracy, we propose a coarse-to-fine optimization algorithm. Thirdly, we introduce the concept of HDOP to measure the confidence level of localization results. Our experiments demonstrate the effectiveness of proposed algorithms and show RF-MVO can achieve 6.23 cm localization error.

4.1 Introduction

Utilizing battery-free RFID for stationary object localization is of importance for a variety of applications. For example, since RFID tags have been widely used in many libraries for automatic book identification, accurate stationary RFID localization would help librarians easily obtain the order of RFID-tagged books on the shelves and find out those misshelved books. Another example is object recognition for pick-and-place robots. The fundamental process behind the task is to accurately identify and locate the objects to be picked up. Since RFID uses wireless signals to capture the globally unique identity (i.e., EPC) of an object from its affixed tag, accurate stationary RFID localization would match the EPCs to the corresponding objects based on position information even in NLOS cases where the objects of interest are occluded.

At present, instead of deploying many high-cost RFID readers around regions of interest for low-cost RFID tag localization like [96, 110, 61, 106], many solutions [95, 79, 60, 93, 19] build a series of virtual antennas by moving a robot binding with multiple reader antennas along a linear trajectory or a pre-deployed magnetic track at a uniform speed. These systems can easily determine the antenna position at each sampling time. However, RFID localization accuracy is subject to uncertainty in antenna position estimation. In dynamic environments, the mobile robot might

rely on its self-adaptive trajectory planning algorithm to avoid obstacles, which will inevitably produce an unpredictable trajectory. And for most of the commonly-used low-cost robots (e.g., about 200 US dollars iRobot Create 2 [46]), they employ wheel odometry [16] to track the robot movement, which cannot provide sufficiently accurate trajectory estimates due to wheel slip in uneven terrain or other adverse conditions [31]. Recently, some researchers [64] explore the potential of a drone equipped with a fisheye camera and a customized RFID signal relay for indoor RFID localization. In indoor environments, the drone cannot depend on its GPS sensor or other sensors to obtain an accurate flight trajectory. Instead, the system deploys a complicated and expensive CV-based tracker consisting of many infrared cameras pre-mounted on a ceiling for flight trajectory tracking.

As the first attempt, we present RF-MVO, a CV-assisted RFID hybrid system for stationary RFID localization in 3-dimensional (3D) space by attaching a light-weight 2-dimensional (2D) monocular camera to COTS reader antennas. *1) Basic Idea.* Our basic idea is to fuse 2D images and RFID data to recover the antenna/camera trajectory and then locate RFID tags. Camera-based visual odometry (VO) (*See Section 4.2 for details*) would allow us to obtain more accurate camera/antenna trajectory estimates than wheel odometry [31]. However, without additional depth information for the monocular camera, the existing MVO [68, 69] only recovers the camera trajectory in the camera view from 2D images. To transform it to the real-world version, we need to know the relationship between distances in the camera view and the real world, which is called *scale factor*. Fortunately, RF phase is a function of tag-to-antenna distance, providing an opportunity to estimate the scale factor. As the antenna moves across the region of interest, RF-MVO relies on the its mobility to build a series of virtual antenna arrays. In each array, it transforms a piece of the camera/antenna trajectory in the camera view and depth-enabled RF phase into a spatial power spectrum that characterizes the likelihood of each scale factor and spatial DOA (i.e., azimuth and elevation angles) of an RFID tag relative to the antenna array in space. Then the scale factor and DOA can be estimated by searching the peak of the proposed spectrum. After recovering the real-world camera/antenna trajectory using the estimated scale factor, RF-MVO would then determine a stationary RFID tag position. *2) RFID-CV Fusion Advantage.* At a high level, RF-MVO would perform accurate localization without specific robot platforms that could carry reader antennas to move with a known trajectory compared to RFID-only solutions. A benefit is that, for example, it may help the existing drone-based RFID localization systems reduce the cost of capturing the flight trajectory of a drone. Also, based on position matching, RF-MVO associates the collected EPCs with the corresponding objects for object recognition. Compared to CV-only technique, RF-MVO would provide information-rich recognition among multiple objects with the same appearance without any pre-training, and deal with NLOS cases.

To realize RF-MVO, we need to address three challenges:

Challenge 1: To our knowledge, most of the existing motion-based localization methods perform with the holographic technique that pre-specifies the region of interest in 3D space where the target RFID tag may exist, partitions the spatial region

Table 4.1 : Terms used in the description of RF-MVO

Term	Definition
L, M, λ	Number of elements in an antenna array, Number of reader antennas, RF signal wavelength
α_A, β_A, γ	Azimuth angle, Elevation angle, Scale factor
$\overline{\alpha}_A$, $\overline{\beta}_A$	Averaged azimuth and elevation angles
γ_{min}, γ_{max}	Minimum and maximum of scale factor
$\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$	Searching spacings of azimuth angle, elevation angle and scale factor
$\Delta\alpha_{opt}$, $\Delta\beta_{opt}$	Searching spacings of azimuth and elevation angles for optimization algorithm
γ^*, μ	Optimized scale factor, Searching threshold for optimization algorithm
$\Delta\boldsymbol{X}_A$	Displacement between adjacent antenna positions in the real world
$\Delta\boldsymbol{X}_C$	Displacement between adjacent camera positions in the camera view
$\Delta\boldsymbol{X}_{A,C}$, $\Delta\boldsymbol{X}_{A_{ref}}$	Displacement of antenna relative to camera, Displacement of antenna relative to reference one
$\boldsymbol{X}_T$, $\boldsymbol{X}_A$	RFID tag and antenna coordinates in the local world coordinate system before optimization
$\boldsymbol{X}_C$	Camera coordinate in the camera view calculated by MVO
$\widehat{\boldsymbol{X}_T}$, $\widehat{\boldsymbol{X}_T^*}$	RFID tag coordinates in the global world coordinate system before and after optimization
$\boldsymbol{X}_T^*$	RFID tag coordinate in the local world coordinate system after optimization
d_A, Δd_A	Tag-to-antenna distance, Tag-to-antenna distance difference
φ_A, $\Delta\varphi_A$, φ_h	RF phase reported by an RFID reader, RF phase difference, Phase shift caused by RFID hardware
N_A, ΔN_A	Phase ambiguity (i.e., an unknown integral multiple of 2π), Phase ambiguity difference

into many cuboids with the centimeter-level size and then calculates the likelihood of containing the target in each cuboid based on RF phase. The localization accuracy depends on the resolution in the cuboid size. And as the surveillance region of interest increases in 3D space, huge computation will affect the system's real-time performance. To deal with this issue, we propose a novel DOA-based RFID localization algorithm that builds a series of spatial lines passing through antenna positions in an antenna array and whose direction vectors are estimated by DOAs, and then calculates the interaction of these lines as the final RFID tag position.

Challenge 2: To obtain a more accurate scale factor and DOA, we require setting smaller searching granularities in the proposed spatial power spectrum. However, it will inevitably jeopardize the real-time performance. To address this issue, we propose a coarse-to-fine optimization algorithm to accelerate our task, including the following steps: (1) the step of DOA refinement is to refine the low-resolution DOA given the high-resolution scale factor; (2) the step of 3D RFID localization is to locate the tag position given the refined DOA; (3) the step of nonlinear optimization is to simultaneously refine the 3D tag position and scale factor that are interrelated with each other. The algorithm recursively takes iterations before it converges.

Challenge 3: As we move the antennas across RFID tags, our system will obtain multiple tag position candidates over antenna arrays. And in an antenna array, since multiple RFID tags may be simultaneously read, our system will produce multiple scale factor candidates. It is still challenging to select out an optimal tag position and scale factor. To deal with it, a key intuition is that the tag localization error is sensitive to the change in tag-antenna geometry, which means antenna elements widely separated can provide higher estimation accuracy than those closely spaced. Hence, we present HDOP, widely used in Global Positioning System (GPS) [55], to measure the confidence level of the estimated tag position and scale factor due to the effect of tag-antenna geometry.

This work makes the following contributions:

1) To the best of our knowledge, RF-MVO is the first CV-assisted RFID system for stationary RFID localization in 3D space. Unlike RFID-only systems, RF-MVO works without driving reader antennas along a pre-defined trajectory. And RF-MVO can also overcome the NLOS problem in CV-only object recognition by associating the reported EPCs with the corresponding objects based on the position information.

2) This work introduces a novel 3D RFID localization algorithm, without exhaustively searching each divided region (e.g., a cuboid) of being likely to contain a target RFID tag in the pre-specified region of interest.

3) This work introduces HDOP to determine the effect of tag-antenna geometry on estimation accuracy in order to find out the optimal localization result.

4) We build a prototype using COTS RFID devices and a 2D monocular camera to demonstrate that when only using two reader antennas, RF-MVO can achieve the average of 6.23 cm tag localization error in 3D space and the average of 0.0158 scale-factor estimation error.

4.2 Visual Odometry Background

VO technique [31, 68, 69] is to estimate camera poses (i.e., positions and orientations) with respect to its surroundings by analyzing a sequence of images.

1) Visual Odometry Types. According to different camera types, there are monocular VO, stereo VO, and RGB-D VO. Compared to MVO, stereo and RGB-D VO can recover a camera trajectory in the real world with the help of depth information. However, large-size stereo or RGB-D cameras are unsuitable for small robots or drones. And in the scene where the distance from a stereo camera to a working region is much higher than the distance between two camera lens, stereo VO will degenerate to the monocular case. For an RGB-D camera (e.g., Kinect V2), it suffers from the measurement distance limit and errors in depth estimation. Hence, we mainly focus on MVO in this work.

2) MVO Workflow. The steps to estimate camera trajectory are summarized as follows: *1) Estimate Camera Pose of the Second Image.* MVO extracts 2D image features from the second image, and then matches them to the first image. The geometric relation of rotation and translation between the two images is calculated from the 2D-to-2D feature correspondences. MVO only obtains the camera pose of the second image relative to the first image up to an unknown scale factor in the real world.*(2) Recover Remaining Camera Trajectory.* The transformation between two different images induces different scale factors from 2D-to-2D feature correspondences. To unify the scale factor for camera trajectory recovery, MVO then finds the 3D-to-2D correspondences between feature points triangulated from the previous two images and the current image. *3) Bundle Adjustment.* The errors in camera pose estimation will accumulate over time. MVO uses windowed bundle adjustment [32, 86] to optimize the camera poses over the last L images. This process will not change the initial scale factor.

3) Camera Trajectory Recovery Accuracy. In VO, the estimated trajectory error is defined as the difference between the ground truth provided by the dataset and estimated trajectory. Many existing stereo or RGB-D VO systems (e.g., ORB-SLAM2 [69]) can achieve centimeter- or millimeter-level trajectory error following different datasets. Hence, it is possible to transform a camera trajectory in the camera view estimated by MVO to the real-world version with such high precision as long as an accurate scale factor is provided.

4.3 DOA and Scale Factor Searching for 3D RFID Localization

In this section, we first utilize the mobility of reader antennas to emulate a sequence of virtual antenna arrays. In each antenna array, we then combine tag-to-antenna path differences of two adjacent antennas calculated by DOA estimation theory and RF phase measurements to build a DOA-based spatial power spectrum. It can simultaneously search the spatial DOA (i.e., azimuth and elevation angles) of an RFID tag relative to an antenna array and the scale factor of the camera/antenna trajectory transformation from the camera view to the real world. After that, we construct the spatial line passing through the RFID tag and each antenna element based on estimated DOA and real-world antenna trajectory. By calculating the interaction of all spatial lines, the system can locate stationary tags in 3D space.

4.3.1 DOA and Scale Factor Searching

Two Basic Assumptions for DOA Estimation. At first, we present two essential preconditions that enable DOA estimation of RFID tags:

1) Narrowband assumption means that the bandwidth inverse is much larger than the signal propagation time across an antenna array. In the backscatter communication, suppose that the Miller-modulated subcarrier coding performs with the subcarrier cycles of M_{cycle} and the backscatter link frequency of BLF Hz. The time duration of receiving the length of L_{EPC} bits EPC packet from an RFID tag is calculated by $\frac{M_{cycle} \times L_{EPC}}{BLF}$, which is equal to the bandwidth. For example, given a LLRP GEN 2 mode '3' configuration (i.e., Dense Reader, $M_{cycle} = 8$, $BLF = 170.6$ kHz) in our RFID reader, the time duration for a 128-bits EPC packet is about 6×10^{-3}s. If the length of the antenna array is $D = 0.5$ meters, the signal will take the time of about 1.67×10^{-9}s to travel through the array, which is far less than the bandwidth. Thus, the signal arriving at each antenna element can be regarded as a narrow signal.

2) Far-field assumption means the signals transmitting from an RFID tag in a given direction to each antenna element are approximately parallel. To satisfy that, the distance between the tag and each antenna element is required to be larger than $\frac{2D^2}{\lambda}$, where λ is the signal wavelength. Again, given $D = 0.5$ meters, the minimum working distance for our system is $\frac{2D^2}{\lambda} \approx \frac{2 \times 0.5^2}{0.32} \approx 1.56$ meters. Increasing the tag-to-antenna distance can make the far-field assumption be better complied. However, it will accordingly reduce the sampling rate and increase the effect of multipath interference on RF phase. Hence, we need to choose an appropriate antenna array

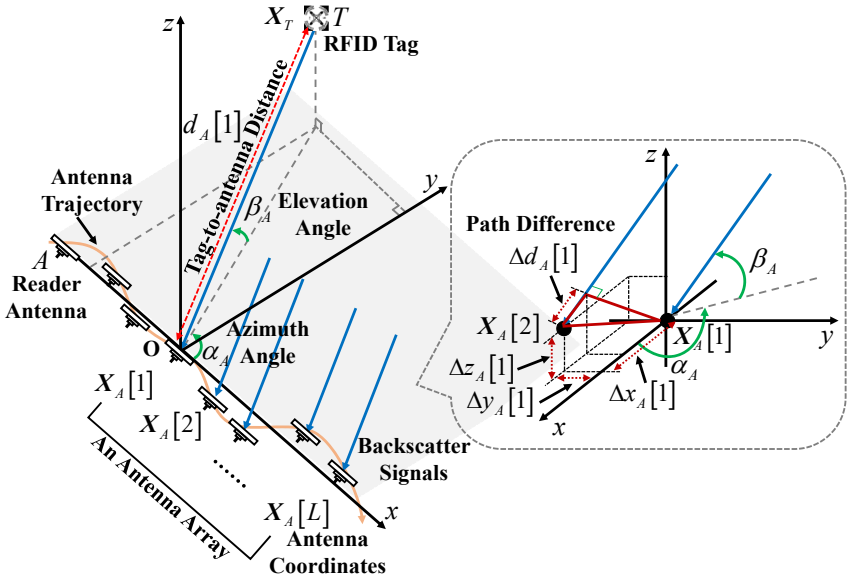


Figure 4.1 : DOA estimation model.

length to ensure the DOA estimation accuracy, which will be discussed for detail in Section 4.7.5.

Obtain Tag-to-Antenna Path Difference based on DOA Theory. To intuitively introduce our basic idea, a geometric model in 3D space is shown in Fig. 4.1. As a reader antenna A moves over time, it reads a stationary RFID tag at different positions just like deploying multiple virtual antennas at each of these positions. We assume that the virtual antenna array contains the length of L elements and two successive antenna arrays have $(L-1)$ overlapping elements. Each antenna array corresponds to a local world coordinate system whose origin is located at the position of the first element. In this coordinate system, a reported RF phase for fusion use is measured at an antenna position $\boldsymbol{X}_A[i]$, where $\boldsymbol{X}_A[1] = \mathbf{0}$. The displacement between two adjacent antenna positions is $\Delta\boldsymbol{X}_A[i] = \boldsymbol{X}_A[i+1] - \boldsymbol{X}_A[i]$, where $\Delta\boldsymbol{X}_A[i] = (\Delta x_A[i], \Delta y_A[i], \Delta z_A[i])$. Suppose that the DOA (i.e., azimuth and elevation angles) of RFID backscatter signals arriving at the antenna array are denoted as α_A and β_A, and the tag-to-antenna distance from a target RFID tag at $\boldsymbol{X}_T$ to the antenna element at $\boldsymbol{X}_A[i]$ is denoted by $d_A[i]$. According to DOA theory[87], the path difference $\Delta d_A[i]$ between $d_A[i]$ and $d_A[i+1]$ is approximately,

$$\begin{aligned}\Delta d_A[i] =& d_A[i+1] - d_A[i] \\ \approx& \Delta x_A[i] \times cos\alpha_A cos\beta_A + \\ & \Delta y_A[i] \times sin\alpha_A cos\beta_A + \\ & \Delta z_A[i] \times sin\beta_A\end{aligned} \tag{4.1}$$

Since MVO can output the camera position $\boldsymbol{X}_C[i]$ in the camera view and $\boldsymbol{X}_C[1] = \mathbf{0}$, the camera displacement is $\Delta\boldsymbol{X}_C[i] = \boldsymbol{X}_C[i+1] - \boldsymbol{X}_C[i]$, where $\Delta\boldsymbol{X}_C[i] = (\Delta x_C[i], \Delta y_C[i], \Delta z_C[i])$. In the real world, the antenna displacement

is the same as the camera displacement. Given a scale factor $\gamma[k]$, we have

$$\Delta \boldsymbol{X}_A[i] = \gamma \times \Delta \boldsymbol{X}_C[i] \tag{4.2}$$

Note that once receiving an RFID and CV fusion sample to build a new antenna array, MVO estimates the corresponding camera trajectory in this array. Due to different different geometric relationship of the first two images, the scale factor in each antenna array might not be the same.

Obtain Tag-to-Antenna Path Difference based on RF Phase. According to the standard phase-distance model in an ideal environment [43, 105, 18, 112], if the measured RF phase is $\varphi_A[i]$ at the antenna position $\boldsymbol{X}_A[i]$, we have

$$\varphi_A[i] + N_A[i] = \frac{4\pi}{\lambda} d_A[i] + \varphi_h \tag{4.3}$$

where φ_h represents the phase shift caused by RFID hardware (i.e., RFID tag circuits, reader antenna reflections, feed cables, and reader circuits). The unknown parameter N_A, called phase ambiguity, is an integral multiple of 2π to make φ_A fall within $[0, 2\pi]$.

By taking a difference between RF phase $\varphi_A[i]$ and $\varphi_A[i+1]$, we can remove the phase shift due to the hardware characteristics and then obtain the path difference,

$$\Delta d_A[i] = \frac{\lambda}{4\pi}\left(\Delta\varphi_A[i] + \Delta N_A[i]\right) \tag{4.4}$$

where

$$\begin{cases} \Delta\varphi_A[i] = \varphi_A[i+1] - \varphi_A[i] \\ \Delta N_A[i] = N_A[i+1] - N_A[i] \end{cases} \tag{4.5}$$

Determine the Range of Scale Factor. We assume all of the antenna displacements in an antenna array is within $[0, \lambda/4]$, i.e., $|\Delta d_A[i]| < \lambda/4$ (about 8 cm). Due to $\Delta\varphi_A[i] \in [-2\pi, 2\pi]$, $\Delta N_A[i]$ can be determined as follows:

$$\Delta N_A[i] = \begin{cases} 2\pi, & -2\pi \le \Delta\varphi_A[i] < -\pi \\ 0, & |\Delta\varphi_A[i]| \le \pi \\ -2\pi, & \pi < \Delta\varphi_A[i] \le 2\pi \end{cases} \tag{4.6}$$

According to the triangle rule that the absolute difference between the lengths of two sides is less than the length of the third side, we combine Eq. 4.2 with Eq. 4.4 to obtain

$$\gamma > \frac{\lambda}{4\pi} \frac{|\Delta\varphi_A[i] + \Delta N_A[i]|}{\|\Delta \boldsymbol{X}_C[i]\|} \tag{4.7}$$

Thus, the minimum scale factor is calculated by

$$\gamma_{min} = \max_{i \in [1, L-1]} \left\{ \frac{\lambda}{4\pi} \frac{|\Delta\varphi_A[i] + \Delta N_A[i]|}{\|\Delta \boldsymbol{X}_C[i]\|} \right\} \tag{4.8}$$

In addition, when the speed at which the camera captures images is set to 30 frames per second (FPS) and $|\Delta d_A[i]| < \lambda/4$, the maximum motion speed of the

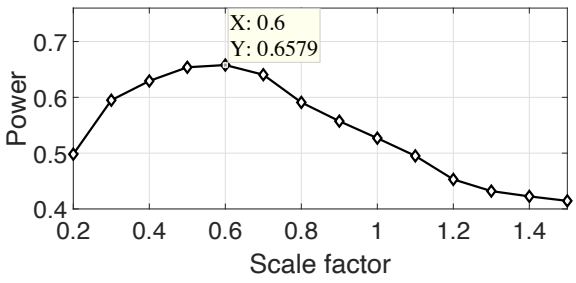


Figure 4.2 : Peak searching.

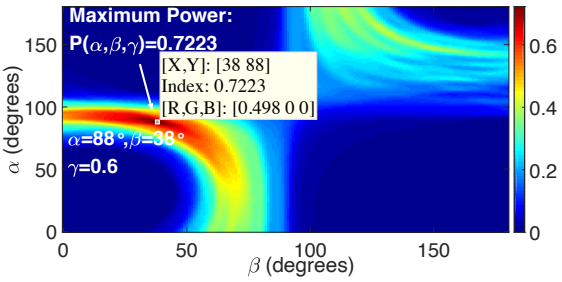


Figure 4.3 : Power spectrum.

reader antenna can achieve as high as 240 cm/s, which can meet the requirement for most of applications. However, previous work [115] indicates that when an RFID tag moves at high speed, most RFID systems may miss a lot of reported packets from the tag. And the RFID system will suffer from serious Doppler frequency shift [100], thereby producing a non-negligible phase shift in RF phase measurement. To minimize the effect of Doppler frequency shift and guarantee that RFID tags can be read at a sufficiently high sampling rate (i.e., 10 times/s at least), we move the antennas with a relatively low speed (i.e., about 20 cm/s $\sim$ 35 cm/s in our experiment). Here, we set the maximum speed at 40 cm/s. Due to $\|\Delta \boldsymbol{X}_A[i]\| \leq 40/10$, the maximum scale factor is

$$\gamma_{max} = \min_{i \in [1, L-1]} \left\{ \frac{4}{\|\Delta \boldsymbol{X}_C[i]\|} \right\} \tag{4.9}$$

Build Spatial Power Spectrum for DOA and Scale Factor Searching. Given a scale factor $\gamma \in (\gamma_{min}, \gamma_{max})$ in an antenna array, the spatial power spectrum of backscatter signals along an azimuth angle $\alpha_A \in [0°, 180°]$ and an elevation angle $\beta_A \in [0°, 180°]$ is a 2D intensity graph in which each pixel characterizes the likelihood of DOA (i.e., azimuth and elevation angles) under a scale factor,

$$\boldsymbol{P}(\alpha_A, \beta_A, \gamma) = \frac{1}{L-1} \sum_{i=1}^{L-1} \cos\left(\Delta\varphi_A[i] - \frac{4\pi}{\lambda} \Delta d_A[i] \right) \tag{4.10}$$

where $\Delta d_A[i]$ is computed by Eq. 4.1 and Eq. 4.2,

$$\begin{aligned} \Delta d_A[i] \approx & \gamma \times \Delta x_C[i] \times cos\alpha_A cos\beta_A + \\ & \gamma \times \Delta y_C[i] \times sin\alpha_A cos\beta_A] + \\ & \gamma \times \Delta z_C[i] \times sin\beta_A \end{aligned} \tag{4.11}$$

The purpose of introducing the cosine function is described as follows. According to Eq. 4.6, we can rely on the RF phase difference $\Delta\varphi_A[i]$ to estimate the phase ambiguity difference $\Delta N_A[i]$. However, when the ground-truth of $\Delta\varphi_A[i]$ is close to $\pm\pi$ or $\pm 2\pi$, measurement error in RF phase may make $\Delta N_A[i]$ deviate far from the ground truth. To deal with this problem, the cosine function is used here to remove unknown phase ambiguity. Also, this searching model is imperfect and will be mainly affected by multipath interference [72], Doppler frequency shift, tag mutual

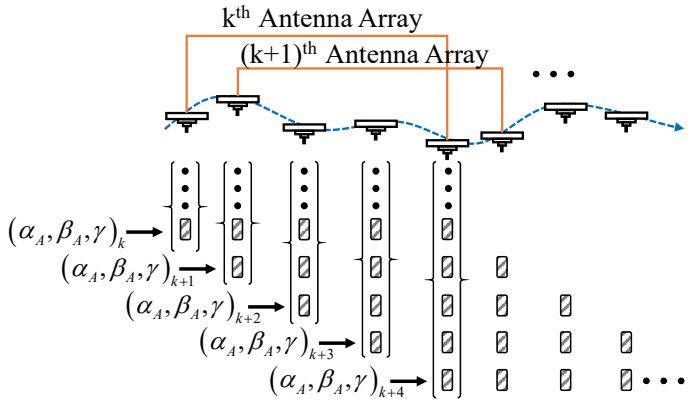


Figure 4.4 : DOA averaging.

coupling [105, 38], and accumulation error in camera pose estimation. In this case, the closer the parameters α_A, β_A and γ to the ground truth, the closer the pixel intensity maximum is to 1. Otherwise, the minimum is close to 0.

