UNIVERSITY OF CALIFORNIA
SANTA CRUZ

TOWARDS RFID-BASED CYBER-PHYSICAL CONNECTION SYSTEMS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER ENGINEERING

by

Haofan Cai

June 2023

The Dissertation of Haofan Cai
is approved:

Professor Chen Qian, Chair

Professor Roberto Manduchi

Professor Yang Liu

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Towards RFID-based Cyber-Physical Connection Systems

by

Haofan Cai

With the advances of new technologies such as Radio-frequency identification (RFID) systems, pervasive sensing, and cyber-physical systems (CPS), the Internet-of-Things enables inter-connection of small objects (such as tagged items, embedded systems and mobile devices) and data collection, delivery, as well as processing among them. This thesis focuses on integrating large-scale and low-cost passive RFID tags into existing pervasive-sensing applications to enable multi-functional and cost-efficient IoT systems. We believe the inherent advantages of RFID systems to identify, trace, and track information using easily deployable tags provide unique opportunities to enable many novel IoT applications in new areas of sensing, actuation, and user interaction, which is far beyond its traditional use in supply chain management. However, the design of these applications involves challenges, as the limited computation ability and simple functionality of passive tags may make them ill-fitted for meeting the diverse requirements. To tackle these issues, this thesis takes a deep look at exploring the potential of commercial off-the-shelf (COTS) passive tags, and proposes new applications in the following themes:

- Enabling cyber-physical connection using RFID passive tags, where system can

seamlessly detect user-item interaction and gather information from real-world subjects.

- Fusing the information from computer vision and RFID sensing channel to actively find/track the mobile object/person with least training efforts.
- Addressing scalability issue and domain shift challenge in human gesture recognition with RFID via employing domain-adaptive few-shot learning (DA-FSL).

Acknowledgments

When I look back towards my journey as a PhD student, I feel very lucky to be around such good people without whom I would not be able to achieve anything. I feel that I was very lucky to be able to become part of this prestigious university and be able to work with such intelligent and wonderful people. I am really thankful to the entire UCSC community for providing such an awesome research experience and in making the difficult jobs easy for everyone.

I would like to express my gratitude to Prof. Chen Qian for his expert advice, continuing guidance, and strong encouragement during my doctoral studies. The numerous discussions in all our academic works have greatly helped me enhance my academics, guided me on publications, and assisted me in achieving my professional goals. I am very grateful for having received such valuable advice and guidance from Prof. Qian throughout my doctorate, and I will always be indebted to him for his teaching and training.

I would also like to thank the entire Prof. Chen Qian's group, both current and previous members. During my doctorate, I am grateful to have worked with these excellent colleagues and collaborators on multiple topics. I would specially like to mention Ge Wang who has worked with me since the beginning of my first work and has been a part of all the ups and downs during my PhD study. She taught me a lot about RFID system during my initial years at UCSC and I am very grateful to her for that. I would also like to express my thanks to Xiaofeng Shi, Minmei Wang, Junjie Xie and

Shouqian Shi for their help in the experiments. The successful implementations of all systems in this thesis are the result of teamwork. On the research topic of ChopTags, I would like to express my many thanks to Prof. Shigang Chen for his strong support in this research collaboration. This work achieved significant success also thanks to his support.

I also want to thank the Computer Science and Engineering department in arranging and teaching such diverse and in-depth curriculum. I especially like to thank Prof. Roberto Manduchi for teaching computer vision which helped me a lot in my research work, and Prof. Yang Liu for teaching data science, which introduced me to the wonderful world of data science and machine learning. I am also thankful to Prof. Roberto Manduchi, Prof. Shigang Chen and Prof. Yang Liu for their encouraging comments during my candidacy. I got many interesting ideas from the discussions and also got to expand my thinking about the project.

I would like to thank my family, my father Zhihui Cai, my mother Liping Wang and my boy friend Rongwen Zhao who always loved me and encouraged me to follow my dreams. They always took care of all my needs so that I could focus on my studies. I would have never made it without their help.

Finally, I would like to express my thanks to the administrators and faculty members at UC Santa Cruz. The UCSC community has significantly helped me with knowledge understanding and the needed learning advancement throughout my doctoral studies.

Thank you all for being part of my life!

Chapter 1

Introduction

This chapter introduces Radio-frequency identification (RFID) system and its applications in the Internet-of-Things (IoT). It highlights the design challenges of adopting RFID technique in different IoT applications. We give an overview of the research presented in this thesis and outline the author's contributions.

1.1 Background

Thanks to the rapid evolution in chip design and wireless communication technologies, sensors and actuators are currently cheap enough to be embedded in any device (1). The Internet of Things (IoT) emerged with the objective of providing new intelligent services and commodities to facilitate our daily tasks (1; 2). IoT visualizes a completely connected world, where things are able to communicate and interact among each other. In this context, one of the most widely used technologies in the IoT is RFID. RFID is one of the best technologies to perform identification, which has gained a lot of

popularity in applications like access control, payment cards, or logistics (1) in the last decades. Other fields where RFID has a known implication are health care (3), animal identification (4), and the supply chain (5).

A typical RFID system includes RFID tags, a RFID reader and back-end systems (Fig. 1.1,). A tag is a small electronic device made up of antenna, coupling component, and microchip. Enclosed in an adhesive sticker, every tag is attached to an object with a unique identifier (UID). RFID reader, usually connected to antennas, emits radio waves to detect and interact with the tags, capturing their data. The collected data is then sent to a backend server, where it is stored and processed for further real-time monitoring and analysis. In fact, according to the power supply mode, an RFID tag can be categorized as either a passive or active tag (6). A passive tag with small size and low cost has no onboard power supply, its operating energy is from the continuous wave transmitted by the reader. When a reader emits radio frequency signals, nearby passive RFID tags within its range respond by backscattering the incoming radio waves to transmit their stored data to the reader. This bidirectional communication is called backscatter communication. Thus, the transmission distance of passive tags is quite limited. In contrast, an active tag has an internal battery to provide energy for the microchip and ensure communication between tag and reader. The potential transmission range can thus reach several hundred meters. However, the production cost is high, and the service lifetime is short because the battery needs to be periodically replaced. In the following parts of this thesis, we only focus on integrating **passive RFID tags** into existing pervasive-sensing applications to enable novel IoT systems. The research

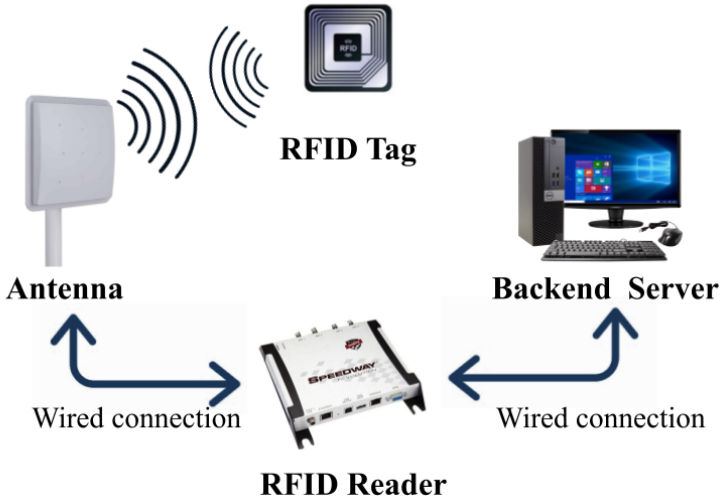


Figure 1.1: RFID system

of active tags is beyond the scope of this thesis.

In fact, RFID serves as a crucial bridge connecting the realms of ***cyber-space*** and ***physical-space***. By enabling the wireless exchange of data between objects and computer systems, RFID facilitates seamless communication and integration in these two domains. By attaching a RFID tag to a physical object, such as a product or living beings, object-related information can be collected and processed in real-time, creating a digital representation of the physical world. We refer this convergence between the virtual and tangible domains as the ***cyber-physical connection***. Moreover, the ability of RFID to identify, trace, and track information using easily deployable tags is now enabling innovative applications that bridge the gap between the cyber and physical space beyond its traditional use in supply chain management or access control: it is now employed in new areas of sensing, actuation, and even user interaction (6). However, the design of these applications involves challenges, as the limited computation ability and

simple functionality of passive tags may make them ill-fitted for meeting the diverse requirements. For example, transforming RFID system with identification capability into human sensing and tracking platform is quite difficult, as the backscattered signal is sparse and coarse, where the useful feature can hardly be extracted.

1.1.1 Research Challenges

The idea of employing RFID in new applications of sensing, actuation, and user interaction involves challenges, as the limited computation ability and simple functionality of passive tags may make them ill-fitted for meeting the diverse requirements. We identify 3 main obstacles that prevent RFID sensing from becoming pervasive adoption in different IoT applications, detailed in the following:

Limited energy harvester and read range: These are considered to be two of the most important limitations because RFID tags are made of scarce resources (7). Existing RFID platforms implemented in industrial systems are mostly passive, indicating that tags cannot operate or sense data without being placed inside the reader's reading zone. The integrated circuit (IC), the microcontroller unit, and the antenna module on a passive tag are powered by harvesting the RF energy transmitted by the reader, and communicate by backscattering the incident signal. This implementation reduces the manufacturing cost by keeping IC costs low. However, the long-range communication and power hungry sensing capabilities will be limited by the power available at the tag.

Lack of scalability: The current RFID protocol uses slotted ALOHA as

the MAC layer solution. Hence, tags compete with the time slots to reply to the reader. The commodity RFID readers support a constant number of successful read operations per second (around 400 on our device) regardless of the number of tags with the interrogate range. When the number of tags increases, the share of time slots of every tag decreases. We define individual reading rate (IRR) as the number of readings obtained from a particular tag per second (8). For RFID-based sensing systems (8; 9), to achieve high sensing accuracy, the IRR of a tag should be sufficiently high, as the collected wireless signal not only contains information of the performed gesture but also carries substantial information of environment. Missing RSS and phase samples will consequently reduce the accuracy of sensing tasks.

Lack of flexibility: Current RFID tags are small electronic devices made up of antennas and microchips. Once they are manufactured, they cannot be replaced or reconfigured without a costly redesign and reproduction. Since the IoT is an open, dynamic, and versatile global networking and sensing system, the generality, modularity, and reconfigurability of the sensing nodes/platform are essential for the their adoption in the future IoT architecture. Furthermore, commercial RFID readers are generally black box systems that only allow limited configuration (1) and are only capable of implementing the current UHF RFID communication standard named EPCglobal Class 1 Generation 2 (EPC C1G2) (8).

1.1.2 Research Contributions

Taking the aforementioned issues into consideration, the major contributions of this dissertation are outlined in the following contents.

Low-Cost and Convenient Tag Mutual Identification. TagMii (8), a new approach to enable user–item interactions using passive RFID tags is proposed. Compared with other solutions that require a user to carry non-trivial hardware, TagMii requires every user to carry only a passive tag. The reader deployed in the environment monitors the interaction events and pairs the user tag and corresponding item tag. The key advantage of TagMii is that it is cost-efficient and especially convenient for children, some seniors, people with certain disabilities, and others who do not operate smartphones. TagMii is evaluated in complex environments with rich multi-path, mobility, wireless signals, and magnetic fields, and it is found to be accurate in recognizing user–item interactions in various setups.

Accurate and Low-cost Interface to Identify User/Item Interactions. This work proposes ChopTags (10), which is a novel solution for tag interaction identification developed with COTS passive RFID tags, that may enable many ubiquitous computing applications that require the accurate recognition of user/item interactions. ChopTags is the first to combine the information of both tag ID presence and physical signal features to infer interactions. ChopTags achieves near 100% accuracy and only requires every user or item to carry a passive tag. We implement two application prototypes based on ChopTags and evaluate the prototypes in complex environments. The

results show that ChopTags is highly accurate and reliable with low latency.

Identity Tracking by RFID-Vision Fusion with Small Training Cost. This work presents RVTrack (11), a vision-RFID fusion system which can enable cameras to recognize the physical IDs of persons in their views and track the targets instantly when they appear in the video frames. By asking the users to perform a simple authentication, the system will be aware of the targets' IDs in its FoV. Later by exploiting the target features extracted from a light-weight Re-ID model, it can handle the PID-switch problems which often occur in vision tracking systems. Experiments show that RVTrack can actively identify and track the RFID-tagged target with an accuracy of 90.14%.

Domain-independent Few-Shot Recognition. In (12), we explore the possibility of employing domain-adaptive few-shot learning (DA-FSL) to address the scalability issue and domain shift challenge and present RF-HGR, a novel RFID-based HGR system to recognize unseen gestures in different domain settings. RF- HGR is a lightweight cross-domain sensing approach in which users only need to collect a few signal samples (i.e., few shots) for any unseen class, and the model only needs to be fine-tuned using these samples. Our experiments evaluate RF-HGR with the dataset collected from real-world experiments. The results show that the system is capable of producing high accuracy models for the HAR task. As far as we know, this is one of the first attempts to jointly address the few-shot DA and few-shot recognition problems in device-free HGR.

1.1.3 Dissertation Organization

The rest of this dissertation is organized as follows.

Chapter 2 provides the details of the proposed TagMii approach. A number of user-item interactions and user-user interactions applications that can utilize TagMii to perform simple yet important tasks, which was difficult to achieve in existing RFID systems, are also presented in this paper.

Chapter 3 presents the design of the ChopTags. The practicability and robustness of ChopTags are demonstrated by two prototypes called TagChess and TagArray. The TagChess system can automatically record and take notations of a chess game by identifying every move, while TagArray can support multiple concurrent readings. The performances of these ChopTags-based applications are also presented.

Chapter 4 details the system of RV-Track. Two datasets are collected and manually labeled for detecting user authentication and tracking. RV-Track is also compared with several state-of-the-art Re-ID network to demonstrate its superiority.

Chapter 5 details the design of RF-HGR, and disscuss the future direction for this work.

Chapter 6 provides the overall summary and conclusions of the dissertation.

Chapter 2

TagMii

2.1 Introduction

Passive Radio Frequency Identification (RFID) technology, widely considered as a cheap, energy-efficient, and scalability solution of the Internet of Things (IoT), has been deployed for ubiquitous applications, such as retailing, warehouse, transportation, and manufactures. The basic functions of RFID that have been extensively studied including: 1) Collect the identity information of tagged items or users in a target area (13); 2) Count or estimate the population of items/users (14; 15; 16); 3) Advanced sensing tasks that explore the physical signal features of RFID, such as localization and tracking (17; 18; 19; 20; 21; 22; 23; 24; 25; 26; 27), and human activity sensing (28; 29; 30; 31).

However, current passive RFID applications are mainly restricted to industrial and logistic areas, due to a *fundamental limitation*: an individual user is difficult to

interact with tagged items or other tagged users, if she carries no reader. The interactions with tagged items may include 1) collecting certain information from the digital profile behind a tag, such as the item name, price (for merchandise), packing location and time (for packages); 2) requesting the backend system to connect the user profile with the item profile and conduct certain processing, such as charging this item to the user account. Hence passive RFID is seldom used in consumer applications.

Recent efforts have been made to improve the consumer-experience of interacting with passive tags, such as small reader (32), smart-phone based RFID scanner (33), and reading tags through the smartphone WiFi interface (34). Also smartphones may have NFC interfaces. However these methods still require non-trivial hardware carried by the users – a smartphone at a minimum. As a result, children, some seniors, people with certain disabilities, and other people who are not familiar of operating smartphones cannot utilize these methods. Moreover, having a reader embedded in the mobile device such as smartphone or reading tags through WiFi interface all require complex hardware modifications and increase the cost of deploying the system.

This work provides an extremely low-cost and convenient solution for user-tag interaction by answering this question: *Can a user initiate the interaction with tagged items with nothing but a single passive tag?* The proposed solution is called TagMii (Tag Mutual Identification Interface). TagMii allows two tags 'read' each other: If two tags are placed in a close physical proximity, the reader can identify these interacting tags among many other tags in the environment. The two tags can be the identifiers of a user and an item, or two users. Compared with current RFID applications (14; 15; 16) which

(a) Application 1: artworks display in the museum

(b) Application 2: shopping in cashier-free retail store

Figure 2.1: Two applications of TagMii

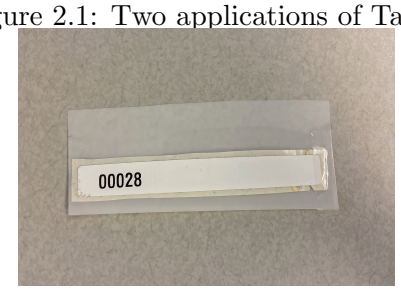


Figure 2.2: TagMii User tag (attached on cardboard)

require user to carry a cumbersome reader to interact with each tag, TagMii only needs to deploy one commercial RFID reader in the environment, which significantly lowers the cost as well as eases the maintenance. One reader with multiple cabled antennas in a room can serve many users (visitors, customers, etc.) and the cost of adding more users is minimal: only passive tags. We present a number of IoT applications that can utilize TagMii to perform simple yet important tasks, which was difficult to achieve in existing RFID systems:

Application 1. A museum displays a collection of artworks. A tag is placed

on the wall next to each artwork. Each user wears a set of wireless headphones. If a user wants to listen to the commentary of an artwork, she puts her user tag close to the tag of the artwork. The backend system then sends the recorded commentary to her headphones, in the preferred language according to the user profile.

Application 2. In a cashier-free retail store, a tag is attached to each type of items. When a user puts her user tag close to an item tag, she can hear the item price, decryption, and reviews from her headphones. She may repeat this operation to request the backend system to put the item in her virtual shopping cart. Warehouse workers put the items that she interacted into a cart at the checkout for her to pick up and charges are made to her account.

Application 3. In a conference, two attendees chatting during a coffee break may put their user tags together. An app running in the backend system will automatically exchange their digital profile (e-business card) and record the time and location. After the conference, the collected digital profiles will be emailed to the user and she can easily review the profiles from the attendees who she has chatted to.

Some existing technologies may also be able to accomplish the tasks in the above applications. For example, a customer can use her smartphone to scan the barcode (35) or QR code (36) attached to the items to retrieve the item information and/or request the payment. Some mobile apps support e-business card exchange. **Compared to these solutions, the key advantage of TagMii is that it is smartphone-free and hence very convenient for children, some seniors, people with certain disabilities, and others who do not operate smartphones.** Other advantages

of TagMii include: 1) Easy to carry. The user tag can be attached to a conference badge (or museum badge, shopping badge in a similar design). 2) Fast interaction. Using smartphones for code scanning or profile exchange usually takes half a minute. TagMii may only take 10 seconds. Such speed advantage is especially important in busy situations like conference breaks. 3) TagMii does not require light conditions compared to code scanning, suitable for museums and warehouses. 4) System manager can easily obtain the useful information monitored by the backend, such as the popular goods, shopping habits of users, and preference of artworks by users with different ages and regions. One might notice that TagMii still requires a reader placed in each room. However this method is scalable: one reader per room can monitor multiple user tags and item tags.

The design of TagMii is based on two important ideas: inductive coupling and phase profile similarity. We observed from experiments that when two tags are placed in physical proximity (e.g., $<$ 2cm), the backscatter signal from either tag would be different from the signal by putting the tag alone, called the *inductive coupling state*. A sudden change of the backscatter signal strength from both tags will occur at the beginning of inductive coupling. TagMii tracks such signal strength change to find out the potential tags that may be in the coupling state, among all tags in the environment. In addition, to further select the pair of tags that are truly in the coupling state. We utilize the tags' *phase profiles* collected by multiple antennas based on the fact that nearby RFID tags experience a similar multipath environment and thus exhibit similar phase profiles. By evaluating and comparing the similarities between phase profile of

potential coupling tags, TagMii can accurately identify the tags placed together and therefore recognize the user and the target item that is interacting with her.

One may ask why we do not simply adopt the existing RFID tag localization protocols (9; 20; 21; 22) to localize all tags, and consider any two tags in very close positions are interacting tags. The main reason of not using this approach is its *false positives*. To identify the interacting tag pair, we need to not only localize the user tag and item tag accurately, but also be able to identify and differentiate different users in the space if more than one user exists. Existing tag localization methods, if using commodity readers, are not able to achieve location errors $<$ 10cm (21; 22). Hence it is likely that two tags within 10-20cm distance are localized to a same position. These tags are not necessarily the user tag and item tag that are interacting. Though PinIt(9) can locate the tagged object with a median accuracy of 11cm, however it requires the usage of sophisticated SDR device for localization, and the antenna need to be moved back and forth during the experiment.

In this work, we address and resolve three main challenges in design and implementation of TagMii. 1) We design accurate tag coupling detection algorithm for **complex environments with rich multipath, mobility, wireless signals, electrical devices, and magnetic fields**. 2) TagMii should identify the coupling tags in a short time even though there are a large number of tags in the environment. 3) TagMii should be able to extract sufficient information from signal dynamics using commodity RFID readers, which have low measurement resolution.

We implement TagMii using commodity off-the-shelf (COTS) RFID devices

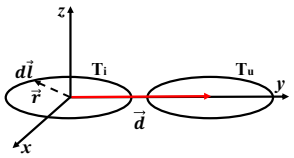


Figure 2.3: Model of two coupling tags

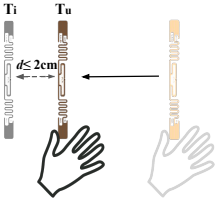


Figure 2.4: Inductive coupling

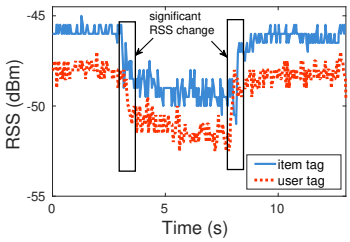


Figure 2.5: Signal strength change of two coupling tags

only. Although we are the first to study mutual identification (interacting a tag with another tag), we do *not* conduct our study in a laboratory condition that is free of multipath reflectors and moving objects. In fact, all implementation and experiments are conducted in various complex environments with rich multipath, mobility, wireless signals, electrical devices, and magnetic fields, in order to valid TagMii for practical applications. Even in these environments, TagMii provides high accuracy. We believe TagMii serves as an important extension of current RFID applications and tag mutual identification will attract further research due to its low-cost and convenience to enable consumer experience of interacting with tagged items.

The remaining paper is organized as follows. Section 2.2 presents the related work. Section 2.3 introduces the background and models. We present the detailed design of TagMii in Section 4.3. We show the system implementation and evaluation results in Section 2.5. We provide some discussion in Section 2.7 and conclude this work in Section 4.6.