Given M physical reader antennas on the cart, the estimated DOA and scale factor are:

$$\begin{cases} \boldsymbol{S}\left(\alpha_{A_i}, \beta_{A_i}, \gamma\right) = \max\limits_{\alpha_{A_i}, \beta_{A_i} \in [0, 180°]} \boldsymbol{P}\left(\alpha_{A_i}, \beta_{A_i}, \gamma\right) \\ \gamma = \arg\max\limits_{\gamma \in (\gamma_{min}, \gamma_{max})} \dfrac{1}{M} \sum\limits_{i=1}^{M} \boldsymbol{S}\left(\alpha_{A_i}, \beta_{A_i}, \gamma\right) \\ \left(\alpha_{A_i}, \beta_{A_i}\right) = \arg\max\limits_{\alpha_{A_i}, \beta_{A_i} \in [0, 180°]} \boldsymbol{P}\left(\alpha_{A_i}, \beta_{A_i}, \gamma\right) \end{cases} \tag{4.12}$$

Here we take an example to intuitively introduce above equations. Suppose that the searching spacings of DOA and scale factor searching are $\Delta\alpha = 1°$, $\Delta\beta = 1°$ and $\Delta\gamma = 0.1$, respectively. The range of the scale factor is within $[0.2, 1.5]$. In Fig. 4.2, the highest power peak corresponds to $\gamma = 0.6$, while the ground-truth scale factor is 0.568. Then we substitute the estimated scale factor to calculate $P\left(\alpha_A, \beta_A, \gamma\right)$ over all possible azimuth and elevation angles. In Fig. 4.3, the spatial power spectrum shows that the DOA corresponding to the maximum power are $\alpha_A = 88°$ and $\beta_A = 38°$ while the ground-truth angles are 80.54° and 31.08°.

4.3.2 RFID Tag Localization in 3D Space

Take the average of DOAs. As shown in Fig. 4.4, a reader antenna is involved in multiple overlapping antenna arrays. To obtain the DOA of an RFID tag relative to each antenna, we take the average of DOAs over each overlapping antenna array that contains the antenna. For the k^{th} reader antenna, its averaged DOA (i.e., $\overline{\alpha}_A[k]$ and $\overline{\beta}_A[k]$) is calculated by

1) If $1 \leq k < L$, we have

$$\begin{cases} \overline{\alpha}_A[k] = \dfrac{1}{k}\sum_{l=1}^{k} \alpha_A[l] \\ \overline{\beta}_A[k] = \dfrac{1}{k}\sum_{l=1}^{k} \beta_A[l] \end{cases} \tag{4.13}$$

2) If $k \geq L$, we have

$$\begin{cases} \overline{\alpha}_A[k] = \dfrac{1}{L}\sum_{l=k-L+1}^{k} \alpha_A[l] \\ \overline{\beta}_A[k] = \dfrac{1}{L}\sum_{l=k-L+1}^{k} \beta_A[l] \end{cases} \tag{4.14}$$

Locate RFID Tag in a Local Coordinate System. In the local coordinate system, the equation of a spatial line $\ell[k,i]$ passing through the target RFID tag coordinate $\boldsymbol{X}_T[k]$ and the antenna element coordinate $\boldsymbol{X}_A[k,i]$ is given by

$$\frac{x_T[k]-x_A[k,i]}{u_A[k,i]} = \frac{y_T[k]-y_A[k,i]}{v_A[k,i]} = \frac{z_T[k]-z_A[k,i]}{w_A[k,i]} \tag{4.15}$$

where

$$\begin{cases} \boldsymbol{X}_T[k] = (x_T[k], y_T[k], z_T[k]) \\ \boldsymbol{X}_A[k,i] = (x_A[k,i], y_A[k,i], z_A[k,i]) \end{cases} \tag{4.16}$$

According to Eq. 4.2, we have

$$\boldsymbol{X}_A[k,i] = \gamma[k] \times (\boldsymbol{X}_C[k,i] - \boldsymbol{X}_C[k,1]) \tag{4.17}$$

The spatial direction vector$(u_A[k,i], v_A[k,i], w_A[k,i])$ is calculated by:

$$\begin{cases} u_A[k,i] = \cos\overline{\alpha}_A[k+i-1]\cos\overline{\beta}_A[k+i-1] \\ v_A[k,i] = \sin\overline{\alpha}_A[k+i-1]\cos\overline{\beta}_A[k+i-1] \\ w_A[k,i] = sin\overline{\beta}_A[k+i-1] \end{cases} \tag{4.18}$$

As shown in Fig. 4.5, we construct a total of $M \times L$ spatial lines to solve their intersection. In practice, however, there might exist multiple non-intersecting lines and even parallel lines in 3D space. We regard the point nearest to these lines as the optimal result, which can be solved using Singular Value Decomposition (SVD) algorithm [39].

Transform to a Global Coordinate System. To obtain the position relationship among all RFID tags for application use, we transform the tag position in this local coordinate system to that in a global coordinate system whose origin is at the initial reference antenna position when the cart starts moving. Hence, the RFID tag position $\widehat{\boldsymbol{X}}_T[k+1]$ $(k \geq 1)$ in the global world coordinate system is

$$\widehat{\boldsymbol{X}}_T[k+1] = \boldsymbol{X}_T[k+1] + \sum_{l=1}^{k} \gamma[l] \times \Delta\boldsymbol{X}_C[l,1] \tag{4.19}$$

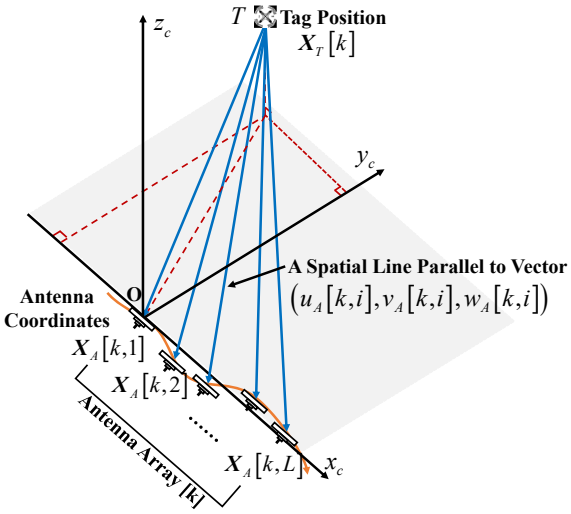


Figure 4.5 : RFID tag localization.

4.4 Coarse-to-Fine Tag Position and Scale Factor Optimization

In the spatial power spectrum, exhaustively searching the high-resolution DOA from 0° to 180° and scale factor from γ_{min} to γ_{max} is a time-consuming process. The time complexity is $\boldsymbol{O}\left(\frac{180}{\Delta\alpha}\times\frac{180}{\Delta\beta}\times\frac{\gamma_{max}-\gamma_{min}}{\Delta\gamma}\right)$, where $\Delta\alpha$, $\Delta\beta$ and $\Delta\gamma$ represent the corresponding searching granularities. In this section, a coarse-to-fine optimization algorithm is proposed to speed up our system. Initially, the algorithm starts by obtaining a coarse-resolution DOA and fine-resolution scale factor based on the proposed spatial power spectrum. It then zooms in the DOA with smaller searching granularity to refine the tag position based on the proposed DOA-based localization algorithm. After that, it performs a joint optimization of the scale factor (as well as tag position) over RFID and CV fusion data in an antenna array, which can recursively refine the DOA and tag position with the optimized scale factor. In the following, we introduce the coarse-to-fine optimization algorithm in detail.

1) Initialization. To reduce the searching computation, we firstly obtain a high-resolution scale factor γ^1 and low-resolution DOA (α_A^1 and β_A^1) given a small searching granularity of $\Delta\gamma$ and a relatively large searching granularity of $\Delta\alpha$ and $\Delta\beta$, instead of simultaneously searching DOA and a scale factor with fine granularities.

2) DOA Refinement. A coarse-to-fine DOA refinement is to refine the DOA using small searching granularities of $\Delta\alpha_{opt}$ and $\Delta\beta_{opt}$. The angular searching ranges are $\left(\alpha_A^l \pm \mu\right) \in [0°, 180°]$ and $\left(\beta_A^l \pm \mu\right) \in [0°, 180°]$, where μ is the predefined searching threshold. In the l^{th} iteration, α_A^{l+1} and β_A^{l+1} are updated given γ^l:

$$\left(\alpha_A^{l+1}, \beta_A^{l+1}\right) = \underset{\substack{\alpha_A\in\left[\alpha_A^l-\mu,\alpha_A^l+\mu\right]\\ \beta_A\in\left[\beta_A^l-\mu,\beta_A^l+\mu\right]}}{\arg\max} \boldsymbol{P}\left(\alpha_A^l, \beta_A^l, \gamma^l\right) \tag{4.20}$$

We update the averaged DOA according to Eq.(13) and (14), so we can obtain $\overline{\alpha}_A^{l+1}$

and$\overline{\beta}_A^{l+1}$.

3) 3D RFID Localization. In the l^{th} iteration, the 3D tag position $\boldsymbol{X}_T^l$ is calculated given γ^l and averaged DOAs $\left\{\overline{\alpha}_{A_j}^{l+1}, \overline{\beta}_{A_j}^{l+1}\right\}(j=1,...,M)$ corresponding to M reader antennas (refer to Section 4.3.2):

$$\boldsymbol{X}_T^l = \underset{\substack{i\in[1,L]\\ j\in[1,M]}}{SVD}\left\{\ell\left(\overline{\alpha}_{A_j}^{l+1}[i], \overline{\beta}_{A_j}^{l+1}[i], \gamma^l\right)\right\} \tag{4.21}$$

4) Joint Optimization. In the l^{th} iteration, the tag position $\boldsymbol{X}_T^{l+1}$ and scale factor γ^{l+1} are both updated by minimizing the distance error given $\boldsymbol{X}_T{}^l$ and γ^l:

$$\left(\boldsymbol{X}_T^{l+1}, \gamma^{l+1}\right) = \underset{\boldsymbol{X}_T^l, \gamma^l}{\arg\min} \sum_{i=1}^{L-1}\sum_{j=1}^{M}\left\|\Delta d_{A_j}^l[i] - \Delta d_{C_j}^l[i]\right\| \tag{4.22}$$

where

$$\begin{cases} \Delta d_{A_j}^l[i] = \dfrac{\lambda}{4\pi}\left(\Delta\varphi_{A_j}[i] + \Delta N_{A_j}^l[i]\right) \\ \Delta d_{C_j}^l[i] = \left\|\boldsymbol{X}_{A_j}^l[i+1] - \boldsymbol{X}_T^l\right\| - \left\|\boldsymbol{X}_{A_j}^l[i] - \boldsymbol{X}_T^l\right\| \\ \boldsymbol{X}_{A_j}^l[i] = \Delta\boldsymbol{X}_{A_{ref}}[j] + \gamma^l \times (\boldsymbol{X}_C[i] - \boldsymbol{X}_C[1]) \end{cases} \tag{4.23}$$

Due to $\left|\Delta d_{A_j}^l[i]\right| < \frac{\lambda}{4}$, $\Delta N_A^l[i]$ can be determined according to Eq. 4.6. And $\Delta\boldsymbol{X}_{A_{ref}}[j]$ is the displacement of an antenna A_j relative to a reference one, which can be manually measured in advance. The nonlinear optimization problem is solved by Levenberg-Marquardt algorithm [67]. Once the estimates of tag position and scale factor are updated, we repeat the optimization from step 2 until it converges. The optimized RFID tag position $\widehat{\boldsymbol{X}}_T^*$ in the global world coordinate system can be calculated according to Eq. 4.19.

Analysis: The pseudocode of our joint optimization algorithm is presented in Algorithm 1. For a non-linear optimization problem, a good initial guess closer to the ground truth can achieve faster convergence so that the tag position refinement can speed up the optimization process. Given $\Delta\alpha_{opt} = 1^\circ$, $\Delta\beta_{opt} = 1^\circ$ and $\mu = 20^\circ$ in our experiment, the number of updating the scale factor and tag position is 4 times on average, taking the average of 16 milliseconds in our platform. Also, the proposed algorithm just provides a local optimum in each antenna array and can not achieve a global optimum over all antenna arrays. Due to adjacent antenna arrays with $(L-1)$ overlapping samples, our algorithm can still achieve high-precision estimation results shown in the following experiments.

4.5 Optimal Tag Position and Scale Factor Selection from Multiple Antenna Arrays

As the cart moves across an RFID tag, our system will produce a series of antenna arrays and output the corresponding tag positions. As shown in Fig. 4.6, when the average distance from the tag to each antenna element becomes larger and/or the

Algorithm 1 Coarse-to-Fine Tag Position and Scale Factor Optimization Algorithm

α_A^1, β_A^1, γ^1 ▷ Initial DOA and scale factor estimated from Spatial Power Spectrum
$\xi \leftarrow 10^{-6}$ ▷ Stopping threshold
$\rho^1 \leftarrow 0.01$ ▷ Initial damping factor
$\mathbb{L} \leftarrow 30$ ▷ Maximum loop count
$\mathbb{N} \leftarrow M \times (L-1)$ ▷ M: Number of reader antennas, L: Number of RFID-CV fusion data in an antenna array
$\mathbb{J}^1 \leftarrow 1$ ▷ Select nonlinear optimization algorithm, 0: Gradient-Descent, 1: Gauss-Newton
for $l \leftarrow 1 : \mathbb{L}$ **do**
  if $\mathbb{J} = 1$ **then**
    $\left(\alpha_A^{l+1}, \beta_A^{l+1}\right) \leftarrow DOARefinement\left(\alpha_A^l, \beta_A^l, \gamma^l\right)$
    $\boldsymbol{X}_T^l \leftarrow 3DRFIDLocalization\left(\alpha_A^{l+1}, \beta_A^{l+1}, \gamma^l\right)$
    $\left(\Delta \boldsymbol{D}_A^l\right)_{1\times\mathbb{N}} \leftarrow \frac{\lambda}{4\pi}\left(\Delta\varphi_A + \Delta N_A^l\right)$ ▷ Update distance difference matrix based on phase measurements
    $\left(\Delta \boldsymbol{D}_C^l\right)_{1\times\mathbb{N}} \leftarrow \left\|\boldsymbol{X}_A^l[2:L] - \boldsymbol{X}_T^l\right\| - \left\|\boldsymbol{X}_A^l[1:L-1] - \boldsymbol{X}_T^l\right\|$ ▷ Update distance difference matrix based on tag position and scale factor
    $\boldsymbol{H}_{4\times4}^l \leftarrow \left(\boldsymbol{J}_{\mathbb{N}\times1}^l\right)^{\mathrm{T}} \boldsymbol{J}_{\mathbb{N}\times1}^l$ ▷ J is the Jacobian matrix of $\Delta \boldsymbol{D}_C^l$
    if $l = 1$ **then**
      $e^1 \leftarrow \|\Delta \boldsymbol{D}_A^1 - \Delta \boldsymbol{D}_C^1\|$ ▷ Initial distance error
    end if
  end if
  $\widetilde{\boldsymbol{H}}_{4\times4}^l \leftarrow \boldsymbol{H}_{4\times4}^l + \rho^l \times \boldsymbol{I}$ ▷ Update Hessian matrix, $\boldsymbol{I}$: 4×4 identity matrix
  $\boldsymbol{h}_{1\times4}^l \leftarrow -\widetilde{\boldsymbol{H}}^l \times \left[\left(\boldsymbol{J}^l\right)^{\mathrm{T}}\left(\Delta \boldsymbol{D}_A^l - \Delta \boldsymbol{D}_C^l\right)\right]$ ▷ Update l^{th} descent direction
  $\boldsymbol{X}_T^{l+1} \leftarrow \boldsymbol{X}_T^l + \boldsymbol{h}^l[1:3]$ ▷ Update tag position
  $\gamma^{l+1} \leftarrow \gamma^l + \boldsymbol{h}^l[4]$ ▷ Update scale factor
  $e^{l+1} \leftarrow \left\|\Delta \boldsymbol{D}_A^{l+1} - \Delta \boldsymbol{D}_C^{l+1}\right\|$ ▷ Update distance error
  if $e^{l+1} < e^l$ **then**
    if $e^{l+1} \leqslant \xi$ **then**
      return;
    else
      $\boldsymbol{X}_T^{l+1} \leftarrow \boldsymbol{X}_T^l$, $\gamma^{l+1} \leftarrow \gamma^l$
      $\rho^{l+1} \leftarrow \rho^l/10$, $\mathbb{J} \leftarrow 1$, $e^{l+1} \leftarrow e^l$
    end if
  else
    $\rho^{l+1} \leftarrow \rho^l \times 10$, $\mathbb{J} \leftarrow 0$
  end if
end for
return refined scale factor γ^* and tag position $\boldsymbol{X}_T^*$.

average spacing between consecutive antenna elements becomes smaller, the size of the shaded region (called uncertainty region) will become larger, and the tag

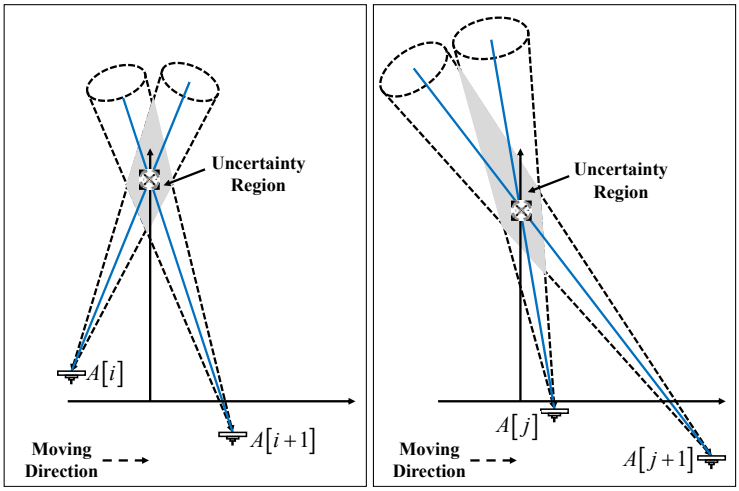


Figure 4.6 : Uncertainty region of RFID tag localization.

localization accuracy will be reduced (without considering the measurement error in RF phase). Inspired by this observation, the concept of HDOP to measure the relative satellite-receiver geometry in GPS [55] is introduced into our system. By measuring the size of the uncertainty region, we can select the optimal tag position and scale factor.

According to the definition of HDOP, we linearize a tag-to-antenna distance $d_A[i]$ in the local coordinate system by expanding a Taylor series at the calculated tag coordinate $\boldsymbol{X}_T^*$. Ignoring second and higher order terms, we have

$$\begin{aligned} d_A[i] \approx & \frac{\partial d_A[i]}{\partial x}\Delta x_T + \frac{\partial d_A[i]}{\partial y}\Delta y_T + \\ & \frac{\partial d_A[i]}{\partial z}\Delta z_T + d_A^*[i] \end{aligned} \tag{4.24}$$

where $\Delta \boldsymbol{X}_T = (\Delta x_T, \Delta y_T, \Delta z_T)$ represents the tag position's measurement error relative to the ground truth. The distance $d_A^*[i]$ of the estimated tag position to the i^{th} antenna element is

$$d_A^*[i] = \|\gamma^* \times (\boldsymbol{X}_C[i] - \boldsymbol{X}_C[1]) - \boldsymbol{X}_T^*\| \tag{4.25}$$

And $\frac{\partial d_A[i]}{\partial \boldsymbol{X}}$ represents the first partial derivatives, i.e.,

$$\begin{cases} \dfrac{\partial d_A[i]}{\partial x} = \dfrac{x_A[i] - x_T^*}{d_A^*[i]} \\ \dfrac{\partial d_A[i]}{\partial y} = \dfrac{y_A[i] - y_T^*}{d_A^*[i]} \\ \dfrac{\partial d_A[i]}{\partial z} = \dfrac{z_A[i] - z_T^*}{d_A^*[i]} \end{cases} \tag{4.26}$$

Given the number of M antennas, we group $M \times L$ equations together and represent them in matrix form,

$$\boldsymbol{D} = \boldsymbol{G}\Delta \boldsymbol{X}_T + \widetilde{\boldsymbol{D}} \tag{4.27}$$

Since the antennas move along the horizontal plane, we mainly focus on horizontal position errors. $\boldsymbol{G}$ is a matrix of the partial derivatives in X-Y dimension with $M \times L$ rows and 2 columns,

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_{A_1} & \cdots & \boldsymbol{G}_{A_M} \end{bmatrix}^T \tag{4.28}$$

and the matrix corresponding to the j^{th} reader antenna is

$$\boldsymbol{G}_{A_j} = \begin{bmatrix} \frac{\partial d_{A_j}[1]}{\partial x} & \frac{\partial d_{A_j}[1]}{\partial y} \\ \vdots & \vdots \\ \frac{\partial d_{A_j}[L]}{\partial x} & \frac{\partial d_{A_j}[L]}{\partial y} \end{bmatrix} \tag{4.29}$$

We construct the covariance matrix Q for localization error analysis,

$$\boldsymbol{Q} = \left[\boldsymbol{G}^T \boldsymbol{G}\right]^{-1} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} \tag{4.30}$$

Since σ_x^2 and σ_y^2 are the variances of X and Y-axis components of the tag position estimates, HDOP is

$$HDOP = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{4.31}$$

Lower HDOP means better tag localization accuracy due to strong tag-antenna geometry. We select the estimated tag position in the antenna array with the minimum HDOP as the optimal one. Since there might be many RFID tags to be located, our system will output multiple scale-factor candidates in an antenna array. The optimal scale factor corresponds to the RFID tag with the minimum HDOP. Additionally, we need to indicate that given a more accurate initial guess of tag position, the proposed coarse-to-fine optimization algorithm can output a more accurate scale factor.

4.6 Implementation

Hardware: The settings of an Impinj R420 RFID reader without any hardware or firmware modification are shown in Table. 4.2. Since no reader collision interference exists, the 'Max Throughput' reading mode is selected to provide the highest sampling rate. The reader works in the operating frequency band of 920∼926 MHz with 500 kHz channel spacing. Two 8dBi circular polarization antennas with about 6∼10m reading range connect to the reader. Dual-dipole Impinj H47 battery-free RFID tags are used by default, which can minimize tag-antenna orientation sensitivity. Single-dipole Impinj E51 and Alien AZ-9640 RFID tags are also used to evaluate system performance. Also, we configure a Microsoft Kinect V2 at the frame rate of 30 FPS to capture images. *Note that we only input 2D images into RF-MVO. The depth information of Kinect V2 is used to calculate the ground-truth camera trajectory.* These devices are all deployed on a mobile utility cart.