2.2 Related Work

User-item interaction: Automatic identification and data capture (AIDC) techniques like barcodes (35) and quick response (QR) codes (37) provide fast, easy and accurate data collection approaches for inventory control, product management (38). The barcode (35) is an optical machine-readable representation of data relating to the object to which it is attached. It systematically represented data by varying the width and spacing of parallel lines.The Quick response (QR) (39) code is a two-dimensional barcode developed to improve the reading speed of complex-structured 2D barcodes. It provides a very convenient way to display information, such as a URL, which can be easily scanned and processed by mobile devices (37). A barcode or QR code requires to be read by a scanner or camera. Different from them, RFID does not require such line-of-sight scanning as RF signals can penetrate through non-metal objects (40). RFID sensing techniques have been widely used for detecting the gesture-based inputs (41; 42). IDSense(43) recognizes the physical movements and touch events of everyday objects using real-time classification of the RSSI and phase angles with only one reader, however, it requires training and calibration and is insufficiently precise. PaperID (44) is a similar work that uses supervised machine learning to detect different types of on-tag and free-air interactions with custom-designed RFID tags. And Pradhan *et al.* (45) show how changes in the received signal phase caused by touching on RFID tag can be leveraged to detect the finger swipe or touch gesture without any pre-training stage. (46) proposes and designs a passive, thin metal tag called LiveTag that can be printed on paper-like

substrates and attached on objects. When being touched by fingers, LiveTag will disturb ambient WiFi channel in a deterministic way, hence making such interaction observable by the system. However it requires elaborate tag design and implementation. There also exists a research direction trying to solve the issue that users have to use RFID readers to query tags, and thus,cannot benefit from the convenience provided by their mobile devices. However, as mentioned in Section 3.1, individual users who have no reader are difficult to interact with tagged items. TiFi (34) allows a 2.4 GHz WiFi receiver to identify 800 MHz UHF RFID tags. This system leverages the underlying harmonic backscattering of tags to open a second channel and uses it to communicate with WiFi receivers. Other related techniques include small RFID reader (32) and smart-phone based RFID scanner (33). However, all of them require the user to carry a smartphone in minimum. Compared with these works, TagMii only requires each user to carry a tag and/or other necessary feedback devices of the application (such as headphones in the museum commentary case).

Recent studies have also proposed to study tag to tag interactions. RFIBricks (47) is a building-block system which symmetric 2D patterns of RFID contact switches are deployed on the top and bottom of each block. Such design enables the backend system to recognize which block is stacked on which, as well as the stacking orientation. RFIMatch (48) detects finger-touching on a tag based on the correlated state change between the tag and an RFIMatch fingerstall worn by user. Tip-Tap(49) recognizes 2-dimensional discrete touch events by sensing the intersection between two arrays of RFID tags. However, these works require physically modifying the tag circuitry and

detecting the correlated state changes of electromechanical sensors in the tag circuitry caused by human behaviors. Trio(50) mainly focuses on the refined localization in a small area using RSS, which is fragile when being applied in large dynamic space.

RFID Localization: Using RFID for Localization is a possible research direction for TagMii, considering any two tags in very close positions are interacting tags. Prior work can be divided into two categories: absolute localization and relative localization. Early RFID-based localization methods mainly rely on RSSI information to acquire tag location. They usually rely on densely pre-deploying tags in the area of monitoring and leverage the pre-collected RSSI values of these tags as references to locate a specific tag (17; 51; 52). The major limitation of RSSI-based methods is that they are highly sensitive to multi-path propagation, and thus can not achieve high-accuracy localization in multi-path environment. Meanwhile, there is a growing interest in using phase differences (21; 42; 53; 54) or Angle of Arrival(AoA) information (20) to estimate absolute locations of tags. Localization using advanced electromagnetism and communication techniques like synthetic aperture radar (SAR) (9) or multiple antennas (22) have also been studied. However, simply adopting the existing RFID tag localization protocols (9; 20; 21; 22) to localize all tags in TagMii will not work, as existing tag localization methods using commodity readers are not able to achieve location errors $<$ 10cm (21; 22). Hence it is likely that two tags within 10-20cm distance are localized to a same position. These tags are not necessarily the user tag and item tag that are interacting.

2.3 Background and Model

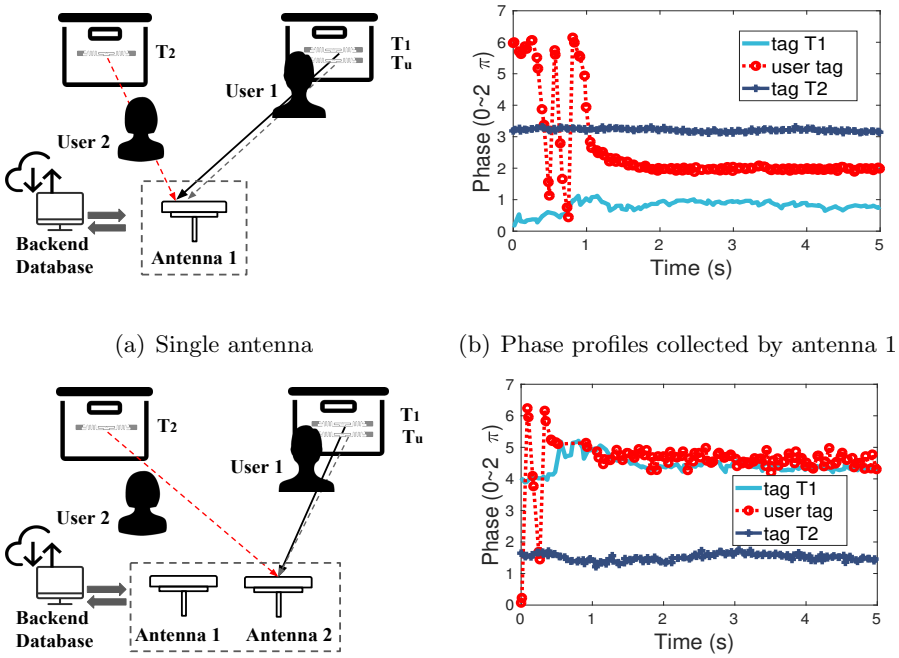
2.3.1 Problem specification

TagMii targets on making convenient user-item interaction using RFID tags. In TagMii, we employ two types of tags, the first type are called *user tags* and the other are *item tags*. Each user registered to TagMii holds a user tag, and each item (or each type of items) in the application is labeled by an item tag. When a user wants to interact an item, she puts her user tag in a close distance (< 2cm) to the corresponding item tag.

TagMii can be completely implemented with COTS RFID devices. It requires *no modification* on current RFID tag or reader hardware. It is implemented as a software program on the backend server connected to the reader, which stores the profile information of all users and tags that could appear in the system. TagMii analyzes the physical-layer signals collected by the reader antennas from all tags in the environments and determines the interacting ones.

2.3.2 Basic ideas

TagMii is based on two important facts and observations from practical RFID communication: *inductive coupling* and *channel similarity*. If two tags are put in a physical proximity, they will interfere with each other and cause changes on their backscatter signals. This phenomenon is called inductive coupling(50)(55). **Our key innovative idea is to use the occurrence of the coupling phenomenon as an indicator of**



(a) Single antenna

(b) Phase profiles collected by antenna 1

(c) Multiple antennas

(d) Multiple antennas

Figure 2.6: Phase profiles and differences captured by multiple antennas

user-tag interaction, which can be detected by analyzing the signals collected by the reader. One additional challenge is that, when users are moving their bodies to 'scan' the item tags using their user tags, the body movement may cause signal changes on other tags in the environments that are not involved in this interaction. By monitoring the signal variations, we can find out a list of tags called *candidate tags* that may be coupling with the user tag. To figure out which candidate tag is indeed the *target tag* that is being read by the user, TagMii further uses the *channel similarity* between the user tag and those candidate tags. Channel similarity is based on the observation that two coupling tags will show high similarity on their received phase information. In fact, the wireless channel is determined by a series of factors, including tags' locations, out-

side environment, and moving objects. Only the target tag, which actually has a close distance with the user tag will experience similar communication channels and shows high similarity on their received signal phases.

2.3.3 Inductive coupling

In the environment, we may have dozens of or even hundreds of tags that represent different items and people. Each of them may be the one that the user is interacting. It is extremely time-consuming and compute-intensive for our system to compare all the tags' signals with that of the user tag. Fortunately, we observe that when the user tag is close to the item tag, there is a sudden decrease on both of their received signal strength (RSS)(56) . We specify this phenomenon with both theoretical model and practical experiments.

In theory, a passive RFID tag can be modeled as a circular loop. If there is no user tag, for each item tag T_i, a steady current I_i will be induced by the electronic waves send by the RFID reader. After transmitting in the wireless channel, which has a channel parameter of $h_{t\leftrightarrow R}$, the received signal strength RSS_i of the tag T_i should be:

$$RSS_i = h_{t\leftrightarrow R} \cdot 10 \cdot log_{10}(1000 \cdot P_i), \tag{2.1}$$

where P_i is the power of tag T_i and U is the voltage only related to the reader's signals:

$$P_i = I_i \cdot U. \tag{2.2}$$

However, when a user tag is placed close enough (Fig. 2.4), *i.e.*, 2cm away from the item tag, the RSSes of the two tags will significantly change. As shown in Fig. 2.3, the

user tag and the item tag can be modeled as two circular loops, whose radius vectors are $\overrightarrow{r}$. Let vector $\overrightarrow{d}$ denotes the directional vector from the center of item tag's circular to that of the user tag. Once the RFID reader sends an electronic signal, it will induce a steady current of I_i to tag T_i. The steady current will generate a magnetic field around it. The mutual inductance M between item tag T_i and the user tag T_u can be calculated by:

$$M = \frac{\mu_0}{4\pi} \oint_c \frac{1}{|\vec{d} - \vec{r}|^2}. \tag{2.3}$$

Due to the influence of the magnetic field, the current I_u in tag T_u changes to I'_u:

$$I'_u = I_u - \frac{1}{R_u} \cdot \frac{dM}{dt} \cdot I_i, \tag{2.4}$$

where R_u is the resistance of user tag T_u. Observing Eq. 2.4 and 2.3, we find that both the resistance R_u and mutual inductance M are positive values. In other words, we have a conclusion that:

$$I'_u < I_u \tag{2.5}$$

The situation is similar for the item tag T_i, *i.e.* $I'_i < I_i$. In other words, for a stable channel, the received signal strength RSS_i and RSS_u for these two tags will decrease due to the inductive coupling.

We also conduct a set of experiments to verify the modelling results. We first put one tag T_u in Fig. 2.4 with a distance of 2cm from a static tag T_i. After keeping the two tags stable for two seconds, we take away tag T_u. The signals of the two tags are shown in Fig. 2.5. Obviously, due to inductive coupling, both tags experience significant decreases on their RSSes when they are putting together. During the coupling

their RSSes keep a lower values than those before and after.

By observing the sharp and simultaneous decreases of their RSSes, TagMii can find out a few candidate tags, whose population is much smaller than the entire set of tags. Then finding the actually coupling tags from the candidate tags is much easier and time-efficient. In Section 4.3, we will specify the detailed algorithm to select the candidate tags.