Software: According to Impinj LLRP Toolkit (LTK) [2], we program an RFID data collection application in C#. Each RFID data contains EPC, RF phase, antenna port number, operating frequency and reading timestamp. We adopt the

Table 4.2 : Impinj R420 RFID Reader Settings

Settings	**Values**
Operating Frequency	All Channels
Reader Mode	Max Throughput
Search Mode	Dual Target
Session	Session 0
RF Phase Reporting	Enabled
LastSeen Timestamp	Enabled

phase calibration mechanism [105] to eliminate the effect of frequency hopping on RF phase. Also, ORB-SLAM2 is a real-time simultaneous localization and mapping (SLAM) system to estimate camera trajectory and 3D reconstruction for monocular, stereo and RGB-D cameras [3, 13]. It can achieve loop closure detection and camera re-localization, which is of essential importance in visual SLAM systems to reduce accumulated errors over time. We run the RGB-D component on Ubuntu and save the camera trajectory as ground truth. However, ORB-SLAM2 can only save keyframe camera poses for a monocular camera rather than all frames. Hence, we record a video stream to a file by running ROS tool [4] and then calculate the 2D monocular poses (i.e., position and orientation) in Matlab [1] by analyzing these 2D images. Finally, the system performance of tag localization and trajectory recovery are all evaluated in Matlab, running on our laptop with 2.3 GHz CPU (Intel Core i5-6200U) and 4G memory.

Sampling Synchronization: According to Gen2 standard [33], RFID readers rely on a slotted-aloha access scheme to randomly read RFID tags. The sampling time between consecutive inventories of the same tag is determined by reader settings (i.e., reader mode, search mode, and session), tag population, and environment interference. We synchronize the reader clock with an Internet time server and then match RFID data to 2D images by minimizing the sampling time difference like [25].

Antenna Position Calibration. To evaluate our system's localization performance, we use Kinect V2 to capture the ground-truth tag position. However, recall from our system, the estimated tag positions are located in the antenna-centered coordinate system rather than in the camera-centered one. In this case, we need to conduct a coordinate system transformation. Due to an unknown camera view angle relative to each reader antenna, it is unable to directly measure antenna positions in the camera-centered coordinate system. In CV field, camera calibration is a fundamental step to estimate camera parameters and remove lens distortion. The commonly-used camera calibration method [116] performs with many 2D images of an asymmetric checkerboard with black and white squares under different positions and angles relative to the fixed camera. We calculate the coordinates of checkerboard corners in the camera-centered coordinate system and manually measure the distance of each corner to the casing surface center of each reader antenna. By building more than three sphere equations, we rely on least squares method to obtain the optimal antenna position. *Note that measuring tag-to-antenna distance based on*

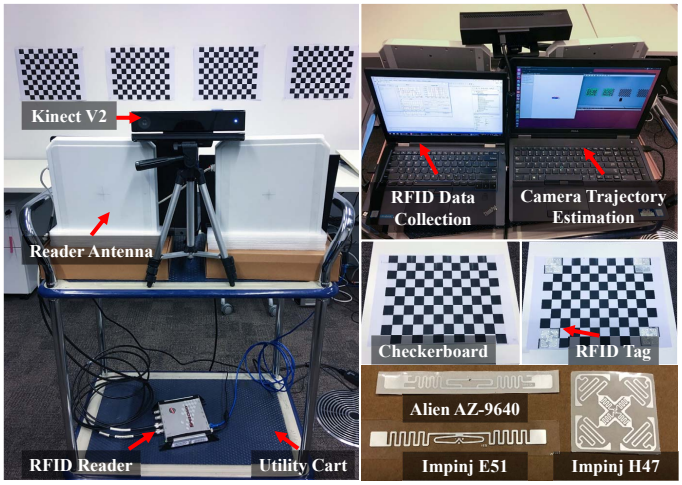


Figure 4.7 : Experiment setup.

RF phase is an effective way to reduce time cost and minimize measurement error in manual calibration, which is not discussed here due to the space limit.

4.7 Evaluation

In this section, we first introduce the experiment setup and metrics, followed by the detailed experiment results.

4.7.1 Experiment setup

Methodology: The experiment setup is shown in Fig. 4.7. We print the same checkerboards for camera calibration on both sides of a work, where the size of each square is 2.9 cm. Four double-sided checkerboards are affixed on the flat wood wall, two of which have 4 RFID tags (dual-dipole Impinj H47 RFID tags by default) to be located at specified corners on its back. The cart moves across the checkerboards with a speed of about 20 cm/s∼35 cm/s. The length of each cart movement trajectory along the x-axis (referring to the xyz axes in Fig. 4.1) is about 2.3 m. The distances of the wall to reader antennas vary within about 0.7 m∼1 m. Since the cart moves on the floor, the distance deviation along the z-axis is relatively small. The default algorithm parameters are: the number of antenna elements in an antenna array is $L = 60$; the spacings for initial DOA and scale-factor searching are $\Delta\alpha = 5°$, $\Delta\beta = 5°$ and $\Delta\gamma = 0.1$; the searching granularities for DOA refinement are $\Delta\alpha_{opt} = 1°$ and $\Delta\beta_{opt} = 1°$; the searching threshold is $\mu = 20°$.

Metrics: To verify RF-MVO performance, we focus on the errors in the tag localization and scale-factor estimation by comparing the estimates with ground-truth. To evaluate HDOP performance, we refer to the average distance of an RFID tag to antenna elements in an antenna array. Recall from Section 4.5, when an antenna moves close to an RFID tag, the average of tag-to-antenna distances is lower, which means the size of the tag position uncertainty region is smaller. Due to stronger tag-antenna geometry, the HDOP value is lower, and the corresponding tag localization accuracy is higher. In contrast, when the antenna becomes far away

from the tag, the HDOP value is higher, and the tag localization accuracy is lower.

1) Tag Position Error. In the i^{th} antenna array (corresponding to the i^{th} image), the ground-truth position $\mathbb{X}_T[i]$ of an RFID tag attached on the specified corner can be measured in the world coordinate system whose origin is at the RGB camera center of Kinect V2. Given the antenna-to-camera displacement $\boldsymbol{\Delta X}_{A,C}$ and the optimized tag position $\boldsymbol{X}_T^*[i]$, the tag position error is

$$T_{err}[i] = \|\mathbb{X}_T[i] - (\boldsymbol{\Delta X}_{A,C} + \boldsymbol{X}_T^*[i])\| \tag{4.32}$$

2) Scale-factor Error. Let $\mathbb{X}_C[i]$ and $\mathbb{X}_C[i+1]$ be the ground-truth camera positions corresponding to the i^{th} and $(i+1)^{th}$ 2D images measured by Kinect V2-powered ORB-SLAM2 in the real world view. Recall that $\boldsymbol{X}_C[i,1]$ and $\boldsymbol{X}_C[i,2]$ are the first two camera positions output by MVO technique in the camera view in the i^{th} antenna array, then the ground-truth scale factor $\Gamma[i]$ is calculated by

$$\Gamma[i] = \frac{\mathbb{X}_C[i+1] - \mathbb{X}_C[i]}{\boldsymbol{X}_C[i,2] - \boldsymbol{X}_C[i,1]} \tag{4.33}$$

And the scale-factor error $F_{err}[i]$ is

$$F_{err}[i] = |\Gamma[i] - \gamma^*[i]| \tag{4.34}$$

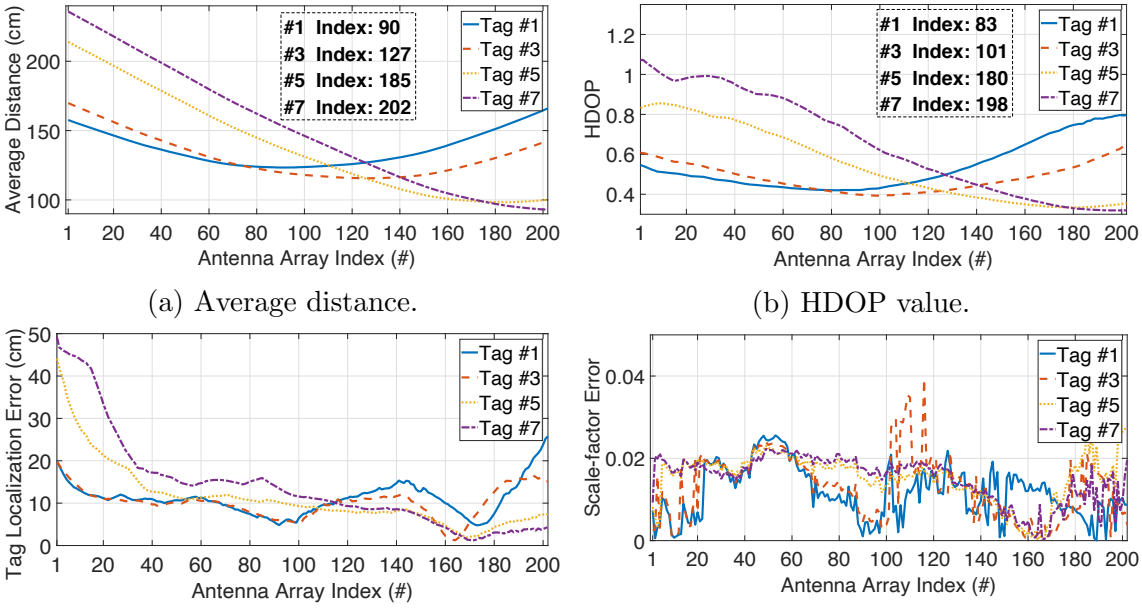
3) Average Distance. In the camera-centered world coordinate system, the i^{th} antenna position is $(\mathbb{X}_C[i] - \boldsymbol{\Delta X}_{A,C})$ while the tag position is $(\mathbb{X}_T[i] - \boldsymbol{\Delta X}_{A,C})$. Thus, we can calculate the average distance $D[i]$ by

$$D[i] = \frac{1}{L}\sum_{l=1}^{L} \|\mathbb{X}_T[i+l-1] - \mathbb{X}_C[i+l-1] - 2\boldsymbol{\Delta X}_{A,C})\| \tag{4.35}$$

4.7.2 HDOP Performance

In this experiment, we evaluate the impact of tag-antenna geometry. We move the cart at almost the same speed across RFID tags to make the displacement between consecutive antennas close to each other. Fig. 4.8a and Fig. 4.8b show that the trend between the average distance and HDOP curves is basically the same and the antenna array indexes with the minimum value are also close to each other. This experiment demonstrates the proposed HDOP is an effective indicator for tag localization error. However, Fig. 4.8c shows that the lowest HDOP value does not always mean the highest localization accuracy because the error level in camera pose estimation and RF phase measurement determines the final localization accuracy.

Also, Fig. 4.8d shows that when the tag localization result is close to the ground truth, our joint optimization algorithm can achieve better scale-factor estimation accuracy due to better initial guesses. We can see that the tag-antenna geometry has little impact on scale-factor error, and the overall error keeps at a very low level. As long as we provide enough RFID and CV fusion samples for the proposed spatial power spectrum, our system can output a high-precision scale factor in each antenna array.



(a) Average distance.

(b) HDOP value.

(c) Tag position error.

(d) Scale-factor error.

Figure 4.8 : HDOP performance

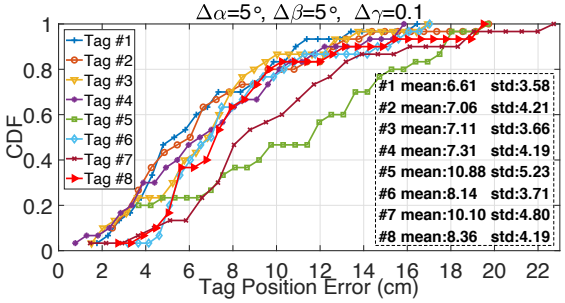
Table 4.3 : Tradeoff between accuracy and latency

Angular Granularity	Scale-factor Granularity	Position Error	Scale-factor Error	Latency
1°	0.1	7.769 cm	0.01594	20.722 s
2°	0.1	7.707 cm	0.01599	5.962 s
10°	0.1	8.047 cm	0.01636	0.629 s
5°	0.05	5.752 cm	0.01486	2.549 s
5°	**0.1**	**7.737 cm**	**0.01595**	**1.376 s**
5°	0.2	9.875 cm	0.01935	0.787 s

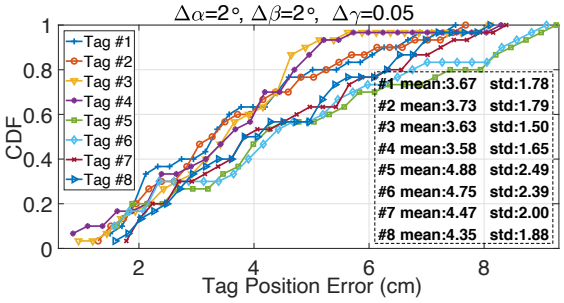
4.7.3 Tag localization performance

We move reader antennas across the RFID tags with different speeds and trajectories each time. For each RFID tag, the tag position with minimum HDOP value over all antenna arrays is selected as the optimal one. We repeat the experiment 30 times. At first, we perform with exhaustive searching (i.e., RF-MVO without optimization) under the spacings of $\Delta\alpha = 5°$, $\Delta\beta = 5°$, $\Delta\gamma = 0.1$ and $\Delta\alpha = 2°$, $\Delta\beta = 2°$, $\Delta\gamma = 0.05$, respectively. Fig. 4.9a and Fig. 4.9b show that the average tag position errors are 8.198 cm and 4.1339 cm in xyz combined dimension. With smaller searching granularities, exhaustive searching can improve estimation accuracy while it will consume much more computations. We discuss the impact of the searching granularities in detail in Section 4.7.5.

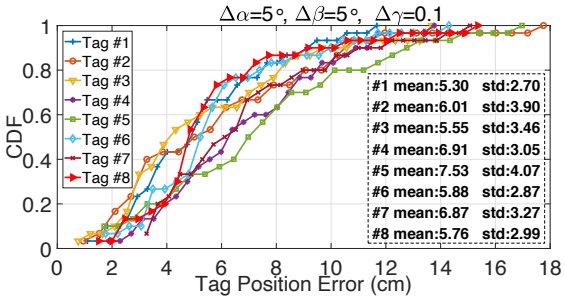
Fig. 4.9c shows that the overall average tag position error of RF-MVO under the same searching granularities drops from 8.198 cm to 6.23 cm in xyz combined dimension after optimization. When we exploit RF-MVO to detect mis-shelved books in a dense book library environment, the improvement of about 2cm is meaningful



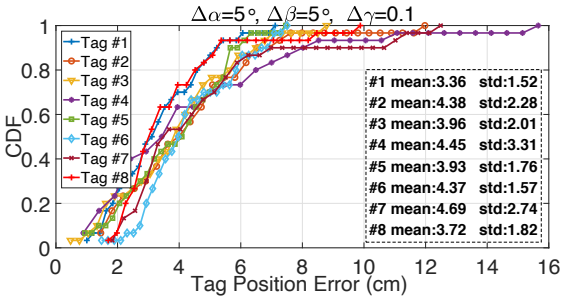
(a) Exhaustive searching with large granularities.



(b) Exhaustive searching with small granularities.



(c) RF-MVO with estimated trajectory.



(d) RF-MVO with actual trajectory.

Figure 4.9 : 3D tag localization performance.

(See Section 4.7.6).

In RF-MVO, the tag localization performance is subject to the scale-factor estimation accuracy. To purely compare RF-MVO's localization performance with previous methods, i.e., Tagoram [110] and MobiTagbot [79], we input the real-world camera trajectory into RF-MVO. Fig. 4.9d shows that RF-MVO achieves the overall average localization error of 4.1 cm, which can achieve 1.52 times accuracy improvement than the case with the estimated scale factors. In Tagoram, an RFID tag is required to move along an already-known trajectory. To match our experiment, we view the tag is fixed at the initial position and the antennas move in the opposite direction. Its enhanced version MobiTagbot has a stronger capability of multipath suppression than Tagoram by exploiting frequency hopping technique. To perform Tagoram and MobiTagbot, we partition the pre-specified region in 3D space into the 60×60×25 cuboids with a width of 1cm around the ground-truth tag positions. Fig. 4.10 shows the overall localization accuracy comparison. Tagoram and MobiTagbot achieve average localization errors of 5.41 cm and 3.64 cm, respectively. Compared to our method, the existing methods need to pre-specify the surveillance region of interest where a target RFID tag is the most likely to exist. And as the region of interest increases, huge searching computations will seriously affect real-time performance.

Fig. 4.11 shows the overall tag position error of 2.322 cm, 3.919 cm and 2.995 cm in x-axis, y-axis, and z-axis with optimization and the estimated trajectory. The localization error along the x-axis is lower than that along the y-axis and z-axis because the tag-antenna geometry along the x-axis in each antenna array is better

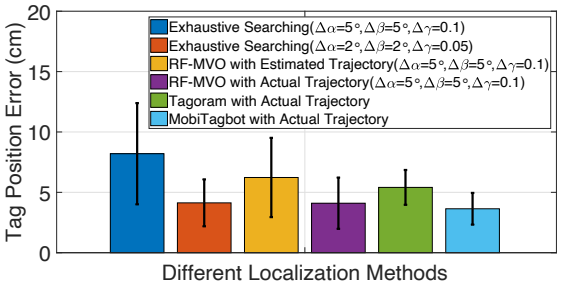


Figure 4.10 : Tag position error comparison.

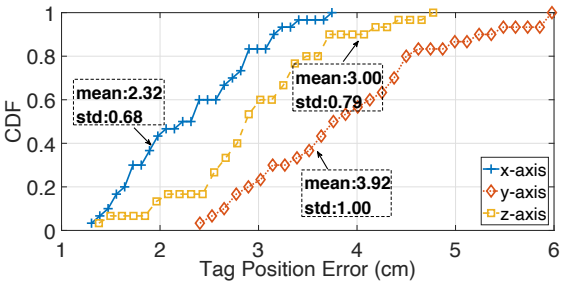


Figure 4.11 : Tag position error in x-axis, y-axis and z-axis.

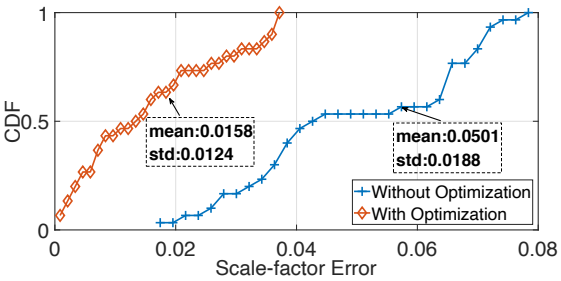


Figure 4.12 : The CDF of scale-factor error.

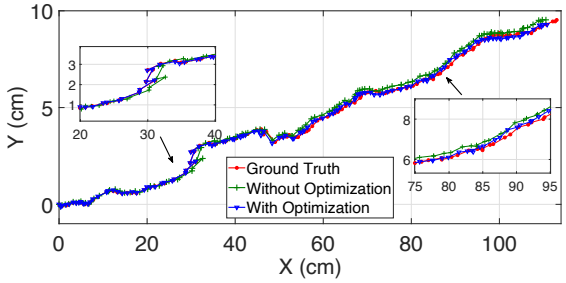


Figure 4.13 : A example of camera trajectory recovery.

than other axes.

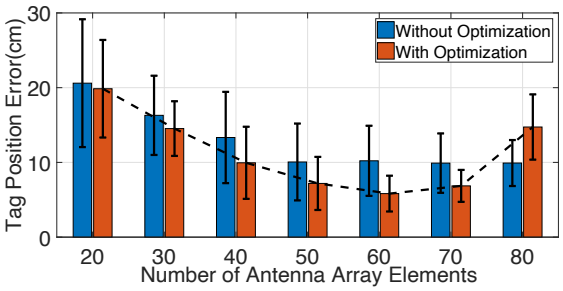
4.7.4 Trajectory recovery performance

We further verify the scale-factor estimation accuracy. Since eight RFID tags may produce multiple candidates, the scale factor with the minimum HDOP in each antenna array is selected as the optimal one. Fig. 4.12 plots the CDF of scale-factor error. Without optimization, the average estimation error is 0.0501, with a standard deviation of 0.0188. After applying the joint optimization, the average is reduced by 3.17 times, down to 0.0158 with a standard deviation of 0.0124. To better show the trajectory recovery performance, we choose an experiment data and plot the corresponding camera trajectory in Fig. 4.13. The estimated trajectory with optimization can better match the ground truth. However, estimating the camera poses and scale factors will inevitably generate accumulated errors over time, making the trajectory increasingly deviate from the ground truth.

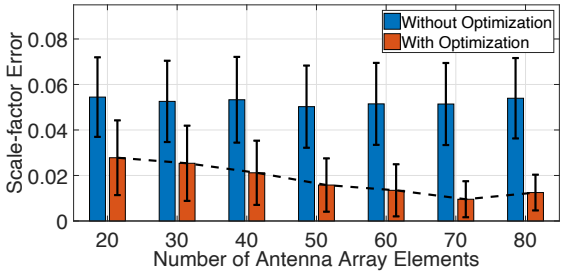
4.7.5 Microbenchmarks

Next, we evaluate RF-MVO performance with different experiment settings.

Impact of array element size. Given a utility cart trajectory, we vary the number of elements in an antenna array from 20 to 80 at 10 steps. Fig. 4.14a and Fig. 4.14b show that more antenna elements can effectively reduce the estimation errors when the element size increases from 20 to 60 (70 in Fig. 4.14b). When an antenna element is involved in more antenna arrays, the estimates of incident angles also become more accurate. However, when the element size increases from 60 (70

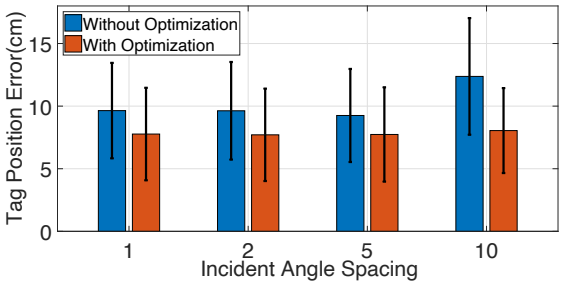


(a) Tag Position error.

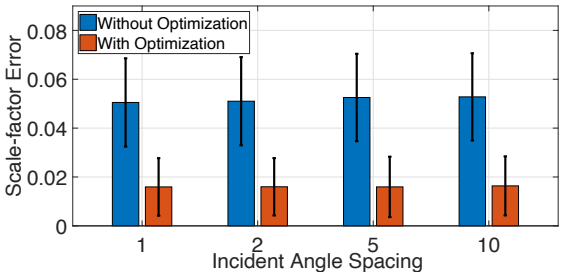


(b) Scale factor error.

Figure 4.14 : Array element size.



(a) Tag Position error.



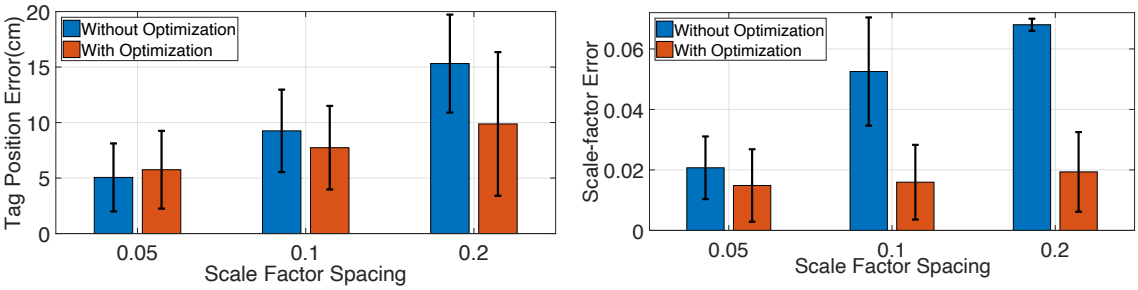
(b) Scale factor error.

Figure 4.15 : Angular spacing.

in Fig. 4.14b) to 80, increasingly accumulated errors in camera pose estimation will reduce our system's estimation accuracy. We suggest setting the size of about 60∼70 elements in an antenna array.

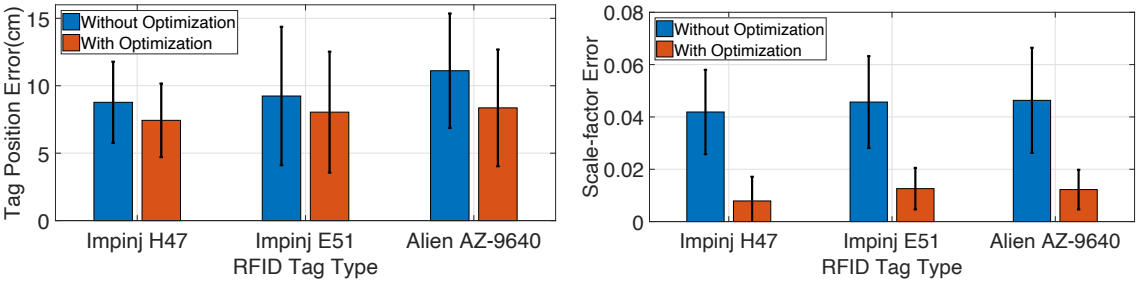
Impact of angular and scale-factor spacings. At first, we vary the DOA searching spacing at 1°, 2°, 5° and 10°, respectively, while fixing the scale-factor spacing at 0.1. Fig. 4.15a shows the tag localization error at the angular granularity of 10° without optimization is much larger than other cases. However, Fig. 4.15b shows it has very little impact on scale-factor estimation performance. We consider that the proposed spatial power spectrum can find out the optimal scale factor if we could provide enough samples for our system. After optimization, the estimation errors are very similar to each other, which is mainly attributed to DOA refinement in the coarse-to-fine optimization. However, the larger angular spacing requires to set a larger search threshold μ, which will inevitably increase the computational load.

Then we vary the scale-factor spacing at 0.05, 0.1, 0.2, respectively, while the angular spacing is fixed at 5°. Fig. 4.16a and Fig. 4.16b show that the smaller scale-factor spacing is helpful for improving estimation accuracy. However, it will inevitably incur much more computations. Table. 4.3 shows a tradeoff between estimation accuracy and response latency. RF-MVO under the angular and scale-factor spacings of 0.05 and 5° will take the runtime of 2.357s and 0.192s (2.549s in total) on our experiment platform for searching and optimization, while the estimation accuracy is higher than other cases. In our experiment, the scale-factor and angular spacings are set to 0.1 and 5° to reduce response latency.



(a) Tag Position error. (b) Scale factor error.

Figure 4.16 : Scale-factor spacing.



(a) Tag Position error. (b) Scale factor error.

Figure 4.17 : RFID tag type.

Impact of RFID tag type. In general, there are two types of RFID tag antennas, i.e., single-dipole and dual-dipole. We use dual-dipole Impinj H47, single-dipole Impinj E51, and single-dipole Alien AZ-9640 to study the impact of different RFID tag types. Fig. 4.17a and Fig. 4.17b show that dual-dipole RFID tag can achieve a little higher accuracy of tag localization and scale-factor estimation than single-dipole tags. The weaker tag backscatter signal power results in a lower tag sampling rate, thereby degrading our system performance. A single-dipole RFID tag is more sensitive to the change in tag-antenna orientation than a dual-dipole tag. Tag-antenna orientation mismatch can reduce the backscatter power observed at the reader antenna. In the extreme case, the single-dipole tag oriented along the electric field from the reader antenna may receive the maximum RF signal while the tag oriented perpendicular to the field may receive no RF signal at all, which can not be read. Fortunately, the dual-dipole tag can efficiently minimize tag-antenna orientation sensitivity.

Impact of Multipath Interference. In actual environments, the signal emitted from a reader antenna will be reflected off the floor, walls, furniture and other objects besides an RFID tag back to the antenna, in addition to the expected tag backscatter signal. These reflected signals will induce a change in the measured RF phase. This experiment is to evaluate the impact of multipath interference on localization accuracy. Two checkerboards with eight RFID tags are attached to a door. To simulate a rich-multipath scenario, we ask two people to move behind the door randomly. The distance of the door to the moving persons is about 1 m$\sim$1.5 m, while the tag-to-antenna distance is about 1 m. *Note that the moving*

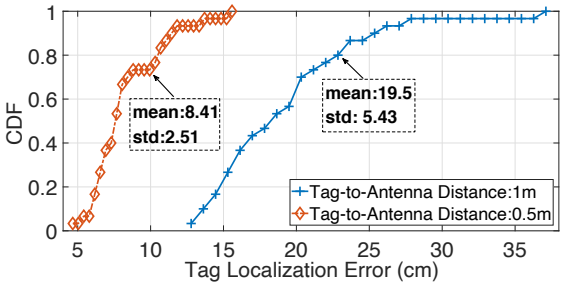


Figure 4.18 : Multipath interference.

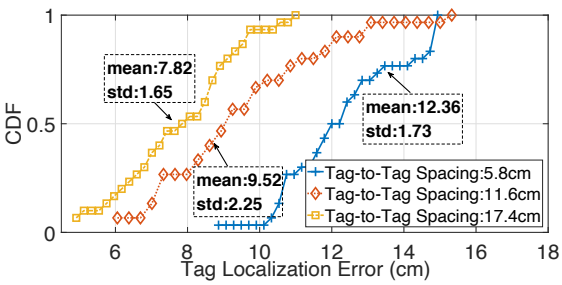


Figure 4.19 : Tag mutual coupling.

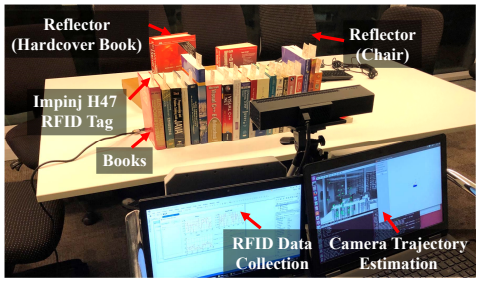


Figure 4.20 : Experiment setup.

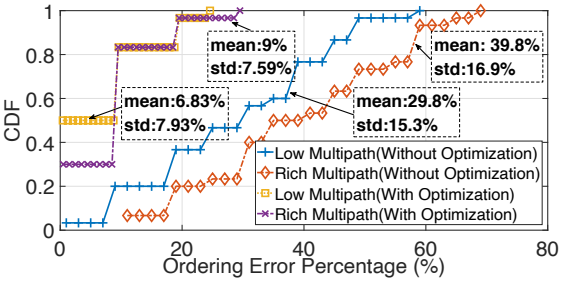


Figure 4.21 : Ordering error.

persons can not appear in the camera view because the ORB-SLAM2 system only performs well in a static environment. Many existing VO studies [120, 53] can deal with motion interference in a dynamic environment by segmenting these moving objects. However, this is beyond the scope of our work. We repeat this experiment 30 times. Fig. 4.18 shows the average localization error under the tag-to-antenna distance of 1m is about 19.50 cm. Compared to previous cases in the absence of dynamic interference (about 6.23cm localization error), the error is increased by more than 3 times. Fortunately, when we shorten the tag-to-antenna distance to about 0.5 m, Fig. 4.18 shows the error is reduced to about 8.41 cm. Since the shorter tag-to-antenna distance will increase the strength of the tag backscatter signal at the antenna receiver, the effect of dynamic interference will accordingly decrease.

Impact of Tag Mutual Coupling. In a dense RFID tag deployment, a responded RFID tag will combine the direct signal from a reader antenna to the tag with additional coupling signals emitted from other nearby RFID tags to produce its backscattered signal. This phenomenon is called tag mutual coupling effect, which will also affect the measured RF phase. In this experiment, six closely-spaced RFID tags are used. We vary the spacing between two adjacent tags, i.e., 5.8 cm, 11.6 cm, and 17.4 cm, to evaluate the impact of tag mutual coupling on localization accuracy. Fig. 4.19 shows as the tag-to-tag spacing increases, the average localization error is reduced from 12.36 cm to 7.82 cm. The closer two RFID tags are placed to each other, the larger the change in RF phase due to tag mutual coupling will be. In practical applications, we should consider the impact of tag mutual coupling.

4.7.6 Case Study for Mis-shelved Book Detection

Here we apply RF-MVO into a case study, i.e., a library scenario where the system detects RFID-tagged book order on a shelf to discover mis-shelved books. The experimental scenario is shown in Fig. 4.20. Twenty RFID-tagged books are placed on the desk for ordering. In a low-multipath scenario, no other reflectors are placed close to these books. In contrast, to generate a rich-multipath scenario, we place some reflectors on the desk and ask two people to randomly move around the RFID-tagged books (they do not occur in the camera view). The thickness of these books ranges from about 2cm to 5cm. The tag-to-antenna distance is about 60 cm∼80 cm. Here we need to indicate that if an RFID tag is directly attached to the spine of each hardcover book, the book surface may produce a strong reflected signal in the opposite direction to the incident signal. RFID-active power for tag demodulation will fall rather rapidly to 0, so RF phase follows the abrupt change in received power and the reading rate will decrease to a very low level. Instead, an RFID tag (i.e., dual-dipole Impinj H47) is clipped on a page in a book in our experiment. In practice, we suggest using an anti-metal RFID tag with a magnetic isolator sheet. We repeat the experiment 30 times. To evaluate book ordering performance, we define the ordering error, i.e., the number of RFID-tagged books ordered incorrectly out of the total book number. Recall from Section 4.3, since the average tag position error along the x-axis is much smaller than that along the y-axis and z-axis, we rely on Eq. (19) to obtain x-axis coordinates of these RFID tags in the global coordinate system for ordering. Fig. 4.21 shows that compared to the cases without optimization, RF-MVO can effectively lower ordering error, which is mainly attributed to the improvement of about 2cm in tag localization after optimization. And the average ordering error of RF-MVO sightly increases from 6.83% in a low-multipath scenario to 9% in a rich-multipath scenario. According to Section 4.7.5, we can further reduce tag-to-antenna distance to improve our system's capability of combating multipath interference.

In addition, recall from Section 4.3 that the average localization error in the x-axis is about 2cm in our default experiment setup. When two adjacent RFID-tagged items are separated by less than 2cm, it could be very challenging to correctly associate the collected EPCs with the corresponding items. To further improve localization accuracy, we can appropriately increase the element size in each antenna array (or deploy more reader antennas), reduce the scale-factor searching spacing, or shorten the tag-to-antenna distance according to our previous experiments. And the computational cost and real-time performance also need to be considered.

4.8 Conclusion

In this work, we present a CV-assisted RFID system to achieve stationary RFID localization in 3D space by introducing a 2D monocular camera. We fuse depth-enabled RF phase and a piece of camera trajectory in the camera view for RFID localization and real-world trajectory recovery. Novel algorithms are proposed to improve real-time performance and localization precision. Experimental results demonstrate the effectiveness of our hybrid system.

Chapter 5

CV-assisted Two-RFID-Tag Labeled Object 3D Orientation

State-of-the-art battery-free RFID systems can achieve orientation tracking by attaching multiple tags on an object. However, the tag signal fingerprints may be distorted by mutual coupling interference. Also, multipath interference in actual environments is another challenge for such systems. In this work, we propose RF-Orien3D that performs 3D orientation tracking in multipath-rich environments only using two tags and one fixed reader antenna. Since the mutual coupling changes the radiation pattern of each tag, the signal fingerprint carries orientation-dependent features. Our basic idea is to translate their fingerprints into spatial orientation. To achieve RF-Orien3D, we introduce additional multipath terms in the existing RSSI/phase-distance models under the coupling effect, where two variables need to be pre-estimated: one is tag radiation pattern, which is obtained by simulating a two-tag array from a 2D image; another is modulation factor, which is calculated using signal fingerprints in non-coupling and coupling cases. On this basis, we simulate all possible multipath impacts on signal fingerprints to pre-training a CNN, and then collect very few actual data to fine-tune the CNN for orientation tracking. The experiments show RF-Orien3D achieves median azimuth and elevation errors of about 29° and 11° in low/rich multipath scenarios.

5.1 Introduction

RFID is becoming a promising technology for wireless sensing, where an RFID tag is generally regarded as a battery-free sensor attached to an object. As the sensor indicator, i.e., backscatter signal fingerprint (i.e., RSSI and RF phase), varies with its distance to a reader antenna, many systems have been proposed for high-precision RFID localization [110, 65, 64, 106, 108]. In addition to location information, the orientation estimation of this RFID-tagged object is also of importance for Internet-of-Things applications. For example, a product (e.g., liquid) is packaged in a cardboard box marked "This Side Up" to call attention to its orientation in shipment and storage. Another is for a robotic pickup task. To achieve that, 6D pose (i.e., location and orientation) of an object is required to be estimated. In this case, the robotic arm can adjust its grip to grasp the target. This work is to design a ubiquitous solution for object orientation tracking with low-cost and battery-free RFID, even when the tracked object is blocked by some obstructions in an NLOS scenario.

Most computer vision-based solutions [52, 85, 73] require large-scale labeled image data. And it is very challenging to address the issues about light and occlusion

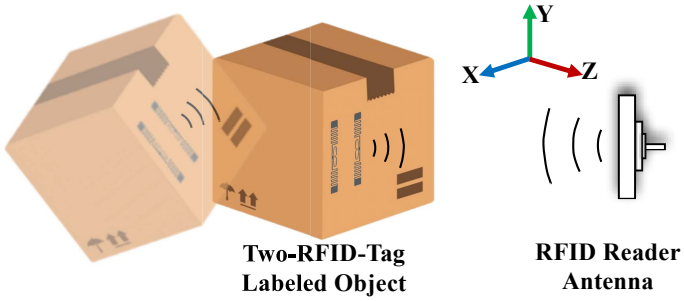


Figure 5.1 : Illustration of RF-Orien3D design. Due to the impact of tag mutual coupling in a two-RFID-tag array, each tag's original radiation pattern is modified, and their backscatter signal fingerprints carry orientation-dependent features, which can be captured by RF-Orien3D for spatial orientation estimation.

and achieve target-specified orientation tracking when multiple objects (e.g., cardboard box) with the same appearance appear in the camera view. Recent works have demonstrated the feasibility of RFID-based solutions. Some single-tag systems [59, 49] affix only one tag on an object and exploit linearly-polarized (for RSSI) or circularly-polarized (for RF phase) antennas to sense the variation of tag-to-antenna polarization for tracking. However, such systems require moving the tagged object along a known trajectory or rotating it at a constant speed, which may not be suitable for ubiquitous applications. Some multi-tag systems [105, 18] require attaching many tags on an object but also deploying multiple reader antennas around the region of interest. By calculating the spatial position relationship among these tags, they infer the tagged object orientation. However, mutual coupling interference [63, 24, 103] is inevitably induced in the multi-tag deployment, which will distort reported RFID fingerprints and thereby challenge orientation tracking accuracy.

In this work, we propose RF-Orien3D, a system that estimates the 3D orientation (i.e., azimuth and elevation) of a two-RFID-tag array relative to a reader antenna (shown in Fig. 5.1, even in multipath-rich scenarios. In the tag array, the tag spacing is fixed, and two tags are deployed in parallel. The system performs with the reported RSSI and RF phase of each tag from a COTS RFID reader. When tracking, it does not require moving a RFID-tagged object or reader antenna like previous works. The key novelty of this work is to leverage the "bad" mutual coupling effect between two tags. For an RFID tag located in space by itself, its radiation pattern (i.e., tag antenna gains in all directions) can be approximately symmetric around the tag. However, when another tag appears near the tag, its original radiation pattern will be changed. In some directions, the tag gains will be strengthened while weakened in others. Correspondingly, its reported signal fingerprint will also be changed. RF-Orien3D would estimate 3D orientation by analyzing the orientation-dependent signal fingerprints. To realize such a system, we need to deal with three challenges:

Challenge 1. Multipath Interference Suppression: Multipath interference is a fundamental challenge in wireless signal-based sensing works. Since the surrounding can reflect an emitted signal to an RFID tag and then reflect its modulated backscat-

ter signals back to an antenna receiver, these unexpected and unpredictable reflected signal may result in the distortion in RFID signal fingerprint. To combat multipath interference, we update the existing RSSI/RF phase-distance models under the mutual coupling effect [103] by introducing multipath items and then propose a convolutional neural network (CNN) based architecture with transfer learning technique. One component is to pre-train a CNN by simulating a large-scale dataset of the two-tag array in all possible multipath cases based on the updated RFID fingerprint models. Another component is to fine-tune the CNN by providing a small number of measured RFID fingerprints in actual scenarios. Due to the strong generalization ability in CNN, multipath interference could be significantly suppressed.

Challenge 2. Radiation Pattern Simulation: In our updated RFID fingerprint models, the modified radiation pattern of each tag is required to be pre-estimated. We firstly build an RFID tag model from an 2D image and then create a two-tag array for radiation pattern simulation in the case of tag impedance matching. In this way, their radiation patterns can be simulated for use.

Challenge 3. Scaled Modulation Factor Estimation: Another variable in the models to be determined is impedance-dependent modulation factor [36] (actually scaled by impedance-dependent tag antenna radiation efficiency), which also suffers from the mutual coupling effect. We design an estimation method only using RFID fingerprints of each tag in the non-coupling and coupling cases.

We build a prototype with an Impinj R420 RFID reader, a circularly-polarized 8dBi antenna and two Impinj E51 single-dipole tags. RF-Orien3D can estimate the spatial orientation of a tagged object, even in multipath-rich and/or NLOS environments. When tag separation distance is fixed at 4 cm in our experiment setup, RF-Orien3D can achieve median azimuth and elevation errors of about 29° and 11° in low or rich multipath cases.

Contributions: This work proposes RF-Orien3D, the first system that captures the change in radiation pattern caused by the mutual coupling effect for object orientation tracking in 3D space. Our design introduces three key innovations:

1) It introduces a CNN-based solution with transfer learning technique to deal with multipath interference, without manually collecting a large number of RFID signal fingerprints in different scenarios;

2) It presents a new way to understand the performance of an RFID tag array in the multi-tag based sensing research that simulates a tag array from a 2D image for tag characteristic analysis;

3) It proposes an approach to analyze the scaled modulation factor of the tag element with a slight impedance mismatch in practice in a tag array only using reported RFID fingerprints in non-coupling and coupling cases, without using specialized RF network analyzers.

5.2 Formulating Mutual Coupling Between Two RFID Tags in Multipath Environment

In realistic environments, the transmitted RF signal from a reader antenna will be reflected off some reflectors (such as tagged objects, walls, and furniture) to an RFID tag, in addition to the signal traveling along the direct tag-to-antenna path. This phenomenon is called *multipath effect*. And when the mutual coupling occurs, the tag will also receive some signals emitting from other nearby RFID tags. In this way, the tag combines all received signals to produce its backscatter signal, thereby sending back to the reader antenna. Compared to a tag located in free space by itself, the multipath and coupling effects both result in unexpected changes in RSSI and RF phase. In this section, we extend the mutual coupling models in [103] by introducing additional multipath interference.

In this work, we focus on the case that two RFID tags are attached to an object in parallel, where such parallel deployment could remove the polarization impact between each tag and a reader antenna. We use the reported RSSI and RF phase of a COTS RFID reader to represent the total channel characteristic between a transmitter and receiver. Let $\mathcal{R}^t$ and $\mathcal{R}^r_{T_1}$ be the transmitted and received power measurement in dBm (take a tag T_1 in the two-tag array for example), and let $\mathcal{P}_{T_1}$ be the RF phase measurement in radians. The total channel characteristic is

$$h^{dir}_{T_1}\left(1+h^{ref}_{T_1}+h^{mul}_{T_1}\right)=10^{\frac{\mathcal{R}^r_{T_1}-\mathcal{R}^t}{20}}e^{\boldsymbol{J}\left(\mathcal{P}'_{T_1}+\pi\boldsymbol{k}_{T_1}\right)} \tag{5.1}$$

where $\mathcal{P}'_{T_1}=\mathcal{P}_{T_1}\left(\mathrm{mod}\left[0,\pi\right)\right)$ and the function $\mathrm{mod}\left(\cdot\right)$ is the modulo operation, which is to eliminate $\boldsymbol{\pi}$-radians phase ambiguity caused by phase reversal-amplitude shift keying modulation. $\boldsymbol{k}_{T_1}$ is an unknown integer. $\boldsymbol{J}$ represent the imaginary unit. The complex exponential $h^{dir}_{T_1}$ is to represent the channel characteristic of the direct path between the reader antenna and the tag T_1; $h^{ref}_{T_1}$ is to denote that of the reflected path off its nearby tag T_2; and $h^{mul}_{T_1}$ is to represent that of the reflected path off its surrounding objects,

$$\begin{cases} h^{dir}_{T_1}=\tau\rho_l\rho^{\Delta}_{T_1}g^2_A g^2_{T_1}\left(\dfrac{\mu_{A,T_1}}{d_{T_1}}\right)^2 e^{-\boldsymbol{J}\left(\frac{4\pi}{\lambda}d_{T_1}+2\theta_{A,T_1}+\phi_l+\phi_{T_1}\right)} \\ h^{ref}_{T_1}=\dfrac{\rho^{\Delta}_{T_2}g_{T_2}g_{T_12}g_{T_21}d_{T_1}}{g_{T_1}\left(d_{T_2}+D\right)}e^{-\boldsymbol{J}\left[\frac{2\pi}{\lambda}\left(d_{T_2}+D-d_{T_1}\right)+\phi_{T_2}\right]} \\ h^{mul}_{T_1}=\displaystyle\sum_{i=1}^{N}\frac{\rho_{o_i}g_{T_1,o_i}d_{T_1}}{g_{T_1}d_{o_i}}e^{-\boldsymbol{J}\left[\frac{2\pi}{\lambda}\left(d_{o_i}-d_{T_1}\right)+\phi_{o_i}\right]} \end{cases} \tag{5.2}$$

where τ is a constant factor; ρ_l and ϕ_l represent the amplitude and phase shift of the leakage signal from the read antenna transmitter to its receiver; $\rho^{\Delta}_{T_1}$ (and $\rho^{\Delta}_{T_2}$) is the modulation factor, i.e., the difference in reflection coefficients when T_1 (and T_2) switches its chip impedance between the non-reflective and reflective states, which is dependent on tag type, operating frequency, and material property of a tagged object [36]; ϕ_{T_1} (and ϕ_{T_2}) is called tag phase shift determined by the propagation delay between the incident and reflected signals on the tag antenna of T_1 and T_2;

g_A is the voltage gain of the reader antenna; g_{T_1} and g_{T_2} are the voltage gains of the tag antennas in the directions of T_1 and T_2 relative to the reader antenna; $g_{T_{12}}$ (and $g_{T_{21}}$) is the voltage gain of the tag antenna in the direction of T_1 towards T_2 (and T_2 towards T_1); g_{T_1,o_i} is the voltage gain of the tag antenna in the direction of the i^{th} object; μ_{A,T_1} and θ_{A,T_1} are the polarization terms in amplitude and phase shift between the reader antenna and T_1; d_{T_1} and d_{T_2} are the distances of T_1 and T_2 to the reader antenna; d_{o_i} is the propagation distance of the reflected signal off the i^{th} object, traveling along the reader antenna to the object and then back to the RFID tag T_1; ρ_{o_i} and ϕ_{o_i} represent the reflection coefficient and phase shift of the i^{th} reflection object; N is the number of surrounding reflectors.