2.3.4 Channel similarity

In fact, only observing the decrease of the RSS can help to select the candidate tags but may not exactly figure out the actual coupling tags. That is because the RSSes of the other tags will also be influenced by many factors. Among these factors, environmental dynamics and user body's movement may be two crucial ones. When a user is continuously moving around a tag, the received signal will exhibit an amplitude increasing/decreasing fluctuation. Obviously, when the moving user blocks the line-of-sight (LOS) propagation path between a tag and the reader antenna, the tag's RSS will also experience a decrease, and such fluctuations may eventually increase the false possibility. In addition, if multiple users are using their user tags to scan different item tags, all of the involving tags will fall in the coupling state with RSS decrease. To resolve this problem, we further compare the phase data of the candidate tags. The basic idea behind this is based on the channel similarity of two tags in a physical proximity

Actually, besides the phase changes over distance, the measured phase θ_{ij} of tag T_i at antenna A_j also contains the initial phase of the tag and the antenna, *i.e.*, θ_{T_i}

and θ_{A_j}, respectively. We can represent θ_{ij} as follows:

$$\theta_{ij} = (\theta_{d_{ij}} + \theta_{m_{ij}} + \theta_{T_i} + \theta_{A_j}) \mod 2\pi, \tag{2.6}$$

where $\theta_{d_{ij}}$ is the phase changes over distance and $\theta_{m_{ij}}$ represents the phase changes introduced by the multi-path effects. As we know, the tag's phase changes over distance can be calculated as :

$$\theta_{d_{ij}} = 2\pi(\frac{2d_{ij}}{\lambda}) \mod (2\pi), \tag{2.7}$$

where d_{ij} is the line-of-sight (LOS) distance between tag T_i and antenna A_j. Observing Eq. 3.2, we find that the received phase is not only determined by the distance and the outside environment, but also impacted by device diversity. Even if two different tags are very close with each other, *i.e.*, have similar θ_d and θ_m, their measured phases are high likely to be different with each other. To deal with the errors introduced by device diversities, we employ multiple antennas in our system (Fig.2.6). The main idea is that even though different tags have ambiguous and diverse initial phases, such difference can be cancelled using the measurement from two antennas. We calculate difference of the phases of tag T_i that collected at antennas A_1 and A_2:

$$\Delta\theta_i = \theta_{i1} - \theta_{i2} = (\theta_{d_{i1}} - \theta_{d_{i2}}) + (\theta_{m_{i1}} - \theta_{m_{i2}}) + (\theta_{A_1} - \theta_{A_2}). \tag{2.8}$$

In this way, the tag diversity can be cancelled. Similarly, for the user tag T_u, the phase differences of two antennas A_1 and A_2 should be:

$$\Delta\theta_u = \theta_{u1} - \theta_{u2} = (\theta_{d_{u1}} - \theta_{d_{u2}}) + (\theta_{m_{u1}} - \theta_{m_{u2}}) + (\theta_{A_1} - \theta_{A_2}). \tag{2.9}$$

For a user tag T_u and an item tag T_i that are very close to each other, their LOS

path are very similar (the path difference is much less than the LOS propagation path), *i.e.*, $\theta_{d_{uj}} \approx \theta_{d_{ij}}$. Existing work (9; 57) demonstrates that the channel conditions are extremely similar for two tags in physical proximity. Hence we may use the similarity among the phase changes to infer the two coupling tags, *i.e.*, $\theta_{m_{uj}} \approx \theta_{m_{ij}}$. The antenna difference $(\theta_{A_1} - \theta_{A_2})$ is a constant for all the tags. As a result, the phase difference $\Delta\theta_i$ of the target item tag should be close to that of the user tag, *i.e.*, $\Delta\theta_u$. On the other hand, other candidate tags, though have a decrease on their RSSs, are likely to have different LOS distances and multipath effects with the user tag, and hence their phase difference $\Delta\theta_i$ would have a much larger gap with that of the user tag, namely $\Delta\theta_u$. By comparing the phase differences, we can further determine the target tag among all candidate tags. In Section 4.3, we will specify more details of the algorithm to determine the coupling tags.

2.4 System Design

The TagMii program has four modules, namely tag interaction detection, signal preprocessing, candidate tag selection and target tag determination, which are illustrated in Fig. 2.7. Tag interaction detection determines the begin and end time for an interaction action performed by user. The signal preprocessing module performs RSS profile smoothing and phase de-periodicity over the received signal. Then TagMii selects candidate tags by detecting the RSS decreases and finally determines the interacting tags among all the candidate tags by comparing the phase profiles.

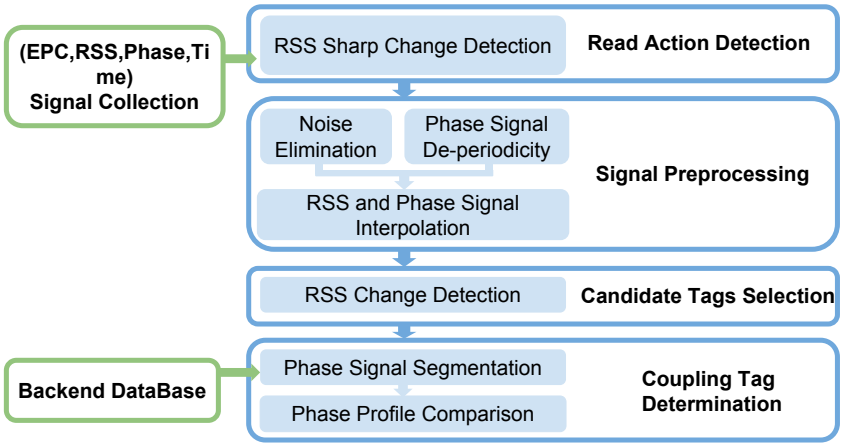
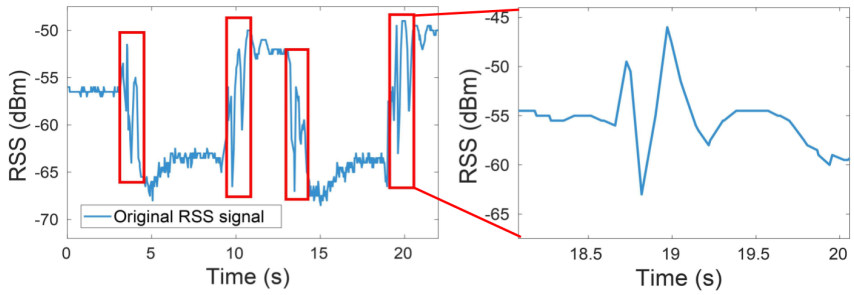


Figure 2.7: System Overview

2.4.1 Detect the action of interaction

When a user holds up the user tag to place it together with a target item tag, the hand movement will result in an abrupt change in the RSS of user tag (shown in Fig.2.4.2). Hence in TagMii, we request that the user waves the user tag slightly in purpose to make this signal change more obviously if the user wants to improve the detection accuracy. TagMii can leverages these time points with RSS discontinuity to locate the starting and finishing time for tag interaction. TagMii uses a threshold R_h to detect interaction events:if the RSS difference between current and following signal points is larger than R_h, then the timestamp of this signal point is considered as potential starting/finishing time for tag interaction. However, sometimes user may suddenly move or rotate the tag , which will also lead to the sudden change of RSS. To resolve this, we can ask the user to wave a tag several times before she starts reading. By doing so, a series of abrupt changes can be caught and we could use such change



(a) RSS profile (b) An abrupt change of RSS

Figure 2.8: RSS profile of user tag for two interactions

pattern to identify the starting time for tag interaction.

After locating these time points, TagMii can leverage the aforementioned inductive coupling phenomenon to determine whether the current moment is the starting time for tag interaction. If the user does begin to interact an item tag, then the RSS will experience a significant decrease.

2.4.2 Signal Preprocessing

Before feeding the received data into the TagMii algorithms, preprocessing is necessary for a better performance. As shown in Fig. 2.7,there are three steps in the signal preprocessing module, namely RSS profile smoothing, phase de-periodicity, and RSS and phase profile interpolation.

RSS profile smoothing: We require TagMii to use commercial RFID devices, which have low resolution on the RSS measurement compared to advanced devices such as the software defined radio. The minimum RSS resolution is 0.5dB by using the ImpinJ R420 reader (33), which is far from sufficient for accurately capturing necessary

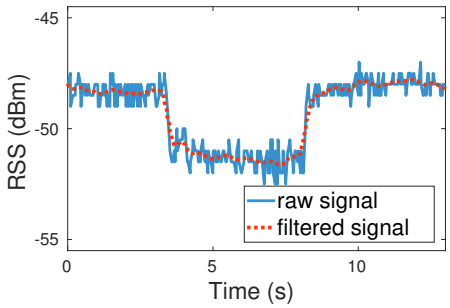


Figure 2.9: RSS signal noise elmination

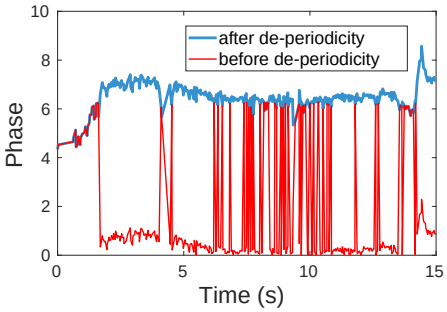


Figure 2.10: Phase signal de-periodicity

tag signal changes. As shown in Fig. 2.9, the raw RSS profile (blue line) is fluctuate and noisy, which is error-prone for the following signal processing. To address this problem, we first smooth the raw RSS profile by employing a low-pass filter. Keep in mind that the RSS variations introduced by inductive coupling has a much lower frequency compared to the Gaussian white noise produced by electronic devices, we choose an appropriate low-pass filter and the performance are shown as the red dot line in Fig. 2.9. Obviously, it removes a lot of fluctuations and make the signal trend more clear. However, the low-pass filter may also obscure the exact change time point of the raw RSS profile. So in the candidate tag selection, we propose to employ a small window to locate the RSS change point.

Phase de-periodicity: From Eq. 3.2 we can find that the phase data is correlated with tag's relative distance from the reader. When a user carries a user tag and move around, the raw phase data is a periodic value ranging from 0 to 2π, which will change gradually. However we may observe that the raw phase data show sharp decreases or increases. We named this phenomenon as 'faked changes'.