RSSI- and RF Phase-distance Models with Mutual Coupling and Multipath Effects. We extract the amplitude and phase shift from Eq. 5.1 to obtain RSSI- and RF phase-distance models,

$$\begin{cases} 10^{\frac{\mathcal{R}^r_{T_1}-\mathcal{R}^t}{20}} = \tau\rho_l\rho^{\Delta}_{T_1}g_A^2g_{T_1}^2\left(\dfrac{\mu_{A,T_1}}{d_{T_1}}\right)^2\Gamma_{T_1} \\ \mathcal{P}'_{T_1} = -\left(\dfrac{4\pi}{\lambda}d_{T_1} + 2\phi_{A,T_1} + \phi_{T_1} + \phi_l\right) + \Theta_{T_1}\,(\mathrm{mod}\,[0,\pi)) \end{cases} \tag{5.3}$$

where

$$\begin{cases} \Gamma_{T_1} = \left\|1 + h^{ref}_{T_1} + h^{mul}_{T_1}\right\| & \text{(5.4a)} \\ \Theta_{T_1} = \angle\left(1 + h^{ref}_{T_1} + h^{mul}_{T_1}\right) & \text{(5.4b)} \end{cases}$$

The function $\|\cdot\|$ is to calculate the modulus, and the function $\angle$ is to calculate the phase shift.

RSSI- and RF Phase-distance Models without Coupling and Multipath Effects. Compared to Eq. 5.3, the models free from the mutual coupling and multipath effect can be denoted as

$$\begin{cases} 10^{\frac{\overline{\mathcal{R}}^r_{T_1}-\mathcal{R}^t}{20}} = \tau\rho_l\overline{\rho}^{\Delta}_{T_1}g_A^2\overline{g}_{T_1}^2\left(\dfrac{\mu_{A,T_1}}{d_{T_1}}\right)^2 \\ \overline{\mathcal{P}}'_{T_1} = -\left(\dfrac{4\pi}{\lambda}d_{T_1} + 2\phi_{A,T_1} + \phi_{T_1} + \phi_l\right)(\mathrm{mod}\,[0,\pi)) \end{cases} \tag{5.5}$$

where $\left\langle\overline{\mathcal{R}}^r_{T_1}, \overline{\mathcal{P}}_{T_1}\right\rangle$ are the RSSI and RF phase in the non-coupling and non-multipath case. The modulation factor $\overline{\rho}^{\Delta}_{T_1}$ and tag gain $\overline{g}_{T_1}$ are different from $\rho^{\Delta}_{T_1}$ and g_{T_1} due to the coupling effect.

Based on above analysis, it is easy to derive the channel characteristic in the case where the responding tag is T_2. In particular, for the RFID tags with the same type, we assume that $\overline{\rho}^{\Delta}_{T_1} = \overline{\rho}^{\Delta}_{T_2}$, $\rho^{\Delta}_{T_1} = \rho^{\Delta}_{T_2}$, and $\phi_{T_1} = \phi_{T_2}$. In addition, once a circularly polarized antenna and single-dipole RFID tags are used, the amplitude polarization terms are constant [24], i.e., $\mu_{A,T_1} = \mu_{A,T_2}$ (*Note that the phase-shift polarization terms θ_{A,T_1} and θ_{A,T_2} changes with different directions with respect to the reader antenna.*). In the rest of this work, we use $\overline{\rho}^{\Delta}_{T}$, ρ^{Δ}_{T}, ϕ_T and $\mu_{A,T}$ to separately represent the corresponding variables for simplicity.

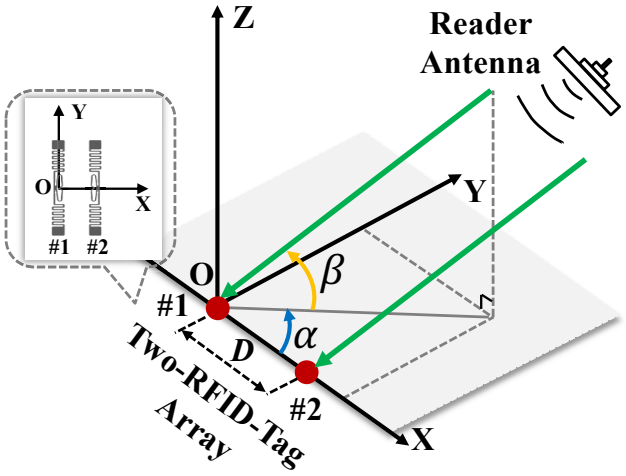


Figure 5.2 : Azimuth and elevation angles in tag-array-centric coordinate system.

5.3 Tracking Object Orientation in 3D Space

In this section, we firstly introduce a simple case without considering the multipath impact for orientation estimation, which is achieved by building an orientation spectrum. It could clearly explain the key principle of RF-Orien3D. After that, we consider the multipath effect and propose a CNN-based method to deal with it.

5.3.1 Orientation Estimation in Low Multipath

At first, we assume that transmitted signals from a reader antenna are radiated far enough from a two-tag array such that the received signals at each tag end exhibit a planar wavefront. As shown in Fig. 5.2, let α and β (in units of degrees) be the azimuth and elevation angles of the reader antenna relative to the two-tag array in the array-centric coordinate system, where $\alpha \in [0°, 180°]$ and $\beta \in [0°, 90°]$ due to the symmetrical radiation patterns (described in Section 5.4). The voltage gains of RFID tags in the direction of α and β can be denoted as $g_{T_1}(\alpha, \beta)$ and $g_{T_2}(\alpha, \beta)$.

In a very low-multipath environment, we ignore the effect of multipath interference and then set $h_{T_i}^{mul} = 0$ in the proposed model in Equation. 5.1. RF-Orien3D performs by building an orientation spectrum of which pixel indicates the likelihood of α and β being the ground truth by

$$\mathbb{P}(\alpha, \beta) = \left\| \frac{\Delta h_{T_1,T_2}^{dir}(\alpha, \beta)\, \Delta h_{T_1,T_2}^{ref}(\alpha, \beta)}{\Delta h_{T_1,T_2}} - 1 \right\| \tag{5.6}$$

where

$$\begin{cases} \Delta h^{dir}_{T_1,T_2}(\alpha,\beta) = \left[\dfrac{g_{T_1}(\alpha,\beta)}{g_{T_2}(\alpha,\beta)}\right]^2 e^{-\boldsymbol{J}\left(\frac{4\pi}{\lambda}D\cos\alpha\cos\beta\right)} \\ \Delta h^{ref}_{T_1,T_2}(\alpha,\beta) = \dfrac{1+h^{ref}_{T_1}(\alpha,\beta)}{1+h^{ref}_{T_2}(\alpha,\beta)} \\ h^{ref}_{T_1}(\alpha,\beta) = \rho^{\Delta}_{T} g_{T12} g_{T21} \dfrac{g_{T_2}(\alpha,\beta)}{g_{T_1}(\alpha,\beta)} e^{-\boldsymbol{J}\left[\frac{2\pi}{\lambda}D(1-\cos\alpha\cos\beta)+\phi_T\right]} \\ h^{ref}_{T_2}(\alpha,\beta) = \rho^{\Delta}_{T} g_{T12} g_{T21} \dfrac{g_{T_1}(\alpha,\beta)}{g_{T_2}(\alpha,\beta)} e^{-\boldsymbol{J}\left[\frac{2\pi}{\lambda}D(1+\cos\alpha\cos\beta)+\phi_T\right]} \\ \Delta h_{T_1,T_2} = 10^{\frac{\mathcal{R}^r_{T_1}-\mathcal{R}^r_{T_2}}{20}} e^{\boldsymbol{J}\left[\left(\mathcal{P}'_{T_1}-\mathcal{P}'_{T_2}\right)+\pi\left(\boldsymbol{k}_{T_1}-\boldsymbol{k}_{T_2}\right)\right]} \end{cases} \tag{5.7}$$

Among them, we need to indicate that:

1) Due to the parallel deployment of T_1 and T_2, they have the same polarizations, i.e., $\theta_{A,T_1} = \theta_{A,T_2}$, which are removed in Eq. 5.6.

2) The gains $g_{T_1}(\alpha,\beta)$, $g_{T_2}(\alpha,\beta)$ as well as g_{T_1,T_2} and g_{T_2,T_1} are obtained by simulating an RFID tag from an 2D image, which will be detailed in the following Section 5.4.

3) The details of how to estimate the tag phase shift ϕ_T and modulation factor ρ^{Δ}_{T} are described in the following Section 5.5.

4) When $D \ll d_{T_1}, d_{T_2}$, we ignore tag-to-antenna distance ratios in calculating $h^{ref}_{T_1}(\alpha,\beta)$ and $h^{ref}_{T_2}(\alpha,\beta)$.

5) According to direction of arrival estimation theory, the difference in the distances of the reader antenna to each RFID tag can be denoted by

$$d_{T_1} - d_{T_2} \approx D\cos\alpha\cos\beta \tag{5.8}$$

6) The integer difference $\Delta\boldsymbol{k} = \boldsymbol{k}_{T_1} - \boldsymbol{k}_{T_2}$ can be determined as follows. According to Eq. 5.4b, we calculate the difference in measured RF phase of T_1 and T_2,

$$\Delta\Lambda(\alpha,\beta) - \pi\Delta\boldsymbol{k} \approx \frac{4\pi}{\lambda}D\cos\alpha\cos\beta \tag{5.9}$$

where $\Delta\Lambda(\alpha,\beta) = \mathcal{P}'_{T_2} - \mathcal{P}'_{T_1} + \angle\Delta h^{ref}_{T_1,T_2}(\alpha,\beta)$. If $|d_{T_1} - d_{T_2}| \approx |Dcos\alpha\cos\beta| < \frac{\lambda}{4} \approx 8cm$, we have

$$\Delta\boldsymbol{k} = \begin{cases} -2, & \Delta\Lambda(\alpha,\beta) \in [-2\pi,-\pi), \cos\alpha\cos\beta > 0 \\ -1, & \Delta\Lambda(\alpha,\beta) \in [-2\pi,-\pi), \cos\alpha\cos\beta < 0 \\ -1, & \Delta\Lambda(\alpha,\beta) \in [-\pi,0), \cos\alpha\cos\beta > 0 \\ 0, & \Delta\Lambda(\alpha,\beta) \in [-\pi,0), \cos\alpha\cos\beta < 0 \\ 0, & \Delta\Lambda(\alpha,\beta) \in [0,\pi), \cos\alpha\cos\beta > 0 \\ 1, & \Delta\Lambda(\alpha,\beta) \in [0,\pi), \cos\alpha\cos\beta < 0 \\ 1, & \Delta\Lambda(\alpha,\beta) \in (\pi,2\pi], \cos\alpha\cos\beta > 0 \\ 2, & \Delta\Lambda(\alpha,\beta) \in (\pi,2\pi], \cos\alpha\cos\beta < 0 \end{cases} \tag{5.10}$$

Thus, the optimal results α^* and β^* are estimated by finding the minimum,

$$(\alpha^*, \beta^*) = \underset{\substack{\alpha\in[0^\circ,180^\circ]\\ \beta\in[0^\circ,90^\circ]}}{\arg\min}\ \mathbb{P}(\alpha, \beta) \tag{5.11}$$

5.3.2 Convolutional Neural Network-assisted Orientation Estimation in Rich Multipath

To deal with the multipath impact, a CNN-based solution is proposed for spatial orientation prediction. The CNN architecture for our regression task is illustrated in Fig. 5.3. At first, we represent object orientation features along with multipath interference using an RGB image. Then to avoid manually collecting a large number of RFID signal fingerprints in coupling in all possible environments for training, we simulate these data to pre-train a CNN and then use only a few actual measurements to fine-tune the CNN. Finally, RF-Orien3D performs with this fine-tuned CNN for orientation tracking.

Building an Image for CNN. By introducing the multipath item, $\Delta h^{ref}_{T_1,T_2}(\alpha, \beta)$ can be rewritten as

$$\Delta h^{ref}_{T_1,T_2}(\alpha, \beta) = \frac{1 + h^{ref}_{T_1}(\alpha, \beta) + h^{mul}_{T_1}}{1 + h^{ref}_{T_2}(\alpha, \beta) + h^{mul}_{T_2}} \tag{5.12}$$

Since the two tags are very close to each other, we assume that the impact of multipath interference on their signal fingerprints is the same, i.e., $h^{mul}_{T_1} \approx h^{mul}_{T_2} = \rho^m_T e^{\boldsymbol{J}\phi^m_T}$, where $\rho^m_T \in (0, 1)$ and $\phi^m_T \in [-180^\circ, 180^\circ]$.

Given a pair of the angle guesses α and β, we use Eq. 5.11 to search the optimal $\rho^m_T|_{\alpha,\beta}$ and $\phi^m_T|_{\alpha,\beta}$,

$$(\rho^{m,*}_T, \phi^{m,*}_T)|_{\alpha,\beta} = \underset{\substack{\rho^m_T\in(0,1)\\ \phi^m_T\in[-180^\circ,180^\circ]}}{\arg\min}\ \mathbb{P}(\rho^m_T, \phi^m_T)|_{\alpha,\beta} \tag{5.13}$$

After searching all possible angles, we use the following rule to build an RGB image that are fed into a CNN,

$$\mathbb{I}(\alpha, \beta) = \begin{cases} norm(\cos\phi^{m,*}_T|_{\alpha,\beta}), & Channel\ 1 \\ norm(\sin\phi^{m,*}_T|_{\alpha,\beta}), & Channel\ 2 \\ norm(\rho^{m,*}_T|_{\alpha,\beta}), & Channel\ 3 \end{cases} \tag{5.14}$$

where the function $norm(\cdot)$ is to normalize a value to the range of 0∼255.

Simulating RFID Fingerprints for CNN Pretraining. Here the angle spacing for searching is set to 4°, so the image size is 46×23 pixels. And the searching spacings of ρ^m_T and ϕ^m_T are set to 0.05 and 20°, respectively. For a pair of spatial angles and multipath variables, we firstly depend on Eq. 5.7 to calculate $\Delta h^{dir}_{T_1,T_2}(\alpha, \beta)$ and $\Delta h^{ref}_{T_1,T_2}(\alpha, \beta, \rho^m_T, \phi^m_T)$, thereby obtaining the corresponding RSSI and phase difference in coupling, i.e., $10^{\frac{\mathcal{R}^r_{T_1} - \mathcal{R}^r_{T_2}}{20}}$ and $\mathcal{P}'_{T_1} - \mathcal{P}'_{T_2}$. Then an image

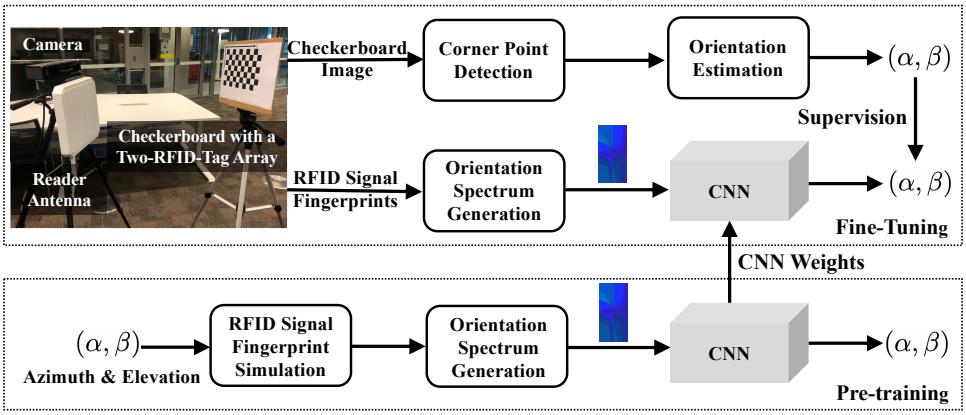


Figure 5.3 : CNN architecture with transfer learning.

$\mathbb{I}(\alpha, \beta)$ is simulated to represent a multipath case under the spatial orientation. Hence, a total number of images inputted into the CNN is $46 \times 23 \times 21 \times 19 = 422,142$ samples, and the CNN responses are specified azimuth and elevation angles normalized to 0∼1 by $\frac{\alpha}{180^\circ}$ and $\frac{\beta}{90^\circ}$.

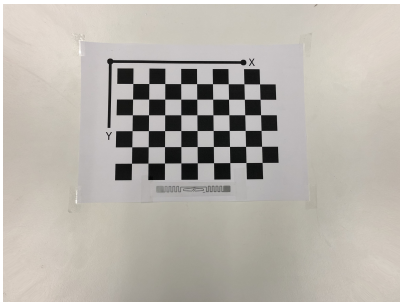
Collecting Actual RFID Fingerprints for CNN Fine-Tuning. Different application scenarios may produce different multipath interference. For example, in free space, the main multipath interference may come from the floor and a distant wall. In a dynamic environments where some people walk around, the dominant interference may be from the human body reflections. In this case, the pre-trained CNN can be fine-tuned for different multipath impacts, just like fine-tuning a pre-trained AlexNet or GoogLeNet to perform image classification on a new task. To achieve that, we need to collect a small number of actual RFID fingerprint measurements by varying the tag-to-antenna direction and/or distance. To label each sample, we adopt the camera-assisted method in Section 5.6 to calculate the azimuth and elevation angles, instead of manual measurement each time.

Performing Orientation Tracking using Fine-tuned CNN. Once the fine-tuning is over, our system only uses RFID fingerprints of a two-tag array measured in practice to generate a feature image $\mathbb{I}(\alpha, \beta)$ as the input of the fine-tuned CNN, without the help of the camera. In particular, our system can perform orientation tracking even in NLOS scenarios where the direct propagation path between a reader antenna and a two-tag array is blocked.

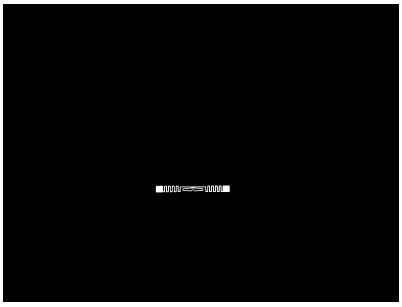
5.4 Simulating Radiation Pattern of Each Element in Two-RFID-Tag Array

In this section, we simulate an RFID tag from an image and then build a two-tag array to observe the changes in the radiation pattern of each element under the mutual coupling effect. The following simulation is conducted in Matlab 2020a using its antenna and image processing components:

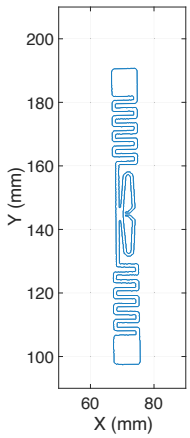
Segment RFID tag from an image. We firstly capture a photo of the checkerboard with an RFID tag to be modeled (a single-dipole Impinj E51 tag with the



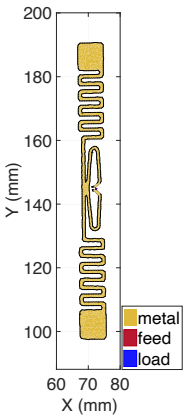
(a) Undistorted image



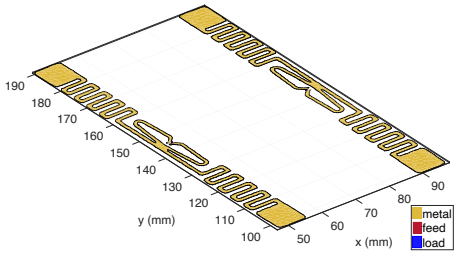
(b) Foreground image



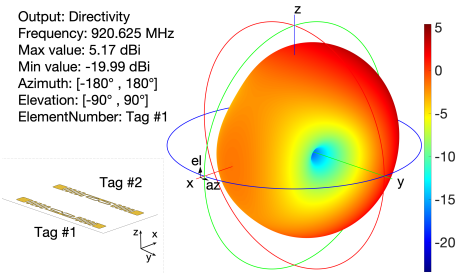
(c) Boundary points



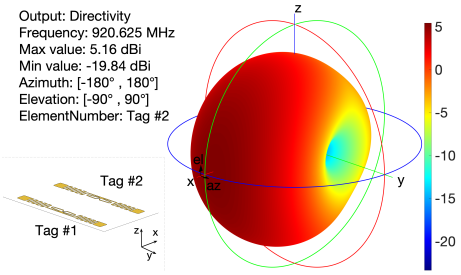
(d) RFID tag model



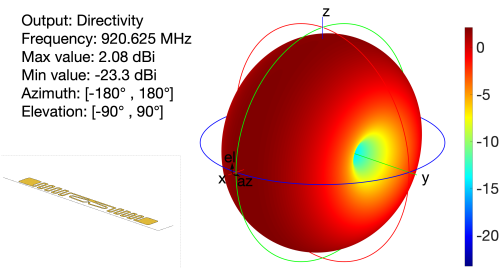
(e) Two-RFID-tag array



(f) Radiation pattern of tag #1



(g) Radiation pattern of tag #2



(h) Isolated radiation pattern

Figure 5.4 : Building two-RFID-tag array for radiation pattern analysis from a 2D photo.

size of 8mm×95mm is used here). After camera calibration [116], we rely on the estimated camera parameters to remove lens distortion from the photo (Fig. 5.4a). After that, we use a foreground segmentation tool in Matlab to segment the tag antenna out (Fig. 5.4b) for the following use.

Build an RFID tag model. We extract boundary points from the segmented foreground and downsample to reduce the computation. By obtaining the rotation and translation metrics of the undistorted image relative to the camera, we map the extracted points in pixel space into the real-world coordinates (Fig. 5.4c). In the real world view, the simulated tag size is 8.02mm×94.05mm, which is very close to the ground truth. Then we create an antenna feed to excite a voltage. In our experiments, the operating frequency is configured at 920.625Mhz, so the simulation result shows the antenna impedance at this frequency is $Z_{ant} = 9.04\Omega + \boldsymbol{J}182.23\Omega$. Suppose that the RFID chip impedance is matched to the tag antenna impedance here. A tag chip load with the reactance of $-\boldsymbol{J}182.23\Omega$ is added to the antenna, so the antenna source impedance is 9.04Ω. Thus, this tag model for MoM analysis is constructed (Fig. 5.4d). Compared to an existing work [48], it maps the boundary points in the image view to those in the real-world view by extracting top-left, top-right, bottom-left and bottom-right pixel points of the foreground and then scaling the tag dimension in the image relative to the actual tag size. However, it is challenging to ensure that the antenna boundary is completely parallel to image coordinate axes, especially after removing the lens distortion, which may introduce more error in obtaining a tag geometry.

Simulate each element pattern in a two-tag array. Based on above tag model, we create a linear array consisting of two mirror-symmetric tag elements with a given tag separation (e. g., 4cm shown in Fig. 5.4e). In RFID systems, an RFID reader communicates with the tags separately to avoid inventory collision. In this case, when a tag is responding, another tag element does not radiate its backscatter signal at the moment. We excite each tag in the array while terminating another one using a resistance of 9.04Ω to obtain each tag pattern (shown in Fig. 5.4f for element 1 and Fig. 5.4g for element 2. For comparison, Fig. 5.4h shows the radiation pattern of a tag placed in space by itself.

The radiation pattern simulation shows that:

1) The radiation pattern of each element in a two-RFID-tag array is modified when a nearby tag occurs. And their radiation patterns are fairly mirror-symmetric as long as the tag deployment underlies the geometry of mirror symmetry. If not, the radiation patterns will be no longer symmetrical (due to the space limit, the patterns are not plotted in this work). And the tag gains are enhanced in some directions while being weakened in others. When a tag is inventoried, another adjacent tag works just like an obstacle blocks the signal from the reader antenna to the responding tag.

2) The actual tag gain is actually equal to the tag antenna directivity multiplied by its radiation efficiency. Let ε be a tag antenna's radiation efficiency, and let $\mathbb{D}$ be its directivity, so we have $g_{T_1} = \varepsilon\mathbb{D}_{T_1}$, $g_{T_2} = \varepsilon\mathbb{D}_{T_2}$, $g_{T_{12}} = \varepsilon\mathbb{D}_{T_{12}}$, and $g_{T_{21}} = \varepsilon\mathbb{D}_{T_{21}}$. In our simulation, this efficiency ε is assumed to be 1. In this case, the simulation

enables to output lossless tag antenna gains (equal to tag antenna directivity) in all directions, including the tag gains g_{T_1} and g_{T_2} in all azimuth and elevation angles, as well as $g_{T_{12}}$ (and $g_{T_{21}}$) in the direction of T_1 (and T_2) towards T_2 (and T_1). The simulated results are the power gains G_{dBi} in units of dBi. To convert them to voltage gains, we use $g = 10^{\frac{G_{dBi}}{20}}$. In fact, the mutual coupling results in tag impedance mismatch so that the real efficiency is less than 1. Fortunately, we can remove this unknown antenna efficiency in our system. Recall from Eq. 5.7, ε is canceled in calculating $\frac{g_{T_1}(\alpha,\beta)}{g_{T_2}(\alpha,\beta)}$ and $\frac{g_{T_2}(\alpha,\beta)}{g_{T_1}(\alpha,\beta)}$. And the orientation-independent constant value $\rho_T^{\Delta} g_{T_{12}} g_{T_{21}}$ is estimated using actual RFID signal fingerprint measurements, not from this simulation.

5.5 Calculating Scaled Modulation Factor

Recall from Section 5.5 that the tag phase shift is a constant value and the modulation factor is dependent on the tag spacing between two tags. In this section, we discuss how to determine them using RFID fingerprints measured in non-coupling and coupling cases.

5.5.1 Tag Phase Shift Estimation

Without considering the multipath effect, Eq. 5.4 can be rewritten as (take T_1 for example)

$$\begin{cases} \Gamma_{T_1} = \sqrt{1 + 2\boldsymbol{U}_{T_1} \cos \boldsymbol{V}_{T_1} + {\boldsymbol{U}_{T_1}}^2} & \text{(5.15a)} \\ \Theta_{T_1} = \arctan\left(\dfrac{-\boldsymbol{U}_{T_1} \sin \boldsymbol{V}_{T_1}}{1 + \boldsymbol{U}_{T_1} \cos \boldsymbol{V}_{T_1}}\right) (\mathrm{mod}\,(-\pi, \pi]) & \text{(5.15b)} \end{cases}$$

where

$$\begin{cases} \boldsymbol{U}_{T_1} = \dfrac{\rho_T^{\Delta} g_{T_2} g_{T_{12}} g_{T_{21}} d_{T_1}}{g_{T_1}\left(d_{T_2} + D\right)} & \text{(5.16a)} \\ \boldsymbol{V}_{T_1} = \dfrac{2\pi}{\lambda}\left(d_{T_2} + D - d_{T_1}\right) + \phi_T & \text{(5.16b)} \end{cases}$$

Given a guess of $\phi_T \in [0, 2\pi]$, we rely on Eq. 5.3 to build a function $\mathbb{J}(\phi_T)$ that determines the likelihood of ϕ_T being the ground truth based on RSSI and RF phase of T_1 and T_2 collected in the non-coupling and coupling cases,

$$\mathbb{J}(\phi_T) = \left| \left[\frac{g_{T_1}(\phi_T)\, d_{T_2}}{g_{T_2}(\phi_T)\, d_{T_1}}\right]^2 \frac{\Gamma_{T_1}(\phi_T)}{\Gamma_{T_2}(\phi_T)} - 10^{\frac{R^r_{T_1} - R^r_{T_2}}{20}} \right| \tag{5.17}$$

where

$$\left(\frac{d_{T_2}}{d_{T_1}}\right)^2 = 10^{\frac{\overline{R}^r_{T_1} - \overline{R}^r_{T_2}}{20}} \tag{5.18}$$

And the methods of calculating $\frac{g_{T_1}(\phi_T)}{g_{T_2}(\phi_T)}$ and $\frac{\Gamma_{T_1}(\phi_T)}{\Gamma_{T_2}(\phi_T)}$ are described in the following. When the guess ϕ_T approaches the ground truth, $\mathbb{J}(\phi_T)$ will be close to 0; otherwise, $\mathbb{J}(\phi_T)$ will be far larger than 0.

Calculate Tag Gain Ratio. At first, the coupling phase shift of T_i, $i = 1, 2$, is calculated based on Eq. 5.3 and Eq. 5.5,

$$\Theta_{T_i} = \left(\mathcal{P}'_{T_i} - \overline{\mathcal{P}}'_{T_i}\right) (\mathrm{mod}\, [-\pi, \pi]) \tag{5.19}$$

Then, the tag separation D is a known value (where $D \leq \frac{\lambda}{4} \approx 8cm$), and the tag-to-antenna distance difference can be determined by RF phase measurements without coupling. Thus, $\boldsymbol{V}_{T_i}(\phi_T)$ in Eq. 5.16b is computed by

$$\begin{cases} \boldsymbol{V}_{T_1}(\phi_T) = \dfrac{2\pi}{\lambda}(d_{T_2} + D - d_{T_1}) + \phi_T \\ \boldsymbol{V}_{T_2}(\phi_T) = \dfrac{2\pi}{\lambda}(d_{T_1} + D - d_{T_2}) + \phi_T \end{cases} \tag{5.20}$$

Given the calculated Θ_{T_i} and $\boldsymbol{V}_{T_i}(\phi_T)$, we use Eq. 5.15 to obtain $\boldsymbol{U}_{T_i}(\phi_T)$,

$$\boldsymbol{U}_{T_i}(\phi_T) = \frac{-\tan\Theta_{T_i}}{\sin \boldsymbol{V}_{T_i}(\phi_T) + \cos \boldsymbol{V}_{T_i}(\phi_T) \tan\Theta_{T_i}} \tag{5.21a}$$

$$\text{s.t.} \begin{cases} \boldsymbol{U}_{T_1}(\phi_T) \in (0, \max(g_{T_2})\, g_{T_{12}} g_{T_{21}}) \\ \boldsymbol{U}_{T_2}(\phi_T) \in (0, \max(g_{T_1})\, g_{T_{12}} g_{T_{21}}) \end{cases} \tag{5.21b}$$

Here the phase periodicity of π radians in Θ_{T_i} is removed by the tangent function. The term $\max(g_{T_1})$ (and $\max(g_{T_2})$) represents the maximum tag gain of T_1 (and T_2). Here $(\max(g_{T_2})\, g_{T_{12}} g_{T_{21}})$ and $(\max(g_{T_1})\, g_{T_{12}} g_{T_{21}})$ are calculated based on our simulated radiation patterns.

Since $D \ll d_{T_1}, d_{T_2}$, the tag gain ratio of T_1 to T_2 is approximately

$$\begin{aligned} \left[\frac{g_{T_1}(\phi_T)}{g_{T_2}(\phi_T)}\right]^2 &= \frac{\boldsymbol{U}_{T_2}(\phi_T) d_{T_1}(d_{T_1} + D)}{\boldsymbol{U}_{T_1}(\phi_T) d_{T_2}(d_{T_2} + D)} \\ &\approx \frac{\boldsymbol{U}_{T_2}(\phi_T)}{\boldsymbol{U}_{T_1}(\phi_T)} \left(\frac{d_{T_1}}{d_{T_2}}\right)^2 \end{aligned} \tag{5.22}$$

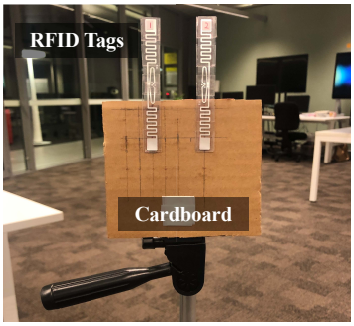
Calculate Coupling Term of RSSI. The coupling term $\Gamma_{T_i}(\phi_T)$, $i = 1, 2$, is determined by

$$\Gamma_{T_i}(\phi_T) = \sqrt{1 + 2\boldsymbol{U}_{T_i}(\phi_T) \cos \boldsymbol{V}_{T_i}(\phi_T) + \left[\boldsymbol{U}_{T_i}(\phi_T)\right]^2} \tag{5.23}$$

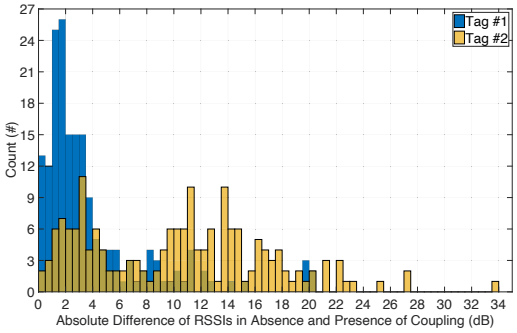
Through above steps, the optimal tag phase shift is estimated by searching the minimum of $\mathbb{J}(\phi_T)$,

$$\phi_T^* = \underset{\phi_T \in (0, 2\pi)}{\arg\min}\, \mathbb{J}(\phi_T) \tag{5.24}$$

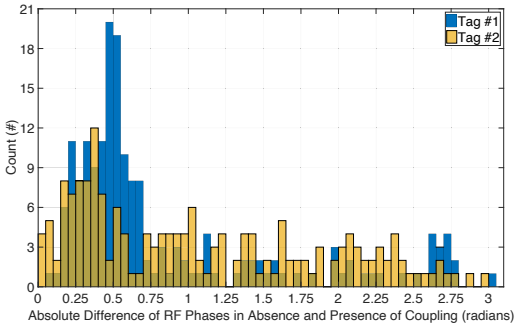
Here we conduct an experiment in free space with negligible multipath interference by varying the tag separation D from 3cm to 8cm. The reading frequency is fixed at 920.625 Mhz. The RFID tag positions relative to the reader antenna are changed each time. Fig. 5.5a plots the experiment setup. Fig. 5.5b and Fig. 5.5c



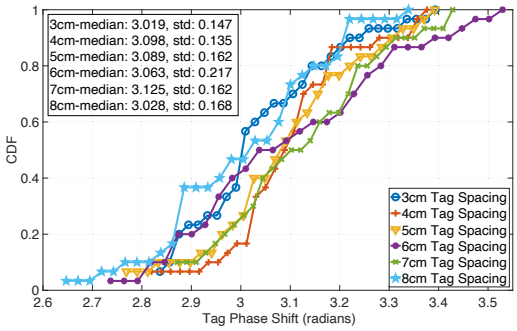
(a) Experiment setup



(b) The change in RSSI



(c) The change in RF Phase



(d) The CDF of tag phase shift

Figure 5.5 : Tag phase shift over different tag separations.

depict the absolute difference in RSSI and RF phase before and after tag mutual coupling occurs for Tag #1 and Tag #2, respectively, which shows the coupling effect results in the unexpected changes in RSSI and RF phase. Based on these collected data, Fig. 5.5d shows the median values of the calculated tag phase shift at different tag separations are 3.019 radians, 3.098 radians, 3.089 radians, 3.063 radians, 3.1245 radians and 3.0280 radians, respectively, which all approaches π radians. In the reflective state of an RFID tag, a short circuit is presented to the tag antenna and no power is delivered to the tag chip. In this case, the RFID tag acts just like a transmission line without carrying any load. Compared to the incident signal, the backscatter signal from the tag is a version with 180° out of phase. Thus, the tag phase shift ϕ_T is equal to π radians.

5.5.2 Scaled Modulation Factor Estimation

First of all, let us consider a scenario without the coupling effect. Suppose that ρ_T^A and ρ_T^B are the the reflection coefficients of the RFID tag in the absorptive and reflective states (i.e, State A and State B), respectively. The antenna impedance is $Z_{ant} = R_a + \boldsymbol{J}X_a$, where the real part R_a is a resistance and the image part X_a is a reactance. In the absorptive state, the chip impedance is $Z_{chip}^A = R_c^A + \boldsymbol{J}X_c^A$. In the reflective state for the short-circuit load chip, the chip impedance is $Z_{chip}^B = R_c^B + \boldsymbol{J}X_c^B \approx 0$. The modulation factor $\overline{\rho}_T^{\Delta}$ in the non-coupling case is

$$\overline{\rho}_T^{\Delta} = \left|\rho_T^A - \rho_T^B\right|^2 \tag{5.25}$$

Table 5.1 : Modulation factor over tag spacing

D	ρ_T^{Δ}	$\varepsilon^2\rho_T^{\Delta}$ **(Median)**	$\varepsilon^2\rho_T^{\Delta}$ **(Std.)**
3 cm	0.582	1.341	0.225
4 cm	0.683	1.027	0.134
5 cm	0.784	0.866	0.109
6 cm	0.905	0.775	0.162
7 cm	1.042	-	-
8 cm	1.180	-	-

where

$$\begin{cases} \rho_T^A = \dfrac{Z_{chip}^A - Z_{ant}^*}{Z_{chip}^A + Z_{ant}} \\ \rho_T^B = \dfrac{Z_{chip}^B - Z_{ant}^*}{Z_{chip}^B + Z_{ant}} \end{cases} \tag{5.26}$$

For example, at 920.625 Mhz, the simulated antenna impedance of the E51 tag is $Z_{ant} = 9.26\Omega + \boldsymbol{J}185.06\Omega$ and $Z_{chip}^A = Z_{ant}^*$, so the modulation factor is $\overline{\rho}_T^{\Delta} = 1$.

When the mutual coupling occurs in the two-tag array, each tag will produce an additional mutual impedance on another one's antenna. The impedance is independent of the direction of a two-tag array relative to a reader antenna and is only determined by its tag spacing. The modified tag antenna impedances are

$$\begin{cases} Z_{ant,1}^{'} = Z_{ant} + Z_{12} \\ Z_{ant,2}^{'} = Z_{ant} + Z_{21} \end{cases} \tag{5.27}$$

where Z_{12} and Z_{21} are mutual impedances, and $Z_{12} = Z_{21}$ for the same tag type, so we have $Z_{ant}^{'} = Z_{ant,1}^{'} = Z_{ant,2}^{'}$. In particular, Z_{12} and Z_{21} can be obtained by calculating the S-parameters of our simulated two-tag array. Then we substitute them into Eq. 5.27 to calculate the modulation factor ρ_T^{Δ} in the coupling case. Table. 5.1 shows the simulated modulation factor over 3cm∼8cm tag spacings, which changes with different tag separations.

However, the antenna-chip impedance matching cannot be satisfied for all frequencies, so accurately estimating the modulation factor by simulation is very challenging. Instead, since $\boldsymbol{U}_{T_1}$ and $\boldsymbol{U}_{T_2}$ can be calculated by substituting $\phi_T = \pi$ radians back to Eq. 5.21, the scaled modulation factor $\varepsilon^2\rho_T^{\Delta}$ is calculated as follows by considering the tag antenna radiation efficiency,

$$\begin{aligned} \rho_T^{\Delta} &= \frac{1}{g_{T_{12}} g_{T_{21}}} \sqrt{\frac{\boldsymbol{U}_{T_1}\boldsymbol{U}_{T_2}\left(d_{T_1}+D\right)\left(d_{T_2}+D\right)}{d_{T_1}d_{T_2}}} \\ &\approx \frac{\sqrt{\boldsymbol{U}_{T_1}\boldsymbol{U}_{T_2}}}{g_{T_{12}} g_{T_{21}}} = \frac{\sqrt{\boldsymbol{U}_{T_1}\boldsymbol{U}_{T_2}}}{\varepsilon^2 \mathbb{D}_{T_{12}} \mathbb{D}_{T_{21}}} \end{aligned} \tag{5.28}$$

where ρ_T^{Δ} cannot be directly estimated due to the unknown ε^2.

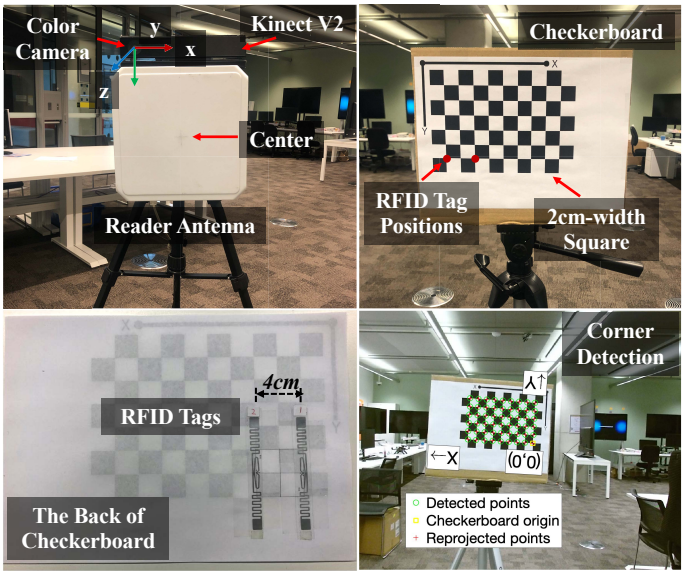


Figure 5.6 : Experiment setup for orientation labelling.

Table. 5.1 also shows the modulation factors multiplied by ε^2 (i. e., $\varepsilon^2\rho_T^{\Delta}$) at 3cm∼6cm tag spacing calculated based on the experiment data in Section 4.6.1. For the separations of 7cm and 8cm, we find the corresponding estimates are weakly trusted because they are sensitive to the measurement noise in RFID fingerprints amplified by the tangent function and thereby result in a fairly large standard deviation. The proposed system performs orientation tracking using these measurements so as to avoid the unknown ε. In addition, compared to simulated modulation factors, we infer that as the tag separation decreases, the mutual impedance value will correspondingly increase, resulting in the lower antenna efficiency ε in the coupling case.

5.6 Generating Spatial Orientation Labels for CNN

Annotating a two-RFID-tag array orientation relative to a fixed reader antenna in 3D space is very challenging. A time-consuming solution is to manually measure the tag positions in the antenna-centric coordinate system for orientation estimation. And its measurement accuracy may not be high enough. In this work, a camera-based solution is proposed to address this challenge. A commonly-used 2D camera binds with an RFID reader antenna. The two-tag array is attached to the back of checkerboard, and each RFID tag center is aligned with the corresponding corner point. The checkerboard pattern is then attached to a cardboard box. The experiment setup is shown in Fig. 5.6. After camera calibration [116], when a new checkerboard image (along with RFID fingerprints of two RFID tags) is obtained, we firstly undistort the image and then compute the rotation matrix $\boldsymbol{R}_{3\times3}$ and the translation matrix $\boldsymbol{t}_{1\times3}$ of the checkerboard relative to the camera in the camera-centric coordinate system. The antenna coordinate is denoted by $\boldsymbol{X}_A = \Delta\boldsymbol{X}$, where $\Delta\boldsymbol{X} = (\Delta x, \Delta y, \Delta z)$ is the manually-measured position displacement of the antenna relative to the camera (the plastic radome center of the antenna with the size of

Table 5.2 : Impinj R420 Reader Configuration

Settings	Values
Reader Mode	Dense Reader M4
Search Mode	Dual Target
Session	Session 0
Transmitted Power	32.5 dBm
Operating Frequency	920.625 Mhz

26cm×26cm×4.5cm is considered as its physical location). The antenna coordinate $\boldsymbol{X}_A^w = (x_A^w, y_A^w, z_A^w)$ in the checkerboard-centric coordinate system is

$$\boldsymbol{X}_A^w = (\Delta \boldsymbol{X} - \boldsymbol{t}_{1\times 3})\, \boldsymbol{R}_{3\times 3}^{'} \tag{5.29}$$

where $\boldsymbol{R}_{3\times 3}^{'}$ is the transpose matrix of $\boldsymbol{R}_{3\times 3}$. Thus, the azimuth and elevation angles are

$$\begin{cases} \alpha = \arctan \dfrac{|y_A^w|}{x_A^w} \\ \beta = \arctan \dfrac{|z_A^w|}{\sqrt{(x_A^w)^2 + (y_A^w)^2}} \end{cases} \tag{5.30}$$

5.7 Implementation & Evaluation

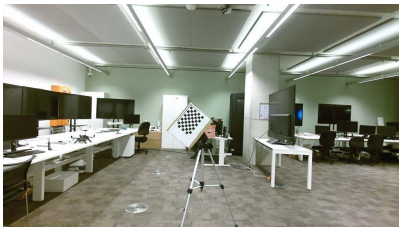
5.7.1 Implementation

RFID Reader. We adopt an Impinj R420 RFID reader. The operating configuration is described in Table. 5.2. Note that the phase ambiguity is caused by PR-ASK modulation when an RFID reader is configured at the operating mode of 'Hybrid', 'Dense Reader M4', 'Dense Reader M8', 'Max Miller' or 'Dense Reader M4 Two'. Although the commonly-used mode 'Max Throughput' using double sideband-amplitude shift keying (DSB-ASK) modulation can avoid the phase ambiguity and provide a much faster reading rate, some RFID tags would fail to be read in the surveillance zone due to worse robustness of the forward and backward communications links.

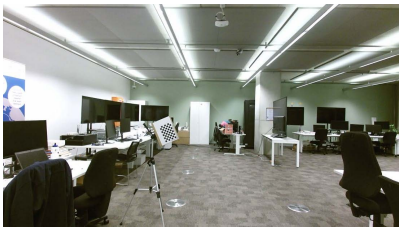
RFID Reader Antenna. A circularly-polarized antenna is linked to the port 1 of the RFID reader. The antenna gain is 8dBi, which can achieve 6∼10m reading range.

RFID Tag. Two Impinj E51 RFID tags are adopted to construct a tag array. The tag has a single dipole. Note that the radiation pattern of a dual-dipole tag (e.g., Impinj H47) is almost isotropic with the same gain in all directions in a non-coupling scenario. When the coupling effect occurs in a two-tag array, the modified pattern has little difference in the azimuth angle. We only use single-dipole tags in our system for spatial orientation tracking.

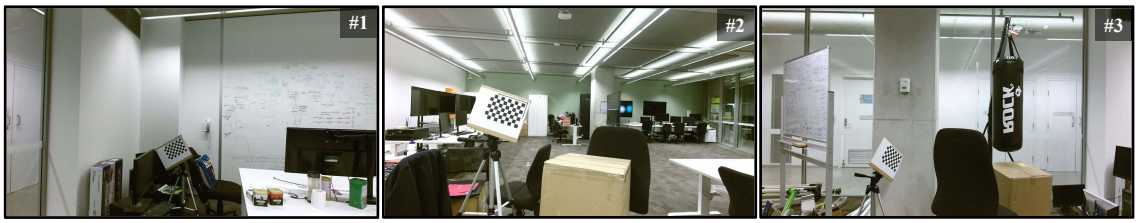
Camera. A Microsoft Kinect V2 is mounted on the reader antenna. We only use its color camera to capture 2D checkerboard images for labeling.



(a) Low multipath



(b) Medium multipath



(c) Rich multipath

Figure 5.7 : Actual RFID fingerprint collection in different scenarios.

5.7.2 Evaluation

Tag-to-Tag Spacing. By default, the tag spacing in a two-tag array is set to 4 cm, because such separation distance can achieve a better balance between tag responding distance and orientation tracking error, which will be discussed in Section 5.8.3. Correspondingly, the values of $\varepsilon^2\rho_T^{\Delta} = 1.027$ and $\phi_T = \pi$ radians are adopted.

Convolutional Neural Network Architecture. Since the constructed RGB image is 46×23×3, the first image input layer is created with the same size as the image. The CNN then creates 4 two-dimensional convolutional layers with 8, 16, 32 and 32 filters, respectively, of size 3×3 and 'same' padding. We use a batch normalization layer followed by a ReLU layer after every layer. We also connect an average pooling layer with pool size 2×2 and stride 2×2 to each of the first two ReLU layers. Finally, a dropout layer with dropout probability 0.2 is created, followed by a fully connected output layer of size 2 (i.e., azimuth and elevation angles) and a regression layer. The network is implemented in Matlab 2020a.