Since the received phases can be easily impacted by outside environments and

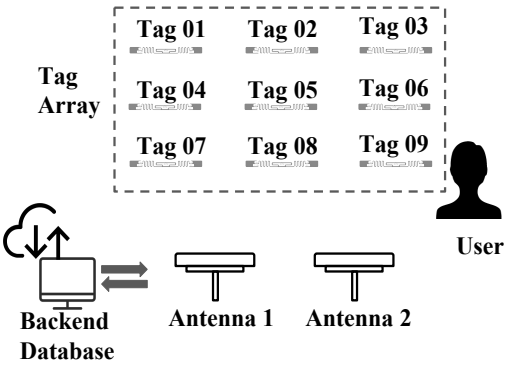
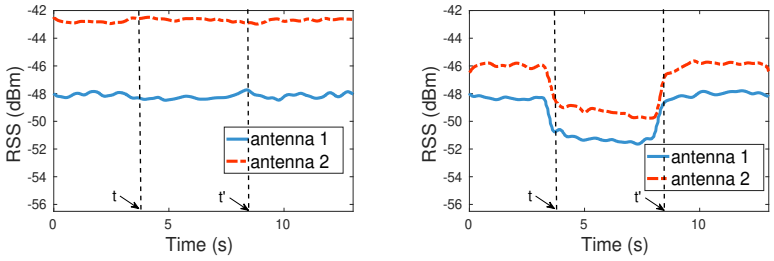


Figure 2.11: Setup of an example

equipments, a sharp fake change between adjacent received phases can be caused by the 2π phase wrapping, or by insufficient reading. Another reason is that when the commercial RFID reader processes the received signal, it introduces some radian of ambiguity such that the reported phase can be either true phase θ or the true phase plus radians $\theta + \pi$ (58) . Hence, We should tell the difference between the faked changes from the true phase changes, and only remove the former one. As illustrated in Fig. 2.10, the received phases are wrapped over cycles and fall into the range of 0 to 2π. We first unwrap the received phase values and retrieve the consecutive phase profile. In this work, we adopt two thresholds, $th_1 = 0.5\pi$ and $th_2 = 1.5\pi$ to detect the π and 2π hops,*i.e.*:

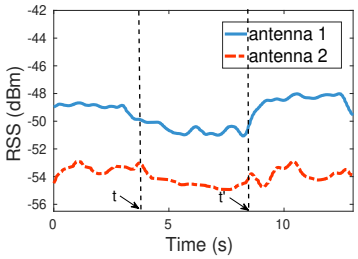
$$\theta(t+1,t_e) = \begin{cases} \theta(t+1,t_e) - 2\pi, & \theta(t+1) - \theta(t) \geq (2-\epsilon_\theta)\cdot\pi \\ \theta(t+1,t_e) - \pi, & (2-\epsilon_\theta)\cdot\pi > \theta(t+1) - \theta(t) \geq (1-\epsilon_\theta)\cdot\pi \\ \theta(t+1,t_e) + \pi, & (\epsilon_\theta - 2)\cdot\pi < \theta(t+1) - \theta(t) < (\epsilon_\theta - 1)\cdot\pi \\ \theta(t+1,t_e) + 2\pi, & \theta(t+1) - \theta(t) \leq (\epsilon_\theta - 2)\cdot\pi \end{cases} \tag{2.10}$$

where $\theta(t+1,t_e)$ represents the phase values from time point $t+1$ to the end

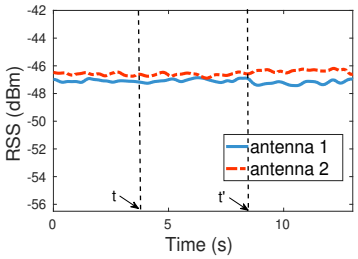


(a) Item tag T_1

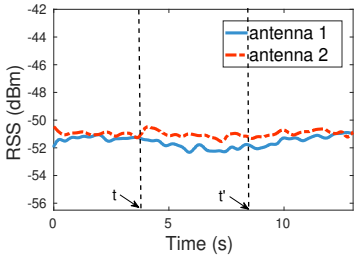
(b) Item tag T_2



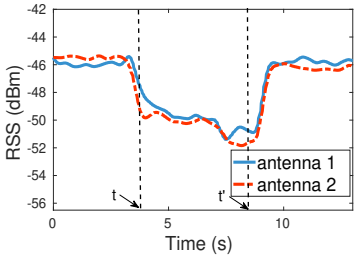
(c) Item tag T_3



(d) Item tag T_4



(e) Item tag T_5



(f) User tag T_u

Figure 2.12: RSS profiles of item tags and user tag acquired by 2 antennas

time point t_e. ϵ_θ is an experimentally-chosen threshold (we set $\epsilon_\theta = 0.5$). The latter phase values will be added or subtracted by π if $th_1 <\mid \theta(t+1) - \theta(t) \mid< th_2$ and by 2π if $th_1 <\mid \theta(t+1) - \theta(t) \mid< th_2$. Note that for this de-periodicity method, we regard the first phase reading as the normal phase reading and the following phase values are unwrapped based on it. The performance of phase de-periodicity can be found in

Fig. 2.10, which shows that TagMii can successfully remove the fake changes from the received phase data.

RSS and Phase signal interpolation: According to the ALOHA-based RFID protocol (59), the sample rate for each tag could be highly different. For simply comparing their RSS and phase profiles, we adopt a cubic spline interpolation method to re-sample the RSS and phase sequence with a uniform sampling rate. After this step, TagMii obtains the same number of samples for the tags that are under comparison in a given time period.

2.4.3 Candidate Tags Selection

As mentioned in Section 2.3.3, we observe that when the user puts her user tag close to an item tag, both tags will show a significant decreases on their RSS profiles. The coupling user tag can be detected using the method presented in Section 2.4.1. Here to present the method to select the candidate tags that **might** be coupling with the user tag. If there are multiple user tags that are detected to have coupling, TagMii computes a set of candidates for every user tag using the same method.

The main idea of the tag selection algorithm is to locate every change points in the RSS profile of each item tag and user tag. Since TagMii has obtained the moment when the user tag starts to be in coupling in Section 2.4.1, if an item tag has experienced a significant RSS change at a similar time as the user tag, we will add this item tag into the candidate tag set S_u of the user tag T_u.

Due to the poor resolution of RSS measurement collected be a COTS reader,

locating the exact change point of the smoothed RSS profile is error-prone. In addition, the changing time points for the user tag and the item tag may not align perfectly. To make our algorithm more tolerant to the practical cases, we adopt a sliding window whose size is T and step is Δt. If there is a significant change occurs in the sliding window, we will record the start time of the current window as the changing time point.

However, as aforementioned, many other factors, such as blocking LOS by moving objects, can also introduce an RSS changes. Only one tag in the candidate set is the correct coupling tag. To reduce the number of candidate tags we choose in mistake, we consider the RSS changes of both antennas. That is inspired by a simple but effective idea, the RSS changes effected by the inductive coupling are simultaneously captured by the two antennas. While for the tags that blocked by moving objects, the RSS decreases may not be observed by both antennas – in most cases, the moving objects will not block the LOS from all antennas. Hence we only consider the tags that have significant and simultaneous RSS changes in both antennas and select them as the candidate tags.

Fig. 2.11 shows an example demonstrating the feasibility of multi-antenna solution. In this example, a 3×3 tag array (namely tag 01-09) is attached on the wall in an office. The distance between each tag and its neighbor is 40cm. A volunteer is asked to browse the tag array and use the user tag T_u to scan the item T_2 he's interested in at about 4-th second. The scanning behavior will last for about 5s till the user leaves away. During the whole process, RSS profile of each tag-antenna link is collected to explore their temporal dynamics. Fig. 2.12 plots the some of the RSS measurements in

this study. In accordance with our analysis, during the first 10 seconds, the RSS profiles of some tags (T_1, T_5) maintain in a relatively stable level, while the RSS trends of T_2, T_3 changes significantly, indicating that the user appoarches the items and impacts these tags's LOS paths sequentially.

2.4.4 Coupling Tag Determination

From Section 2.3.2 we know that for tag i in the TagMii system, its phase profile at antenna j can be modeled as Eq. 3.2. If we compare the phase profiles for tag i received at multiple antennas (antennas A_1 and A_2 in our example), we can eliminate the phase shift introduced by the tag hardware or environment.

Based on Eq. 3.3, if tag i and j are located in a physical proximity, they should have similar values between $\theta_{d_{i1}} - \theta_{d_{i2}}$ and $\theta_{d_{j1}} - \theta_{d_{j2}}$, as well as $\theta_{m_{i1}} - \theta_{m_{i2}}$ and $\theta_{m_{j1}} - \theta_{m_{j2}}$. Hence for samples at any time point $\Delta\theta_i$ should be a small value. We define the *phase difference profile* D_i as a vector where each element is $\Delta\theta_i(T)$ at time T. During the coupling state, D_i and D_j for two tags i and j should be similar. We define an effective distance metric $Dist(D_i, D_u)$ as the Euclidean distance between D_i and D_j to evaluate their similarity. Euclidean distance metric is adopted here since the coupling item-user tag pair should experience similar phase change almost synchronically, and those tags whose phase profiles are similar but locally out of phase with item tag should not be taken into consideration. While Dynamic Time Warping (DTW) can deal with the problem of distortion in the time axis and allows a time series to accommodate sequences that are similar, but locally out of phase.

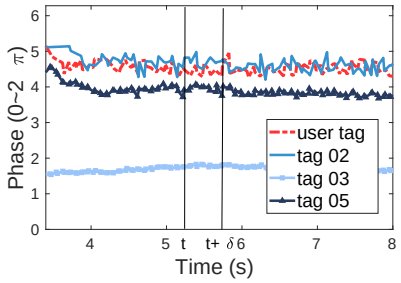


Figure 2.13: Phase difference profile of four tags

Segment / Index	1	2	3	4	5
2	1	1	2	1	1
3	2	2	1	2	2
5	3	4	3	3	3

Figure 2.14: Matrix V

The coupling tag determination algorithm first computes the phase difference profile D_i for each tag i. TagMii measures the similarities between profiles over a time window with fixed size T.The choice of T determines the latency and processing overhead of computation. Once TagMii has determined the set of candidate tags S_c, it immediately begins measuring the similarities between phase difference profiles over these candidate tags and user tag.

TagMii first gets current phase difference profiles $D^1, D^2 ... D^n$ for n candidate tags, where D^i is tag i's phase difference profile in time interval $(t, t+T)$. After interpolation in Section 2.4.2, these n vectors should have equal length, which we denote as L.

TagMii then calculates the Euclidean distance between the phase difference profile vectors of user tag u and each candidate item tag i,

$$L^i = \sqrt{\sum_{k=1}^{L} (e_k^i - e_k^u)^2} \tag{2.11}$$

where e_k^i and e_k^u are the k-th elements of the phase difference profile D^i and D^u, respectively. TagMii performs calculations between user tag u and every candidate tag, and then finds the item tag i whose L^i satisfies $|L^i - L^u| \leq L_h$ and records the top

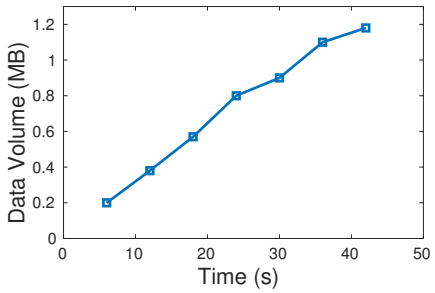


Figure 2.15: Signal data volume vs. time duration

M tags that has the least L^i values from them, which are most likely in the coupling state with user tag. L_h is a empirically pre-defined threshold for comparison. After choosing the top M tags, TagMii uses a $N_c \times T_l$ matrix V to records the comparison result. The one with the smallest L_i will be assigned with 1 and the second one with 2 and so on. Hence, each element (i, t) in V represents the tag i's result in the top M list at t-th segment ($V(i, t) = 0$ if i is not in the list).Take the result in Section. 2.4.4 as an example. As can be seen in Fig.2.14, the element $V(3, 2)$ is 1, which indicates that tag 3's has the highest similarity with the user tag in 2-th profile segment.

After finding the candidate tag that minimizes the distance, we can obtain an array V for all T_l segments. Our goal is to determine the tag with the most occurrences of 1 in V , which is most likely the one coupling with the user tag.

TagMii estimates the similarities after every time interval T. The choice of T determines the latency and processing overhead of computation. In our final implementation of TagMii, we use $T = 2s$ as we have empirically determined it to be suitable for a practical deployment. We have also evaluated the performance of TagMii for other values of T and the results will be presented in the following section.

In the example shown in Fig. 2.11, the user is interaction with Tag 2 among all nine item tags. The RSS profiles of the tags around Tag 2 as well as the user tag are shown in Fig. 2.12. After candidate tag selection, three tags are selected as the *candidate tags*, namely Tags 2, 3, and 5. The phase different profiles of three candidate tags and the user tag are shown in Fig. 2.13. Although all three candidate tags have signal changes caused by the user movement, TagMii is still able to select Tag 2 as the interacting tag by analyzing the phase profile.

2.5 Implementation and Evaluation

2.5.1 Prototype Implementation

We build a TagMii prototype based on COTS UHF RIFD devices: an ImpinJ Speedway modeled R420 RFID reader, two Laird S9028-PCL directional antennas, and three models of tags: ALN-9740, ImpinJ E41C/B, and Alien 964X. We observed inductive coupling for all models of tags. Even if two tags are in different models, inductive coupling still occurs and can be recognized by TagMii. While conducting the experiment, the user tag is attached on a piece of cardboard (as Fig. 2.2 shown) and each user needs to hold this cardboard to interact with the item tag, in order to eliminate the phase influence brought by human hand (60) We also requires the user not to cover the user tag as possible as they can to obtain a relatively comprehensive signal profile of user tag. In fact, as long as one model of tags work, TagMii can be successfully implemented and used. We only show the results of using ImpinJ E41C/B. Each item for evaluation

is attached with an ImpinJ E41C/B tag. We deploy the reader antennas in a distance about 2m away from the tag array. This distance will be varied in order to be consistent with the real implementation. Note one reader may connect to multiple antennas and the antennas are not necessarily at the same location of the reader – depending on the length of the cable. Hence one reader is sufficient to cover a large indoor area. The transmission gain and receiving gain are both 25dB. The prototype is compatible to the standard EPC Class 1 Generation 2 protocols(C1G2). We run the software components of TagMii at a Dell desktop, equipped with Intel Core i7-7700 CPU at 3.6GHz and 16G memory.

In TagMii, the RFID reader keeps interrogating tags in the environment, hence the collected signal measurement data will gradually increase. Fig. 2.15 shows a snapshot for the volume of accumulated data when TagMii monitors a tag array containing 9 tags. The total volume increases almost linearly with time and becomes over 1MB by the end of 40 second. resulting a speed of $>$ 1MB per minute. Thus, in order to reduce such space consumption, TagMii is designed to run in a real-time manner. It will periodically delete the outdated data every few minutes to reduce the storage cost.

2.6 Evaluation Methodology

We evaluate the performance of TagMii in two complex environments with various multipath reflectors, moving objects, wireless signals (WiFi, LTE, and Bluetooth), electrical devices (servers, workstations, printers, refrigerator), and magnetic

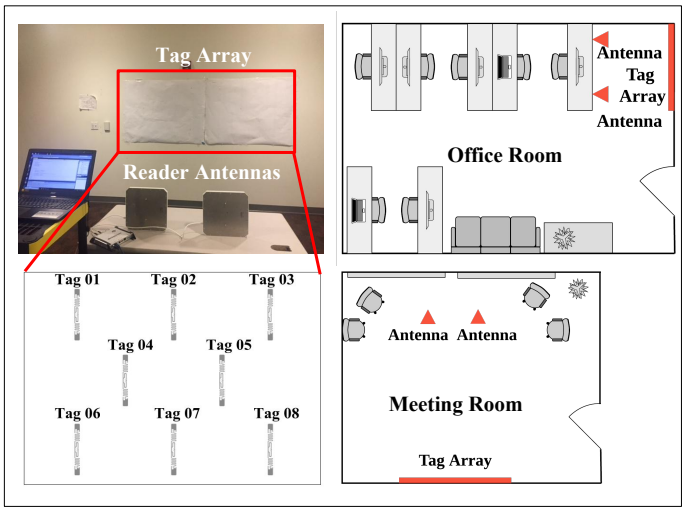


Figure 2.16: System deployment in office/meeting room

fields (whiteboard): (a) an office room and (b) a meeting room as shown in Fig. 2.16. The office environment simulates a retail store where shelves are densely placed, and the meeting room simulates the museum application where artworks are placed on the wall. In office environment we use a tag array that is with 3 rows and 5 columns as illustrated in Fig.2.16 to simulate the tags attached on a rack of commodities or exhibitions, while in the meeting room we further exploit TagMii's performance when tags are attached in a horizontal line. The location of the tag array is highlighted with red markers shown in Fig.2.16

To be consistent with the real implementation, the distance between the reader antennas and the tag array is 1.5m in office room and 3m in meeting room. Note in practice reader antennas can be hung from the ceiling to reduce the probability of blocking LOS signals by moving objects. Note although in theory UHF RFID can operate in 10m distance, practical deployments usually only allow the distance to be

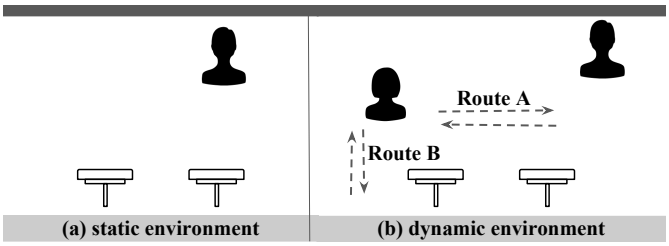


Figure 2.17: Different environments

around 3m or less. However we are not able to reconstruct the ceiling to hang antennas. The distance between two adjacent tags is initially set to be 30cm in the office room and 80cm in meeting room, while the horizontal distance between them are both set to be 30cm. We will show that longer distance between adjacent tags will result in more accurate results. Consider the application in retail stores, it will be rather time- and cost-inefficient if each item is attached with a passive tag. Moreover, a large tag population will degrade the sampling rate for each tag. We may assign one item tag for each type of items. Hence 30cm distance is a practical setting. In our experiments, we invite 4 volunteers with heights varying from 160cm to 180cm. We let volunteer arbitrarily move in the space and use their tags to interact with the item tags in the environment. In the worst case, a moving volunteer may block the LOS path of the signal between tags and the reader antenna as shown in Fig.2.17. Each user is only trained with 3 minutes on what they should do to interact with item tags.

Every accuracy value shown in this section is the **average of 120 production experiments**. Note that all results are analysed offline and there's no direct communication between the reader and user devices (headphone/smartphone *etc.*), meaning that user is unable to learn the information of the item tag she is interacting with during

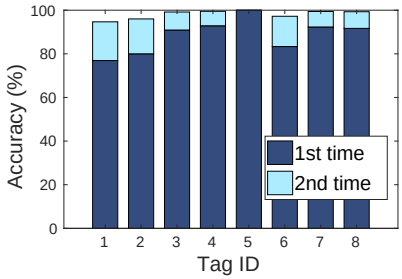


Figure 2.18: Detection accuracy in 1st and 2nd time

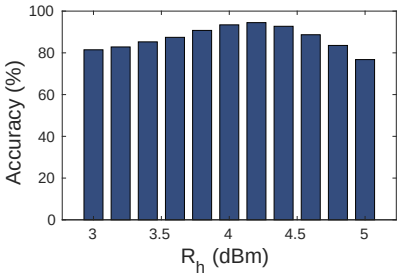


Figure 2.19: 1st-time accuracy vs. R_h

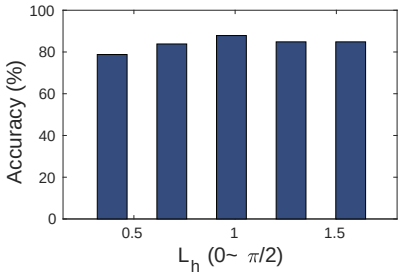


Figure 2.20: Impact of threshold L_h

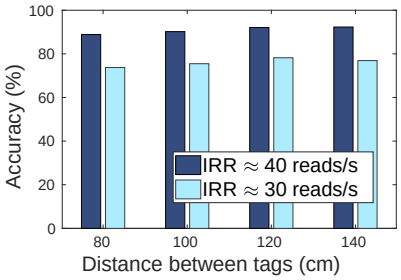


Figure 2.21: Accuracy with smaller IRR

the experiment.

In the following subsections, we will first show the evaluation results of two important steps of TagMii respectively, namely interaction event detection and coupling tag determination.

2.6.1 Interaction Event Detection

Evaluation metrics. The interaction event detection step reports the events of interactions. We use two metrics to evaluate the accuracy of this step: 1) Among all interaction events performed by the users, the *accuracy* is defined as the ratio of events that are successfully detected by TagMii. 2) We also let a user carry a user tag, walk in front of the item tag array, but do not perform any tag interaction. The *false detection*

rate is the number of interaction events reported, which did not actually happen, during a given time period.

Fig. 2.18 shows the accuracy for the 1st-time action and the improvement of a 2nd-time action. If an interaction fails to be detected by TagMii for the 1st time, repeating the interaction for a 2nd time may possible success, similar to the user experience of most input interfaces. The detection accuracy for all 8 tags in the environment can be $> 95\%$ for two-time actions. As discussed in Sec. 2.4.1, TagMii uses a pre-defined threshold R_h to detect interaction events. We vary R_h from 3dBm to 5dBm with a step of 0.2dBm and record the accuracy change of interaction detection. As shown in Fig. 2.19, TagMii achieves a best 1st-time accuracy when the R_h is 4.2dBm. Using other setups, the best-case thresholds are similar. Hence we use this threshold value in the following experiments.

In this set of experiments, we ask four volunteers to carry the user tags and walk in front of the tag array without interacting with any item tag. The false positive rate is measured in number of false positives per minute (FPPM). We vary the distance between the user to the tag array in four different values from 40cm to 100cm. In each experiment we ask each user to walk for 1.5 minutes. We repeat the experiment for four users and two different walk speeds. The results are shown in Table 2.1. The values 0.17 and 0.33 are all in FPPM. Hence the chance of false positives is very small.

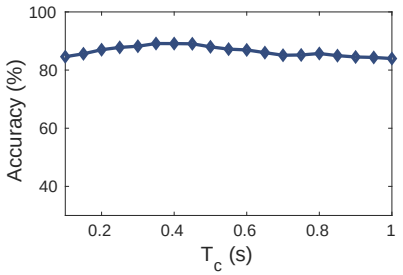


Figure 2.22: Impact of threshold T_c

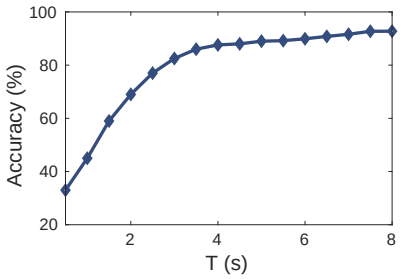


Figure 2.23: Impact of window size T

Walking Speed (m/s)	Distance (cm)			
	40	60	80	100
0.75	0.17	0.17	0	0
1.5	0.33	0	0.17	0

Table 2.1: False positives per minute

2.6.2 Target Tag Determination

Evaluation metric. The *accuracy* μ of the target tag determination step is defined as: $\mu = n_p/n_o$, where n_o is the total number of interaction operations performed by the users, and n_p is the number of tag pairs that are correctly determined as coupling tags among them.