Training Details. The CNN is pre-trained using a batch size of 64 and Adam optimizer with an initial learning rate of 1×10^{-3}. When fine-tuning, we freeze the weights of the first 9 layers in the pre-trained network by setting the learning rates to zero. And its batch size is reduced to 16 and the initial learning rate is 1×10^{-4}.

Measured RFID Fingerprints for Fine-tuning and Validation in Different Scenarios. As shown in Fig. 5.7, RFID fingerprint measurement is conducted in the scenarios of low-multipath LOS, low-multipath NLOS, medium-multipath LOS and rich-multipath LOS, respectively. The corresponding sample size is 1001, 667, 1460 and 2140. In the low-multipath case, many reflection objects are far from the tag array. The medium-multipath case includes tables and chairs located not too far away from the tag array. The rich-multipath case contains some objects with strong reflection ability (such as chairs, tables, computer screen, wall, etc.) around

the tag array. For the NLOS scenario, we use a cardboard box to block the path between the tag array and the antenna when collecting the RFID fingerprints. To capture the orientation by the camera, the box is removed, but the tag-antenna geometry keeps unchanged.

Baseline. We conduct five baselines to compare to our system RF-Orien3D (with simulated data for pre-training and with actual measured data for fine-tuning).

Naive method. The included angle between two sides of Antenna-to-Tag#1 and Tag#1-to-Tag#2 in a 2D plane is directly calculated by

$$\theta_{2D} \approx \arccos\left[\frac{\lambda}{4\pi D}\left(P'_{T_1} - P'_{T_2} + \pi\Delta\boldsymbol{k}\right)\right] \tag{5.31}$$

Tagyro[105]. The system needs to pre-estimate the virtual geometry of these tags to deal with tag mutual coupling. For a two-tag array used in our experiment, we rely on its approach to obtain a virtual tag spacing $D^{'} = 6.6$ cm. Then we substitute it into Eq. 5.31 for estimation.

Orientation-spectrum. Recall from Section 5.2 that an orientation spectrum is proposed to directly track orientation in 3D space.

RF-Orien3D only with simulated data for CNN training (RF-Orien3D W. S.). Only simulated dataset is fed into a CNN for training. Once the training is over, we directly validate our system performance using actual measurements.

RF-Orien3D only with measured data for CNN training (RF-Orien3D W. M.). We directly use actually measured samples to train a CNN for tracking. Then we test our system using validation datasets.

In particular, since both Naive method and Tagyro can only achieve 2D-plane orientation tracking when using a two-tag array, the angle θ_{2D} in a 2D plane for RF-Orien3D can be calculated by substituting the estimated azimuth and elevation angles,

$$\theta_{2D} \approx \arccos\left(\cos\alpha\cos\beta\right) \tag{5.32}$$

And the orientation-spectrum and RF-Orien3D based solutions output 3D-space angles for performance comparison.

Metric and Ground Truth. The angular error in azimuth and elevation is defined as the absolute difference between the estimate and ground truth. The ground-truth orientation is captured by a camera system.

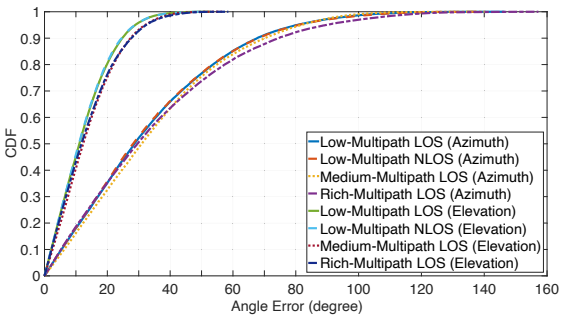
5.8 Results

5.8.1 Orientation Tracking Performance

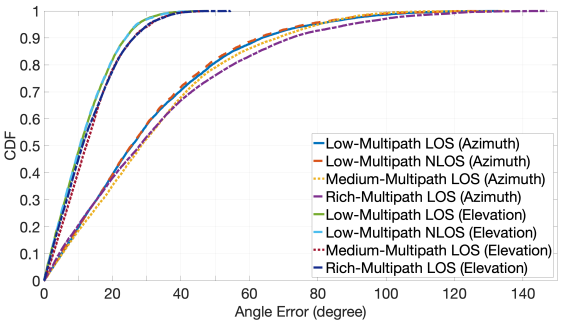
We start by evaluating RF-Orien3D performance in different scenarios. The samples in each case is randomly divided into only 5% of the samples for fine-tuning and 95% for validation in each case. After that, we combine the fine-tuning data (about 250 samples) from different scenarios together to fine-tune the pre-trained CNN. The experiment is repeated 50 times. Fig. 5.8a plots the CDF of orientation

Table 5.3 : Median errors in azimuth and elevation in different scenarios

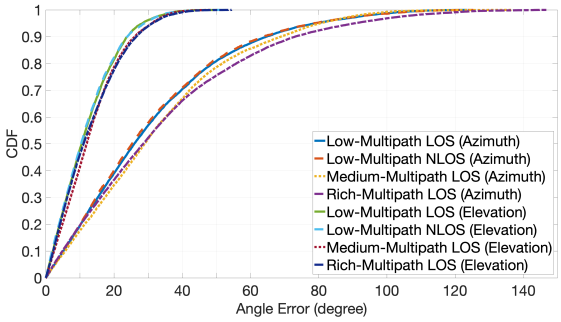
	Low-Multipath LOS	**Low-Multipath NLOS**	**Medium-Multipath LOS**	**Rich-Multipath LOS**
RF-Orien3D (5% of Fine-tuning Samples)	$(28.630°, 11.093°)$	$(28.193°, 10.938°)$	$(30.923°, 12.297°)$	$(29.574°, 11.758°)$
RF-Orien3D (80% of Fine-tuning Samples)	$(25.668°, 10.533°)$	$(25.477°, 10.463°)$	$(28.357°, 12.334°)$	$(27.858°, 11.195°)$
RF-Orien3D W. M. (80% of Training Samples)	$(25.901°, 10.520°)$	$(25.248°, 10.350°)$	$(28.997°, 11.964°)$	$(28.557°, 11.059°)$



(a) RF-Orien3D with 5% of samples



(b) RF-Orien3D with 80% of samples



(c) RF-Orien3D W. M. with 80% of samples

Figure 5.8 : Orientation tracking performance in different scenarios.

tracking errors in azimuth and elevation for each case. Also, Fig. 5.8b and Fig. 5.8c plot the results for RF-Orien3D with 80% fine-tuning samples and RF-Orien3D W. M. with 80% training samples. Table. 5.3 shows the median angular errors. The experiment reveals the following findings:

1) When 5% of samples are fed, the angular errors of RF-Orien3D are slightly lower than other two cases with the large sample dataset by about $2° \sim 3°$ in azimuth and $1°$ in elevation, respectively. It could illustrate that RF-Orien3D can be driven for orientation tracking without collecting large-scale samples in actual environments. In addition, we claim that RF-Orien3D can deal with most of multipath scenarios due to our multipath simulation. However, RF-Orien3D W. M. can also use 80% samples to achieve the similar tracking errors. This means that our orientation spectrum itself has already been equipped with the capability to combat multipath interference. Since the two tags are close to each other, the multipath effect has the same impact on each tag's signal fingerprint and thereby can be suppressed by the differencing method in Eq. 5.6. Even so, RF-Orien3D can achieve better performance than RF-Orien3D W. M. when smaller samples are used (see

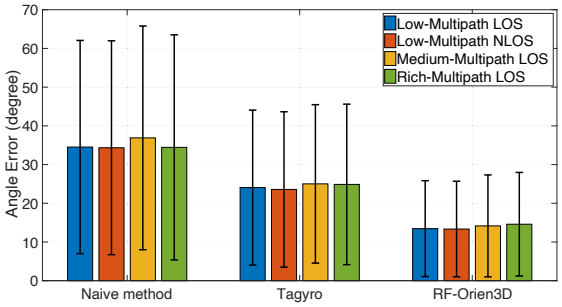


Figure 5.9 : 2D orientation error.

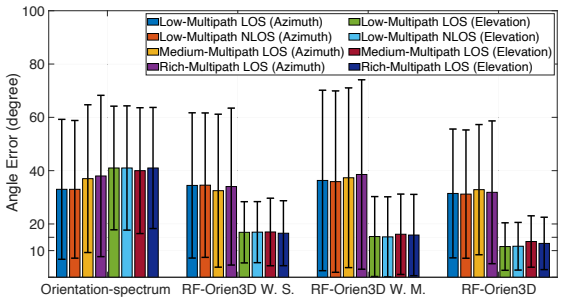


Figure 5.10 : 3D orientation error.

Section 5.8.3).

2) The median errors of RF-Orien3D in different scenarios are very close to each other, which means the impact of multipath interference could be significantly suppressed. And the errors in medium and rich-multipath cases are just slightly higher than in the low-multipath cases. Such an experiment result is expected due to the multipath diversity in the medium and rich multipath cases. Two potential solutions could further improve accuracy. One is to collect more RFID fingerprints in actual application scenarios. Another is to simulate more multipath impacts on RFID fingerprints. However, a tradeoff between the amount of simulated data and pre-training cost should be considered.

3) In low-multipath LOS and NLOS, the average errors are very close, so it could demonstrate that RF-Orien3D can perform even in NLOS. Since the impact of an obstruction on forward and backward communication links of the two tags are the same, our differencing method can completely remove its impact.

4) The tracking error in azimuth is about 2.5 times higher than that in elevation. This is due to the fact that the mutual coupling effect results in a relatively smaller variation of tag gain in azimuth (i.e., X-Y plane in Fig. 5.12) than that in elevation (i.e., X-Z plane in Fig. 5.13). To reduce the azimuth error, we can introduce more RFID tags to build a planar tag array instead of a linear two-tag array. However, accurately calculating the modulation factor in the planar array is very challenging, which is left for our future work.

5.8.2 Performance Comparison

To compare our system RF-Orien3D with RF-Orien3D W. M. that is trained only using actual RFID measurements, we randomly extract 2.5% of samples (about 133 samples) to fine-tune the pre-trained CNN for RF-Orien3D and directly train a new CNN for RF-Orien3D W. M. And the remaining sample are used for validation. Also, we depend on these validation dataset to evaluate the tracking accuracy of Naive method and Tagyro in 2D plane, and that of Orientation-spectrum and RF-Orien3D W. S. in 3D space. The experiment is repeated 50 times. Fig. 5.9 plots the 2D orientation error for Naive method, Tagyro and RF-Orien3D, while Fig. 5.10 plots the 3D orientation errors in azimuth and elevation for Orientation-spectrum and RF-Orien3D based methods. From the experimental results, we have the following conclusion:

1) For 2D direction estimation, Fig. 5.9 shows that RF-Orien3D outperforms Naive method and Tagyro by about 2.61 times and 1.77 times, respectively. In different multipath scenarios, these methods have the similar results due to the fact that multipath interference induces the similar distortion in RFID signal fingerprints of the two closely-spaced tags. However, the Naive method completely ignores mutual coupling interference. And Tagyro has a little improvement in accuracy. However, the virtual tag separation may be changed in different directions of a two-tag array relative to a reader antenna. RF-Orien3D can achieve better performance of combating coupling and multipath interference than the method of purely differencing phase in Naive method and Tagyro.

2) For 3D direction estimation, Fig. 5.10 shows RF-Orien3D can achieve lower estimation errors and standard deviations in azimuth and elevation than other methods. For Orientation-spectrum, it ignores the impact of multipath interference and cannot accurately differentiate possible directions from the tiny variation in tag radiation pattern. Recall from Section 5.8.1 that given enough training samples, RF-Orien3D W. M. can achieve similar performance to RF-Orien3D. However, when only a smaller number of samples are used to train in this experiment, its estimation accuracy is relatively lower, especially in azimuth estimation, because most of the divided RFID fingerprints in the validation dataset are relatively new relative to the training dataset. This experiment could demonstrate that the pre-trained CNN can help our system deal with new multipath cases that never occur in training samples.

3) Surprisingly, RF-Orien3D W. S. just has a little lower accuracy than RF-Orien3D, but higher than RF-Orien3D W. M. It could reveal the effectiveness of our RSSI/phase-distance models with mutual coupling and multipath terms, and thereby the simulated data may be similar to actual measurements. Unfortunately, the estimated modulation factor and the simulated radiation pattern under impedance matching may more or less deviate from the ground truth. After fine-tuning using a few numbers of samples, RF-Orien3D could achieve better performance.

5.8.3 Microbenchmarks

Next, we evaluate RF-Orien3D performance with two key settings: tag-to-tag separation and tag-to-antenna distance.

Impact of tag-to-tag separation. At first, we deploy the centers of a reader antenna and two tags on the same line. A tag (e.g., T_1) is located between the reader antenna and another shadowed tag (e.g., T_2). Here we observe the impact of the tag separation on RSSI of T_2, because it has a lower RSSI than T_1 and thereby cannot be read sometimes. For our system, however, it is required to keep reading both of them for tracking. We adjust the position of T_1 to change the spacing between them in each experiment. Also, we vary the distance of T_2 to the reader antenna from 100cm to 500cm with a spacing of 50cm. Fig. 5.11 and Fig. 5.12 plot tag radiation patterns in the X-Y and X-Z planes under the tag separation distances of 3cm$\sim$8cm. The variation in RSSI over different tag-to-antenna distances are shown in Fig. 5.13. Fig. 5.14 shows the angular errors as the tag spacing increases from

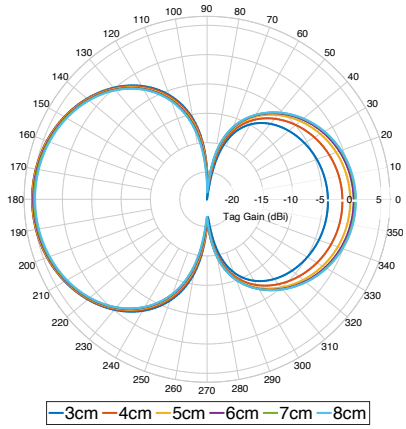


Figure 5.11 : X-Y plane.

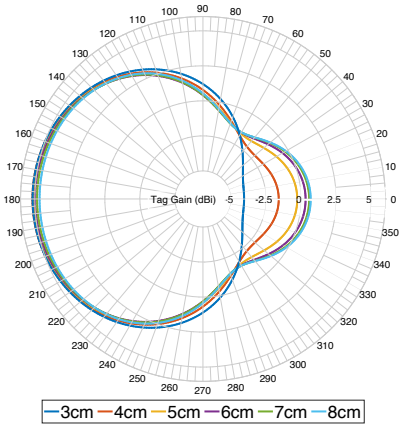


Figure 5.12 : X-Z plane.

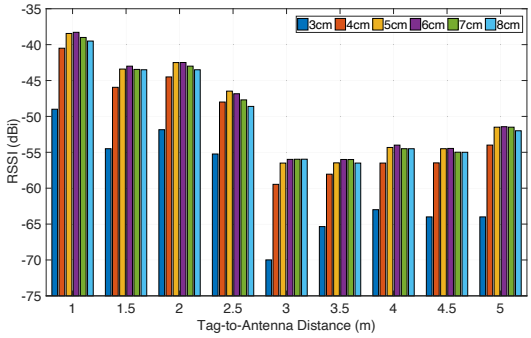


Figure 5.13 : RSSI vs. tag-antenna distance.

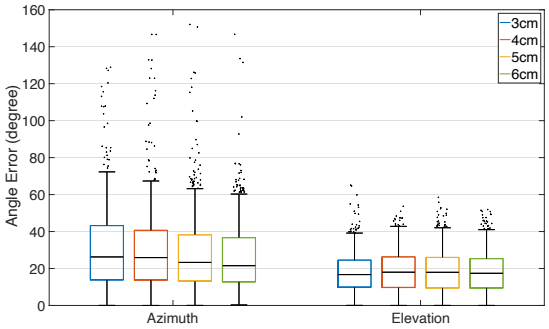


Figure 5.14 : Impact of tag separation.

3cm to 6cm. We can see that:

1) The tag array with a smaller spacing has a substantial decrease in reported RSSI. As the tag separation increases, the change in the radiation pattern of each tag due to tag mutual coupling is correspondingly weakened. And if we ignore the loss in antenna efficiency due to impedance mismatching, the variations in measured RSSI and tag radiation pattern are very similar (2 times loss in tag gain due to forward and backward communication). For example, the RSSI is increased by about 7.5dB from the tag separation of 3cm to 4cm, while the tag gain is increased by about 5dB. It could validate the effectiveness of our radiation pattern simulation.

2) Given the tag spacings of 3cm to 6cm, we obtain each pre-trained model and then test the impact of tag separation based on the simulated data in the same array-to-antenna directions for fairness. Note that the training samples do not contain these data. The sample size is about 1000 for each case. We evaluate the azimuth angle error given the elevation angle of 0° and the elevation angle error given the azimuth angle of 0°, respectively. Fig. 5.14 shows that as the tag separation increases, the azimuth error slightly reduces. However, the elevation error has a slight increase. This result also follows the variation in tag radiation pattern.

3) Our system needs to collect RFID fingerprints of two tags for estimation. A smaller tag spacing results in a larger mutual impedance, thereby decreasing the

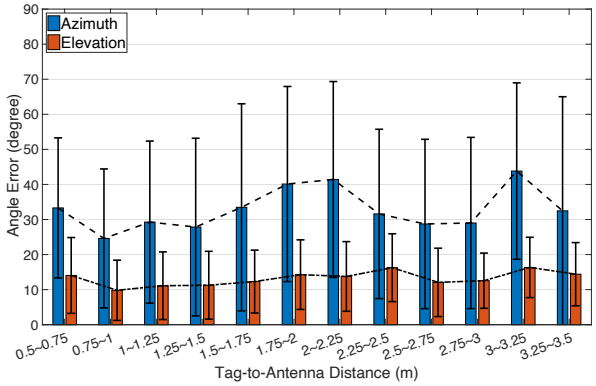


Figure 5.15 : Impact of tag-antenna distance.

power transfer coefficient and modulation factor in forward and backscatter links, respectively. In this case, the shadowed tag will be more difficult to be read in some directions or at longer tag-to-antenna distances. We need to look for a balance between tracking accuracy and responding distance. In our system, we select a 4-cm tag spacing for use.

Impact of tag-to-antenna distance. In this experiment, 5% of samples are used for fine-tuning. The remaining validation dataset is split into different tag-to-antenna distance ranges with a spacing of 25cm. The ground-truth distances can be obtained by the above camera-based solution. Fig. 5.15 shows that the angular errors are almost independent of the change in the tag-to-antenna distance. At first glance, a larger tag-to-antenna distance will diminish backscatter signal power and then increase the impact of multipath interference on signal fingerprints. However, RF-Orien3D can effectively suppress the multipath impact.

5.9 Conclusion

In this work, we present RF-Orien3D that estimates RFID-tagged object orientation in 3D space only with a two-tag array and one reader antenna. The key novelty is to leverage the 'bad' mutual coupling effect when two tags are located closely. It induces the variation in the radiation pattern of each tag in different directions, thereby providing us an opportunity to find out the spatial angles of the tag array relative to the reader antenna. A novel CNN-assisted method is proposed to deal with strong multipath interference. The experiments have demonstrated the effectiveness of our system. We hope that the techniques of leveraging the change in tag radiation pattern and combating multipath could provide new thoughts for RFID sensing research.

Chapter 6

Conclusion

6.1 Contributions

In this thesis, some existing CV-related techniques are introduced to address a series of challenges in purely battery-free RFID sensing systems. In Table. 6.1, we summarize the proposed computer vision-assisted battery-free RFID systems from the aspects of tasks, challenges, novel algorithms, application scenarios, system performance and key improvements compared to the existing methods. And the detailed contributions of this thesis are described as follows:

1) To reduce the impact of multipath interference and frequency hopping, the proposed RF-Focus is the first CV-assisted RFID system for moving RFID tag recognition and localization within the region of interest, without pre-capturing the entire trajectory of each moving object in CV. Novel RSSI/RF phase-distance models with multipath terms are proposed. On this basis, a dual-antenna solution is adopted in RF-Focus to combat multipath interference, and the tag-to-antenna distance can be extracted from the multipath terms so as to estimate the frequency-dependent phase shifts caused by frequency hopping. RF-Focus achieves about 91.67% recognition accuracy and 94.26% position matching accuracy when tracking 5 moving RFID-tagged objects within the ROI.

2) To deal with the issues of acquiring a reader antenna's position at each sampling time and improving real-time performance in 3D localization, RF-MVO is the first CV-assisted RFID system for stationary RFID localization in 3D space, without driving a robot along a pre-defined trajectory or pre-deployed track. A novel DOA-based 3D localization is designed to avoid exhaustively searching all possible positions. And the HDOP referring to GPS can help RF-MVO output the optimal tag position. RF-Focus achieves about 6.23cm 3D localization accuracy.

3) To cope with the impact of tag mutual coupling and multipath interference, RF-Orien3D is the first CV-assisted RFID system for two-RFID-tag labeled object orientation using CNN with transfer learning technique, without manually collecting large-scale RFID fingerprint measurements. In RF-Orien3D, an RFID tag array is simulated in Matlab for radiation pattern analysis, and an approach based on reported RFID fingerprints in non-coupling and coupling cases is proposed to characterize the variation in practical modulation factor under different tag spacing. RF-Orien3D achieves median errors in azimuth and elevation of about 29° and 11° in both low and rich multipath cases.

Table 6.1 : Summary of computer vision-assisted battery-free RFID systems

	RF-Focus	**RF-MVO**	**RF-Orien3D**
Task	Region-of-interest moving object recognition	Stationary object 3D localization	Multi-tag labeled object 3D orientation
Challenges	1) Unpredictable multipath interference; 2) Frequency-dependent hardware diversity; 3) High matching latency in CV and RFID fusion	1) Moving reader antennas on a pre-defined trajectory; 2) Vast computation complexity in 3D localization; 3) Selection of multiple tag position candidates	1) Unknown tag mutual coupling interference; 2) The change in tag radiation pattern and modulation factor
Algorithms	1) Dual-antenna setup for multipath suppression; 2) Multipath-powered hardware diversity removal; 3) Novel matching algorithm for low-latency searching	1) RFID fingerprint and MVO fusion for 3D localization; 2) Coarse-to-fine tag position and scale factor optimization; 3) HDOP-based optimal position and scale-factor selection	1) Transfer learning-based 3D direction estimation; 2) Tag radiation pattern simulation from a 2D image; 3) Impedance-dependent modulation factor estimation
Applications	Conveyor belt-based applications, e.g., parcel sorting, defective product identification	Robotic pick-up task, mis-shelved book detection in library	Direction-required applications, e.g, liquid shipment and storage
Performance	ROI object recognition accuracy: 91.67%	3D localization accuracy: 6.23cm	3D Orientation accuracy: 29° for azimuth and 11° for elevation
Improvement	1) Reducing multipath interference impact; 2) Compensating frequency-dependent phase offset; 3) Fusing 2D images to minimize position uncertainty	1) Fusing 2D images to estimate unknown antenna trajectory; 2) Reducing computation delay in 3D-space position searching; 3) Introducing HDOP to evaluate localization accuracy	1) Leveraging deep learning to deal with multipath interference; 2) Simulating a tag array from a 2D image for tag characteristic analysis; 3) Only using RFID fingerprints to estimate impedance-dependent modulation factor, without specialized RF network analyzers

6.2 Future Work

The future work will further improve the proposed CV-assisted RFID systems:

CV-assisted ROI Moving Object Recognition. In RF-Focus, a potential problem is how to deal with those ROI RFID tags which are blocked by obstructions. In practice, the placement of each RFID-tagged object may be uncontrolled on the conveyor, so not all ROI RFID tags are in LOS scenarios. For example, in the airport baggage transport, an ROI RFID tag may be located on the bottom of baggage on the conveyor. In this case, the effectiveness of the dual-antenna hypothesis to combat multipath interference in the NLOS will weaken as the lower transmit power arrives at the tag end. And the matching score of an object proposal containing the ROI tag may keep at a low level. As a result, our system may regard this tag as a false-positive reading one located outside the ROI. To minimize the effect of the NLOS issue, our future work is to design a scheme to detect those ROI RFID tags in NLOS, i.e., if the matching scores of an object proposal corresponding to all RFID tags read are all lower than a threshold, it indicates the presence of an ROI RFID tag in NLOS and the corresponding object proposal is the position where the tag is located at. Instead of directly determining the tag as a false-positive reading one, our system is to report the anomalous events and ask sorters to manually deal with the NLOS issue of the tag. In addition, we would deploy multiple reader antenna pairs mounted on a gantry with different positions and directions and select RF phase measurements reported by the antenna pair with the maximum RSSI values for fusion use.