Impact of threshold L_h: The comparison threshold L_h plays a important role in judging whether a tag can be considered as a choice in top list. Intuitively, if L_h is too large, then the system may probably take many irrelative item tags into consideration. On the contrary, a small L_h may cause the miss of target item tag. We vary L_h from 0 to $\frac{\pi}{2}$, and show the accuracy μ of TagMii in Fig. 2.20. TagMii can

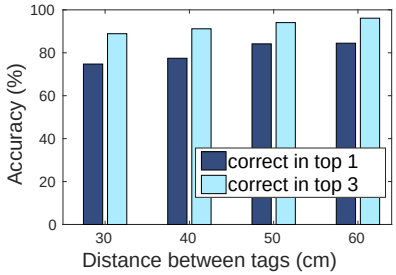


Figure 2.24: Impact of distance between item tags (office room)

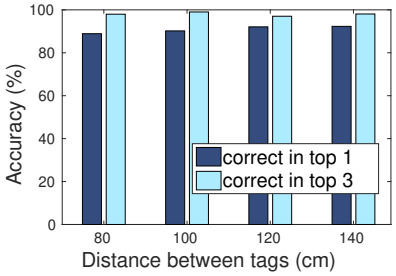


Figure 2.25: Impact of distance between item tags (meeting room)

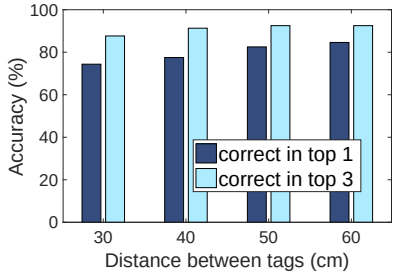


Figure 2.26: Impact of tags' orientations (office room)

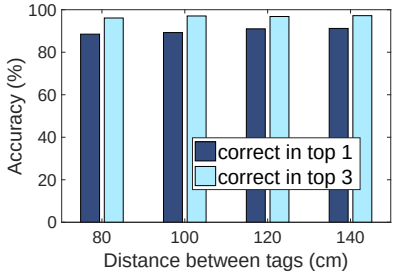


Figure 2.27: Impact of tags' orientations (meeting room)

maintain an accuracy of about 88% when $L_h = \pi/4$ and this value is consistency across different setups, hence we set the comparison threshold to be $\pi/4$.

Impact of threshold T_c: If the time interval threshold T_c too large, TagMii will select too many item tags as candidate tags. Also a small T_c might cause not including the target item tag as a candidate. We vary T_c from 0.1s to 1s, and show the accuracy μ of TagMii in Fig. 2.22. The accuracy of TagMii is relatively stable by varying T_c, but it still achieves a relative high value when $T_c = 0.4$s. We use this threshold value in the other experiments.

Operation duration T: As mentioned in Section 2.4.4, we use a filter with

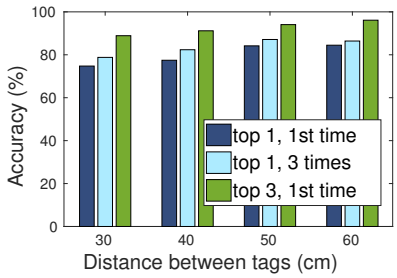


Figure 2.28: Accuracy with retrying (office room)

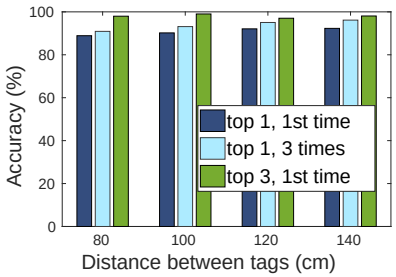


Figure 2.29: Accuracy with retrying (meeting room)

sliding window to process the collected data. Intuitively a longer operation duration T for user to put its user tag close to the item tag will result in higher accuracy. However, the user experience will downgrade if the operation lasts very long. We vary the time window size T from 0.5s to 8s and show the accuracy in Fig. 2.23. It shows that when $T = 4s$, the accuracy of TagMii can stay in a high value for all experimental scenarios. Hence, TagMii requires each user to put the user tag for 4s in the other experiments. Note that a user may receive feedback earlier than 4s.

2.6.3 Overall Performance of TagMii

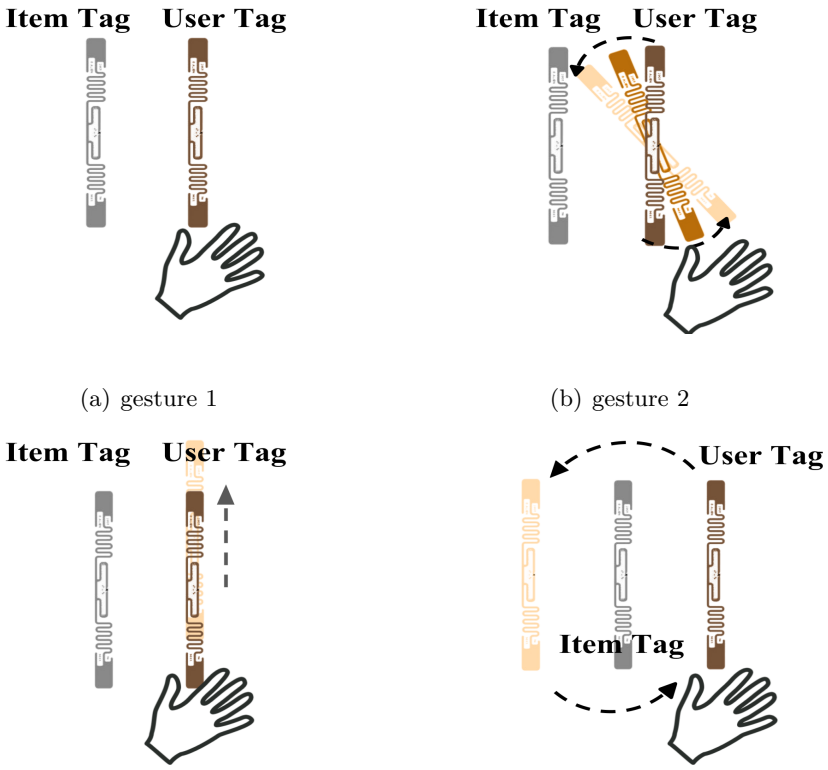
We first evaluate the performance of TagMii in the static environments where there is only one user interacting with the items (Fig. 2.17a). The *accuracy* of TagMii is defined as the ratio of correctly recognizing the two tags involved in each interaction event. In addition, we also show the accuracy of recognizing the tags with up to two repeated interaction actions, if TagMii reports no result to the 1st-time action. Note if TagMii reports a wrong result, then this experiment will be considered failed immediately. We denote this as the *accuracy in 3 times*. Also we are interested in explore

if the target tag is in the top-3 list of candidates of TagMii, if it is not top 1. Some human-computer interaction applications support the following function: if the system is hard to select the most relevant target, it can show the top-3 list and let the user to select the correct one. We denote this as the *accuracy in top-3 list*.

Impact of distance between item tags. We conduct the experiments by varying the distance between the item tags from 30cm to 60cm in the office and from 80cm to 140cm in the meeting room. The results are shown in Fig. 2.24 and Fig. 2.25. We find that for single user case, the accuracy of top 1 results is higher than 75% even if $d_t \approx 30cm$, a very dense placement setup. When we increase the distance, the accuracy of TagMii significantly increase. For distance $>$ 1m, the accuracy of top 1 and top 3 results are around 90% and 98% respectively.

Impact of orientations of tags. Though tags are all deployed vertically at first as Fig. 2.16 shown, we randomly rotate these tags while keeping their centroids unchanged, to check whether different orientations will affect the performance. As can be learned from Fig. 2.26 and Fig. 2.27, there's no obvious difference in accuracy compared to the case when all tags are placed in vertical states.

Accuracy with retrying. From the experiments in Fig. 2.24 we notice that there are two types of failures in TagMii. One is the that TagMii reports no result and the other is that TagMii reports a wrong result. In the no-result cases, we further explore if one or two extra retrying will give a correct result. We ask the volunteer to interact with the target item tag for three times, with a time duration of 30s between two consecutive interactions. We show the accuracy in 3 times in Figs. 2.28 and 2.29.



(a) gesture 1

(b) gesture 2

(c) gesture 3

(d) gesture 4

Figure 2.30: Differen gestures

We find retrying does improves the accuracy though not in a big margin.

Impact of different coupling gestures. We also evaluate the performance of TagMii when the user tag is not placed correctly. We ask the user to rotate the tag (Fig.2.30(b)), to move the tag up and down (Fig. 2.30(c)), or to move the tag around the item tag(Fig.2.30(d)). We show the corresponding accuracy in Fig.2.31. All incorrect gestures will lower the accuracy, but the accuracy is always $> 50\%$. Hence, to maintain a relatively reasonable identification accuracy, we require the users to place the user tag in parallel with the target item tag.

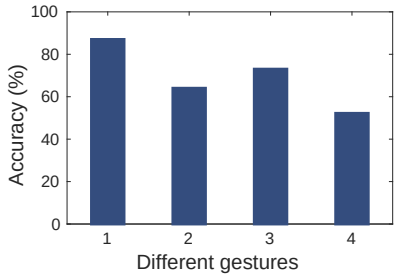


Figure 2.31: Impact of different reading gestures

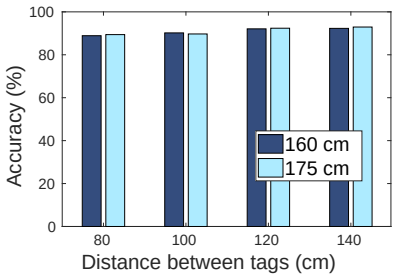


Figure 2.32: Impact of different users' heights

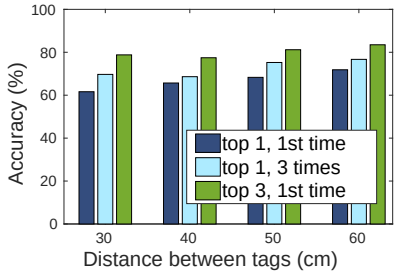


Figure 2.33: Accuracy with route A (office room)

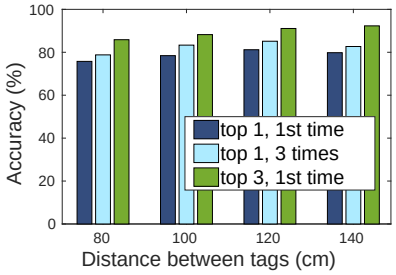


Figure 2.34: Accuracy with route A (meeting room)

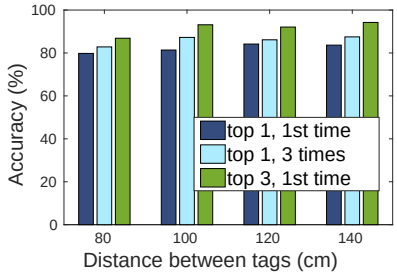


Figure 2.35: Performance of route B in meeting room

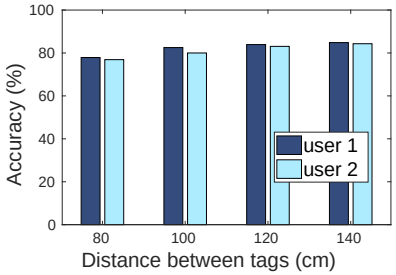


Figure 2.36: Accuracy for concurrent interactions on different tags

Processing time. We show the processing time of TagMii to determine the interacting tags in Table 3.2. We vary the distance between two item tags from 40cm to 120cm. The results show that the processing time of TagMii is very short ($<$ 1sec)

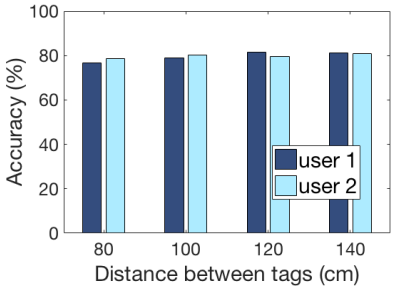


Figure 2.37: Accuracy for concurrent interactions on the same tag

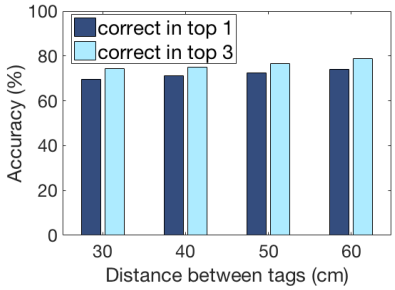


Figure 2.38: Accuracy in unprofessional deployment (single user)

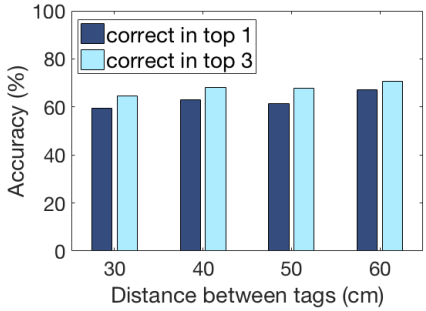


Figure 2.39: Accuracy in unprofessional deployment (average of two users)

for all experiments.