CV-assisted Stationary Object 3D Localization. In RF-MVO, two problems should be considered. *1) Sampling Rate of RF Phase.* To guarantee high-precision tag localization, RF-MVO requires enough fusion samples. However, it is very challenging in a dense tag deployment. One feasible solution is to move reader antennas at a lower speed. Another is to reduce the size of the reading region by using a relatively low transmitted power or a small-size reader antenna like [60]. Moreover, our future work is to design a reading method that improves the tag sampling rate in the main beam of a reader antenna and reduce that in the side beam. *2) Error in Camera Trajectory Recovery.* RF-MVO demonstrates the feasibility of fusing depth-enabled RF phase to estimate scale factors for MVO. However, the RFID tag localization accuracy is subject to the camera trajectory recovery performance (i.e., scale-factor estimation accuracy). In some indoor application scenarios, our future work is to directly recover the trajectory using a stereo or RGB-D camera, which will improve tag localization accuracy but also reduce response latency.

CV-assisted Two-RFID-Tag Labeled Object 3D Orientation. In RF-Orien3D, one drawback is that the azimuth error is around 2.5 times higher than the elevation one. To address this issue, a potential solution is to deploy other parallel tags both horizontally and vertically to build a planar tag array. Then the variation of the tag radiation pattern in azimuth would be similar to that in elevation. Unfortunately, accurately obtaining the modulation factor in this case using the proposed estimation method will be a time-consuming process. In our future work, we would design a faster and more convenient solution without sepa-

rately measuring the RFID fingerprint of each element in a planar tag array in the non-coupling and coupling scenarios.

Bibliography

[1] Monocular visual odometry. https://au.mathworks.com/help/vision/examples/monocular-visual-odometry.html.

[2] Octane sdks for .net. https://support.impinj.com/hc/en-us/articles/202755268-Octane-SDK.

[3] Orb slam2. https://github.com/raulmur/ORBSLAM2.

[4] Ros tool. http://wiki.ros.org/image_view.

[5] Global Standards 1. Regulatory status for using rfid in the epc gen2 (860 to 960 mhz) band of the uhf spectrum. Technical report, 2018.

[6] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics*, 34(6):1–13, 2015.

[7] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *USENIX NSDI*, pages 317–329, 2014.

[8] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-person localization via rf body reflections. In *USENIX NSDI*, pages 279–292, 2015.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. Smart homes that monitor breathing and heart rate. In *ACM CHI Conference on Human Factors in Computing Systems*, pages 837–846, 2015.

[10] Zhenlin An, Qiongzheng Lin, and Lei Yang. Cross-frequency communication: Near-field identification of uhf rfids with wifi! In *ACM MobiCom*, pages 623–638, 2018.

[11] Holger Arthaber, Thomas Faseth, and Florian Galler. Spread-spectrum based ranging of passive uhf epc rfid tags. *IEEE Communications Letters*, 19(10):1734–1737, 2015.

[12] Shadi Ebrahimi Asl, Mohammad Tayeb Ghasr, Maciej Zawodniok, and Kevin Eugene Robinson. Preliminary study of mutual coupling effect on a passive rfid antenna array. In *IEEE I2MTC*, pages 138–141. IEEE, 2013.

[13] Tim Bailey and Hugh Durrant-Whyte. Simultaneous localization and mapping (slam): Part ii. *IEEE Robotics & Automation Magazine*, 13(3):108–117, 2006.

[14] Constantine A Balanis. *Antenna theory: analysis and design*. John wiley & sons, 2016.

[15] Abdelmoula Bekkali, Horacio Sanson, and Mitsuji Matsumoto. Rfid indoor positioning based on probabilistic rfid map and kalman filtering. In *IEEE WiMob*, pages 21–21. IEEE, 2007.

[16] Johann Borenstein and Liqiang Feng. Measurement and correction of systematic odometry errors in mobile robots. *IEEE Transactions on Robotics and Automation*, 12(6):869–880, 1996.

[17] Yanling Bu, Lei Xie, Yinyin Gong, Jia Liu, Bingbing He, Jiannong Cao, Baoliu Ye, and Sanglu Lu. Rf-3dscan: Rfid-based 3d reconstruction on tagged packages. *IEEE Transactions on Mobile Computing*, 2019.

[18] Yanling Bu, Lei Xie, Yinyin Gong, Chuyu Wang, Lei Yang, Jia Liu, and Sanglu Lu. Rf-dial: an rfid-based 2d human-computer interaction via tag array. In *IEEE INFOCOM*, pages 837–845. IEEE, 2018.

[19] Yanling Bu, Lei Xie, Jia Liu, Bingbing He, Yinyin Gong, and Sanglu Lu. 3-dimensional reconstruction on tagged packages via rfid systems. In *IEEE SECON*, pages 1–9. IEEE, 2017.

[20] Stefano Caizzone, Emidio DiGiampaolo, and Gaetano Marrocco. Wireless crack monitoring by stationary phase measurements from coupled rfid tags. *IEEE Transactions on Antennas and Propagation*, 62(12):6412–6419, 2014.

[21] Bekir Sait Ciftler, Abdullah Kadri, and Ismail Güvenç. Iot localization for bistatic passive uhf rfid systems with 3-d radiation pattern. *IEEE Internet of Things Journal*, 4(4):905–916, 2017.

[22] Han Ding, Jinsong Han, Chen Qian, Fu Xiao, Ge Wang, Nan Yang, Wei Xi, and Jian Xiao. Trio: Utilizing tag interference for refined localization of passive rfid. In *IEEE INFOCOM*, pages 828–836. IEEE, 2018.

[23] Han Ding, Chen Qian, Jinsong Han, Ge Wang, Wei Xi, Kun Zhao, and Jizhong Zhao. Rfipad: Enabling cost-efficient and device-free in-air handwriting using passive tags. In *IEEE ICDCS*, pages 447–457. IEEE, 2017.

[24] Daniel M Dobkin. *The RF in RFID: UHF RFID in Practice*. Newnes, 2012.

[25] Chunhui Duan, Xing Rao, Lei Yang, and Yunhao Liu. Fusing rfid and computer vision for fine-grained object tracking. In *IEEE INFOCOM*, pages 1–9. IEEE, 2017.

[26] Chunhui Duan, Lei Yang, Qiongzheng Lin, and Yunhao Liu. Tagspin: High accuracy spatial calibration of rfid antennas via spinning tags. *IEEE Transactions on Mobile Computing*, 17(10):2438–2451, 2018.

[27] Chunhui Duan, Lei Yang, Qiongzheng Lin, Yunhao Liu, and Lei Xie. Robust spinning sensing with dual-rfid-tags in noisy settings. *IEEE Transactions on Mobile Computing*, 18(11):2647–2659, 2018.

[28] David Eberly. Least squares fitting of data. *Chapel Hill, NC: Magic Software*, 2000.

[29] Xiaoyi Fan, Wei Gong, and Jiangchuan Liu. Tagfree activity identification with rfids. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–23, 2018.

[30] Xiaoyi Fan, Fangxin Wang, Feng Wang, Wei Gong, and Jiangchuan Liu. When rfid meets deep learning: Exploring cognitive intelligence for activity identification. *IEEE Wireless Communications*, 26(3):19–25, 2019.

[31] Friedrich Fraundorfer and Davide Scaramuzza. Visual odometry: Part i: The first 30 years and fundamentals. *IEEE Robotics and Automation Magazine*, 18(4):80–92, 2011.

[32] Friedrich Fraundorfer and Davide Scaramuzza. Visual odometry: Part ii: Matching, robustness, optimization, and applications. *IEEE Robotics & Automation Magazine*, 19(2):78–90, 2012.

[33] EPC Gen2. Epcglobal. `www.gs1.org/epcglobal`.

[34] Ross Girshick. Fast r-cnn. In *IEEE ICCV*, pages 1440–1448, 2015.

[35] Enrique González Plaza, Germán León Fernández, Susana Loredo Rodríguez, and Luis Fernando Herrán Ontañón. Calculating the phase center of an antenna: A simple experimental method based on linear near-field measurements. *IEEE Antennas and Propagation Magazine*, 2017.

[36] Joshua D Griffin and Gregory D Durgin. Complete link budgets for backscatter-radio and rfid systems. *IEEE Antennas and Propagation Magazine*, 51(2):11–25, 2009.

[37] Junchen Guo, Ting Wang, Yuan He, Meng Jin, Chengkun Jiang, and Yunhao Liu. Twinleak: Rfid-based liquid leakage detection in industrial environments. In *IEEE INFOCOM*, pages 883–891. IEEE, 2019.

[38] Jinsong Han, Chen Qian, Xing Wang, Dan Ma, Jizhong Zhao, Wei Xi, Zhiping Jiang, and Zhi Wang. Twins: Device-free object tracking using passive tags. *IEEE/ACM Transactions on Networking*, 24(3):1605–1617, 2016.

[39] Lejia Han and John C Bancroft. Nearest approaches to multiple lines in n-dimensional space. *CREWES Research Report*, 22:1–17, 2010.

[40] Sam Hare, Stuart Golodetz, Amir Saffari, Vibhav Vineet, Ming-Ming Cheng, Stephen L Hicks, and Philip HS Torr. Struck: Structured output tracking with kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):2096–2109, 2015.

[41] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *IEEE ICCV*, pages 2961–2969, 2017.

[42] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, 2014.

[43] Impinj. Speedway revolution reader-low level user data support. Technical report, 2013.

[44] Impinj. Indy itk version 1.8.0 release documentation. Technical report, 2016.

[45] Impinj. Optimizing tag throughput using readermode. Technical report, 2019.

[46] iRobot Create 2. `http://www.irobot.com/AboutiRobot/STEM/Create-2.aspx`.

[47] Iso/iec 18000-63 information technology radio frequency identification for item management part 63: Parameters for air interface communications at 860 mhz to 960 mhz type c. Standard, International Organization for Standardization, Oct. 2015.

[48] Vishwanath Iyer and Alex Taylor. Antenna model generation and full-wave analysis from a photo. `https://au.mathworks.com/help/antenna/examples/antenna-model-generation-and-full-wave-analysis-from-a-photo.html`. 2018.

[49] Chengkun Jiang, Yuan He, Xiaolong Zheng, and Yunhao Liu. Omnitrack: Orientation-aware rfid tracking with centimeter-level accuracy. *IEEE Transactions on Mobile Computing*, 2019.

[50] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1409–1422, 2011.

[51] Navjot Kaur, Diana Segura Velandia, William Whittow, David Barwick, Ehidiamen Iredia, Neil Parker, Neil Porter, Paul P Conway, and Andrew A West. Design and performance of a flexible metal mountable uhf rfid tag. In *IEEE ECTC*, pages 2120–2126. IEEE, 2015.

[52] Wadim Kehl, Fabian Manhardt, Federico Tombari, Slobodan Ilic, and Nassir Navab. Ssd-6d: Making rgb-based 3d detection and 6d pose estimation great again. In *IEEE ICCV*, pages 1521–1529. IEEE, 2017.

[53] Deok-Hwa Kim and Jong-Hwan Kim. Effective background model-based rgb-d dense visual odometry in a dynamic environment. *IEEE Transactions on Robotics*, 32(6):1565–1573, 2016.

[54] Rasmus Krigslund, Petar Popovski, Gert Frølund Pedersen, and Kim Olesen. Interference helps to equalize the read range and reduce false positives of passive rfid tags. *IEEE Transactions on Industrial Electronics*, 59(12):4821–4830, 2011.

[55] Richard B Langley. Dilution of precision. *GPS world*, 10(5):52–59, 1999.

[56] Hanchuan Li, Peijin Zhang, Samer Al Moubayed, Shwetak N Patel, and Alanson P Sample. Id-match: A hybrid computer vision and rfid system for recognizing individuals in groups. In *ACM CHI Conference on Human Factors in Computing Systems*, pages 4933–4944, 2016.

[57] Ping Li, Zhenlin An, Lei Yang, Panlong Yang, and QiongZheng Lin. Rfid harmonic for vibration sensing. *IEEE Transactions on Mobile Computing*, 2019.

[58] Qiongzheng Lin, Lei Yang, Chunhui Duan, and Yunhao Liu. Revisiting reading rate with mobility: Rate-adaptive reading of cots rfid systems. *IEEE Transactions on Mobile Computing*, 18(7):1631–1646, 2018.

[59] Jia Liu, Min Chen, Shigang Chen, Qingfeng Pan, and Lijun Chen. Tagcompass: Determining the spatial direction of an object with small dimensions. In *IEEE INFOCOM*, pages 1–9. IEEE, 2017.

[60] Jia Liu, Feng Zhu, Yanyan Wang, Xia Wang, Qingfeng Pan, and Lijun Chen. Rf-scanner: Shelf scanning with robot-assisted rfid systems. In *IEEE INFOCOM*, pages 1–9. IEEE, 2017.

[61] Tianci Liu, Yunhao Liu, Lei Yang, Yi Guo, and Cheng Wang. Backpos: High accuracy backscatter positioning system. *IEEE Transactions on Mobile Computing*, 15(3):586–598, 2015.

[62] David Fernández Llorca, Raúl Quintero, Ignacio Parra, Mario Jimenez, C Fernández, Rubén Izquierdo, and MA Sotelo. Fusing directional passive uhf rfid and stereo vision for tag association in outdoor scenarios. In *IEEE International Conference on Intelligent Transportation Systems*, pages 2527–2532. IEEE, 2016.

[63] Feng Lu, XiaoSheng Chen, and T Ye Terry. Performance analysis of stacked rfid tags. In *IEEE International Conference on RFID*, pages 330–337. IEEE, 2009.

[64] Yunfei Ma, Nicholas Selby, and Fadel Adib. Drone relays for battery-free networks. In *ACM SIGCOMM*, pages 335–347. ACM, 2017.

[65] Yunfei Ma, Nicholas Selby, and Fadel Adib. Minding the billions: Ultra-wideband localization for deployed rfid tags. In *ACM MobiCom*, pages 248–260. ACM, 2017.

[66] Jan Machac, Amine Boussada, Milan Svanda, Jaroslav Havlicek, and Milan Polivka. Influence of mutual coupling on stability of rcs response in chipless rfid. *Technologies*, 6(3):67, 2018.

[67] Kaj Madsen, Hans Bruun Nielsen, and Ole Tingleff. Methods for non-linear least squares problems. 2004.

[68] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.

[69] Raul Mur-Artal and Juan D Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[70] Hyeonseob Nam and Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking. In *IEEE CVPR*, pages 4293–4302, 2016.

[71] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. In *IEEE PerCom*, pages 407–415. IEEE, 2003.

[72] Pavel V Nikitin, Rene Martinez, Shashi Ramamurthy, Hunter Leland, Gary Spiess, and KVS Rao. Phase based spatial identification of uhf rfid tags. In *IEEE RFID*, pages 102–109. IEEE, 2010.

[73] Sida Peng, Yuan Liu, Qixing Huang, Xiaowei Zhou, and Hujun Bao. Pvnet: Pixel-wise voting network for 6dof pose estimation. In *IEEE CVPR*, pages 4561–4570. IEEE, 2019.

[74] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *ACM Mobicom*, pages 27–38, 2013.

[75] Chunfang Qin, Lingfei Mo, Hongliang Zhou, and Hongjian Zhang. Dual-dipole uhf rfid tag antenna with quasi-isotropic patterns based on four-axis reflection symmetry. *International Journal of Antennas and Propagation*, 2013, 2013.

[76] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *IEEE CVPR*, pages 779–788, 2016.

[77] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, pages 91–99, 2015.

[78] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 234–241. Springer, 2015.

[79] Longfei Shangguan and Kyle Jamieson. The design and implementation of a mobile rfid tag sorting robot. In *ACM MobiSys*, pages 31–42. ACM.

[80] Longfei Shangguan and Kyle Jamieson. Leveraging electromagnetic polarization in a two-antenna whiteboard in the air. In *(ACM CoNEXT*, pages 443–456. ACM.

[81] Longfei Shangguan, Zheng Yang, Alex X Liu, Zimu Zhou, and Yunhao Liu. Relative localization of rfid tags using spatial-temporal phase profiling. In *USENIX NSDI*, pages 251–263, 2015.

[82] Longfei Shangguan, Zimu Zhou, and Kyle Jamieson. Enabling gesture-based interactions with objects. In *ACM MobiSys*, pages 239–251. ACM.

[83] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *IEEE CVPR*, volume 2, pages 246–252. IEEE, 1999.

[84] Jianqiang Sun, Lei Xie, Qingliang Cai, Chuyu Wang, Jie Wu, and Sanglu Lu. Rf-isee: Identify and distinguish multiple rfid tagged objects in augmented reality systems. In *IEEE ICDCS*, pages 723–724. IEEE, 2016.

[85] Bugra Tekin, Sudipta N Sinha, and Pascal Fua. Real-time seamless single shot 6d object pose prediction. In *IEEE CVPR*, pages 292–301. IEEE, 2018.

[86] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustmenta modern synthesis. In *International Workshop on Vision Algorithms*, pages 298–372. Springer, 1999.

[87] Tõnu Trump and Björn Ottersten. Estimation of nominal direction of arrival and angular spread using an array of sensors. *Signal Processing*, 50(1-2):57–69, 1996.

[88] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, 2013.

[89] Chuyu Wang, Jian Liu, Yingying Chen, Hongbo Liu, Lei Xie, Wei Wang, Bingbing He, and Sanglu Lu. Multi-touch in the air: Device-free finger tracking and gesture recognition via cots rfid. In *IEEE INFOCOM*, pages 1691–1699. IEEE, 2018.

[90] Chuyu Wang, Lei Xie, Wei Wang, Yingying Chen, Yanling Bu, and Sanglu Lu. Rf-ecg: Heart rate variability assessment based on cots rfid tag array. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–26, 2018.

[91] Chuyu Wang, Lei Xie, Keyan Zhang, Wei Wang, Yanling Bu, and Sanglu Lu. Spin-antenna: 3d motion tracking for tag array labeled objects via spinning antenna. In *IEEE INFOCOM*, pages 1–9. IEEE, 2019.

[92] Fangxin Wang, Jiangchuan Liu, and Wei Gong. Multi-adversarial in-car activity recognition using rfids. *IEEE Transactions on Mobile Computing*, 2020.

[93] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *ACM MobiCom*, pages 288–300, 2017.

[94] Ju Wang, Jie Xiong, Hongbo Jiang, Xiaojiang Chen, and Dingyi Fang. D-watch: Embracing bad multipaths for device-free localization with cots rfid devices. *IEEE/ACM Transactions on Networking*, 25(6):3559–3572, 2017.

[95] Jue Wang and Dina Katabi. Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 51–62. ACM.

[96] Jue Wang, Deepak Vasisht, and Dina Katabi. Rf-idraw: virtual touch screen in the air using rf signals. *ACM SIGCOMM Computer Communication Review*, 44(4):235–246, 2014.

[97] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and modeling of wifi signal based human activity recognition. In *ACM MobiCom*, pages 65–76, 2015.

[98] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *ACM MobiCom*, pages 617–628, 2014.

[99] Yanwen Wang and Yuanqing Zheng. Modeling rfid signal reflection for contact-free activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–22, 2018.

[100] Yanwen Wang and Yuanqing Zheng. Tagbreathe: Monitor breathing with commodity rfid systems. *IEEE Transactions on Mobile Computing*, 19(4):969–981, 2019.

[101] Yuxi Wang, Kaishun Wu, and Lionel M Ni. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594, 2016.

[102] Zhongqin Wang, Fu Xiao, Ning Ye, Ruchuan Wang, and Panlong Yang. A see-through-wall system for device-free human motion sensing based on battery-free rfid. *ACM Transactions on Embedded Computing Systems*, 17(1):1–21, 2017.

[103] Zhongqin Wang, Min Xu, Ning Ye, Ruchuan Wang, and Haiping Huang. Rf-mirror: Mitigating mutual coupling interference in two-tag array labeled rfid systems. In *IEEE SECON*, 2020.

[104] Zhongqin Wang, Ning Ye, Reza Malekian, Fu Xiao, and Ruchuan Wang. Trackt: Accurate tracking of rfid tags with mm-level accuracy using first-order taylor series approximation. *Ad hoc networks*, 53:132–144, 2016.

[105] Teng Wei and Xinyu Zhang. Gyro in the air: tracking 3d orientation of batteryless internet-of-things. In *ACM MobiCom*, pages 55–68. ACM, 2016.

[106] Fu Xiao, Zhongqin Wang, Ning Ye, Ruchuan Wang, and Xiang-Yang Li. One more tag enables fine-grained rfid localization and tracking. *IEEE/ACM Transactions on Networking*, 26(1):161–174, 2018.

[107] Lei Xie, Jianqiang Sun, Qingliang Cai, Chuyu Wang, Jie Wu, and Sanglu Lu. Tell me what i see: Recognize rfid tagged objects in augmented reality systems. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 916–927, 2016.

[108] Huatao Xu, Dong Wang, Run Zhao, and Qian Zhang. Adarf: Adaptive rfid-based indoor localization using deep learning enhanced holography. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):113, 2019.

[109] Chao Yang, Xuyu Wang, and Shiwen Mao. Sparsetag: High-precision backscatter indoor localization with sparse rfid tag arrays. In *IEEE SECON*, pages 1–9. IEEE, 2019.

[110] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices. In *ACM MobiCom*, pages 237–248. ACM, 2014.

[111] Lei Yang, Yao Li, Qiongzheng Lin, Huanyu Jia, Xiang-Yang Li, and Yunhao Liu. Tagbeat: Sensing mechanical vibration period with cots rfid systems. *IEEE/ACM transactions on networking*, 25(6):3823–3835, 2017.

[112] Lei Yang, Yao Li, Qiongzheng Lin, Xiang-Yang Li, and Yunhao Liu. Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals. In *ACM MobiCom*, pages 16–28, 2016.

[113] Lei Yang, Qiongzheng Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. See through walls with cots rfid system! In *ACM MobiCom*, pages 487–499, 2015.

[114] Yanni Yang, Jiannong Cao, and Xiulong Liu. Er-rhythm: Coupling exercise and respiration rhythm using lightweight cots rfid. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4):1–24, 2019.

[115] Pengyu Zhang, Jeremy Gummeson, and Deepak Ganesan. Blink: A high throughput link layer for backscatter communication. In *ACM MobiSys*, pages 99–112, 2012.

[116] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[117] Run Zhao, Dong Wang, Qian Zhang, Haonan Chen, and Anna Huang. Crh: A contactless respiration and heartbeat monitoring system with cots rfid tags. In *IEEE SECON*, pages 1–9. IEEE, 2018.

[118] Yiyang Zhao and Lionel M Ni. Vire: Virtual reference elimination for active rfid-based localization. *Adhoc & Sensor Wireless Networks*, 17, 2013.

[119] C Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, pages 391–405. Springer, 2014.

[120] Danping Zou and Ping Tan. Coslam: Collaborative visual slam in dynamic environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):354–366, 2012.

[121] Yongpan Zou, Jiang Xiao, Jinsong Han, Kaishun Wu, Yun Li, and Lionel M Ni. Grfid: A device-free rfid-based gesture recognition system. *IEEE Transactions on Mobile Computing*, 16(2):381–393, 2016.