We further evaluate the performance of TagMii in dynamic environments where moving people exist (Fig.2.17b).

Impact of moving people. While conducting the experiments, we ask one extra volunteer to walk around the reader antennas with two patterns shown in Fig. 2.17(b). Route A will block the LOS signals. Route B does not but still makes the

Distance (cm)	40	60	80	100	120
Time (sec)	0.7696	0.7364	0.8189	0.7972	0.7783

Table 2.2: Average processing time

RFID signal more complicated. Comparing Fig. 2.33 with Fig. 2.28 and Fig. 2.29 with Fig. 2.34, we find that Route A could possibly lower the accuracy of TagMii by 10% to 15%. Route B provides better results as shown in Fig. 2.35. These results indicate that the proposed system is robust enough against varying walking events.

Concurrent interactions. We study the case that two users concurrently interact with different target item tags. TagMii should be able to recognize which user tag pairs to which item tag. During the trial, we ask the volunteers to perform coupling with two different item tags almost at the same time. The results in Fig. 2.36 shows the accuracy of the two users respectively. Compared to the single-user case, the accuracy is reduced by 10%. However it still maintains in a high level. We further explore the case when two users concurrently interact with the same target item tag. And the result is illustrated in Fig. 2.37. In general cases, when multiple users interact with the same item tag synchronically, their identification results won't change much. TagMii automatically find the specific item tag that has the highest similarity with the user tag from user-tag array. The readings performed by different user tags are independent and won't have much impacts on each other.

Unprofessional deployment. To illustrate the practicability and robustness of our system, we deploy TagMii in a crowded office room filled with furniture and wireless devices in a more informal way. The distance between reader antennas and the tag array is increased to 4.5m, with the antennas not directly pointing to the tag array. We only request volunteer to wave the tag before the interaction and give no extra interaction guidance before the experiment and let her arbitrarily move in the space

and interact with any interested item tag. Hence it's quite possible that the user place her user tag more than 2cm away from the item tag during the interaction, touch the antenna while holding the tag, or randomly move/rotate the tag before interactions *etc.* The results are shown in Fig. 2.38. Compared with the result Fig. 2.24 in professional setup, the accuracy is reduced by 10∼15%. This result is expected because with the LOS distance between tag array and reader antenna increased, the backscattered signals from tags get weaker, making the receiver hard to extract enough information from them. And the unprofessional coupling gestures will also degrade the system performance (as shown in Fig. 2.31). Later we also extend the single-user trial to the multi-user case, and ask two volunteers to walk around the office room. During the trial, these two users randomly interact with the interested item tags, and the results can be found in Fig. 2.39.

2.7 Discussion

Though TagMii provides accurate mutual identification in the aforementioned evaluations, we acknowledge the following limitations of the current system and propose the potential solutions.

Limitations on system settings: Although TagMii can achieve relatively high accuracy in most evaluations, the limitations on experimental setting can not be overlooked. The performance of TagMii relies on the choices of several thresholds (L_h, T_c, *etc.*). To achieve a better system performance,it is recommended for TagMii to

first learn multiple thresholds through a series of analysis (Sec 2.6.2) before the usage.

Limitations on multi-user cases: Multi-user sensing is a well-known challenging research problem in wireless sensing, since the signals reflected from multiple targets will get mixed at the receiver, interfering with each other. As can be seen from the experiment results, current version of TagMii works well with a single user target. However, if multiple users read item tags almost at the same time, it is not easy for our system to distinguish them clearly. We believe it will be a challenging yet interesting research direction to deal with multi-users case in the future work.

Limitations on number of tags: The current RFID protocol uses slotted ALOHA as the MAC layer solution. Hence tags competes the time slots to reply to the reader. The commodity RFID readers support a constant number of successful read operations per second (around 400 on our device), regardless of the number of tags with the interrogate range. When the number of tags increases, the share of time slots of every tag decreases. We define *individual reading rate* (IRR) as the number of readings obtained from a particular tag per second(61). In order to achieve high accuracy, the IRR of a tag should be sufficiently high, such that its state changes can be continuously and correctly captured. Missing RSS and phase samples will consequently reduce the accuracy of both candidate tag selection and coupling tag determination.

Fig. 2.29 shows an example in which eight items and four users are tagged. In this set of experiments we further attach five more tags in the environment without any other changes on experiment settings. Consequently, the average IRR of each tag reduces from about 40 reads/sec to fewer than 30 reads/sec.Fig. 2.21 shows that lower

IRR will result in obvious accuracy decrease.

We leverages several possible methods to improve the IRR, including: (1) *Decrease EPC length.* Shorter packet durations can increase the individual reading rate. One approach is to decrease the Electronic Product Code (EPC) length. Following the commercial EPC C1G2 standard, the length of EPC code can be set to 8-bit at a minimum (13). We may adopt this length in our implementation. (2) *Adopt FM0 reader inventory mode.* As another approach to decrease packet lengths, the reader may use faster reader inventory mode (i.e., FM0 mode) to speed up the communication rate (28). (3) *Utilize PHY-layer filtering.* The PHY-layer filtering feature is supported by the RFID Class 1 Generation 2 (C1G2) protocol, which allows the reader to read only a subsection of the tags(61).

Future Work Different from TagMii which uses an abrupt change in tag's RSS caused by user wave as the starting time indicator for coupling interaction, RFTrack regards the problem of recognizing coupling state from RSS signal stream as a binary time series classification(TSC) problem, where the objective is to determine whether current RSS signal is at coupling state.

2.8 Conclusion

TagMii is a new approach to enable user-item interactions using passive RFID tags. Compared to other solutions that require a user to carry non-trivial hardware, TagMii only requires every user to carry a passive tag. The reader deployed in the

environment monitors the interaction events and pair the user tag and the corresponding item tag. The key advantage of TagMii is that it is cost-efficient and especially convenient for children, some seniors, people with certain disabilities, and others who do not operate smartphones. We evaluate TagMii in complex environments with rich multipath, mobility, wireless signals, and magnetic fields and find TagMii to be accurate in recognizing user-item interactions in various setups. TagMii is the first step of tag mutual identification and it will attract further research on improving its accuracy and designing new applications based this idea.

Chapter 3

ChopTags

3.1 Introduction

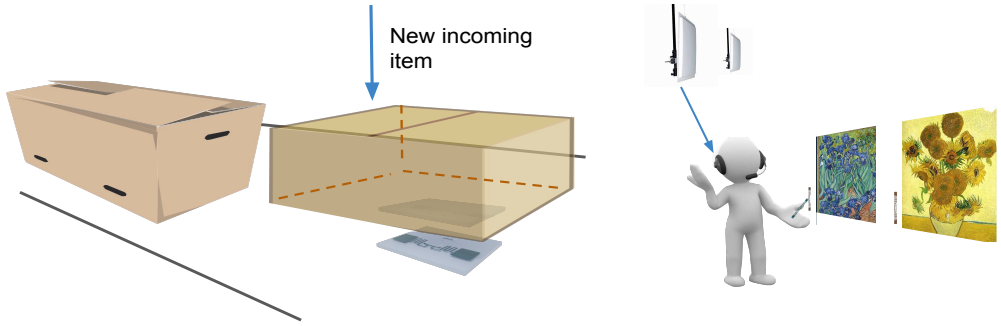
Identifying item-item and user-item interactions is an essential requirement of many ubiquitous computing applications. Here "interaction" is defined as the event that a user touches an item for a particular purpose or two items touch each other to represent some relationship. Interaction identification requires the backend system of this application to recognize both the user and the item or the two items involved in the event.

Passive Radio Frequency Identification (RFID) technology is a cheap and energy-efficient solution for the Internet of Things (IoT) and has been deployed for retailing, warehousing, and transportation applications. However, identifying item-item and user-item interactions based on passive RFID tags is difficult. Despite of the current innovations of RFID sensing and localization based on analysis of physical signal features (41; 42**?**), these methods have inherent limitations to be used for large-scale

and dense deployment. First, they have large errors especially in dynamic environments (9**?**), which are unacceptable for densely deployed tags. Moreover, the signal collection process may cost long time and high data volume. Last, fine-grained signal collection cannot be achieved on commodity off-the-shelf (COTS) RFID readers and need special devices such as software defined radios.

Recently physically altering tag hardware to achieve interaction identification has been proposed (47; 62). However, these solutions only consider IDs as the only sensing source without exploring signal features or optimizations on RFID protocols. Hence it is unclear whether these methods can deal with concurrent interactions where multiple IDs are changing their status at the same time. *For example, our experiments shows that if there are more than 200 tags in the environment with possible concurrent interactions, more than 30% of the interaction events cannot be correctly identified using physical altering tags.*

This paper takes a low-cost and scalable approach called ChopTags, to achieve user/item interaction identification using COTS passive RFID tags. The method is the first to combine physical altering of tags and wireless signal feature analysis. We demonstrate the application of ChopTags by building a prototype called TagChess. We build a set of chess pieces and squares, each attached with a passive tag. The system can automatically record and take notations of a chess game by identifying every move with almost 100% accuracy. It avoids manual chess notations that are known to be necessary but time-consuming and distracting. In addition to the chess application, ChopTags can also be used in the following cases:



(a) Item-item application 1 (b) User-item application 1

Figure 3.1: Applications of ChopTags

Item-item interactions. Suppose each shelf has multiple positions to place items such as shipping boxes and retail products. We attach a ChopTags module on each position as well as a module under each item. By putting all items on the shelves sequentially (Fig. 3.1(a)), the system can automatically record which item is put at which position. It significantly helps manage logistic, retailing, and manufacturing.

User-item interactions. In a museum or exhibition space, a tag is placed on the wall next to each exhibit. A visitor wears wireless headphones. She may put her tagged ticket onto the exhibit tag (Fig. 3.1(b)) Once the reader identifies this interaction, the backend server will stream the commentary of the exhibit to headphone, in the preferred language from the user profile. Such user-service matching requires the server knows both tag IDs. This application can be extended to amusement arcades, interactive classrooms, hospitals, cashier-free stores, and IKEA-style showrooms.

User-user interactions. Attendees of a conference may chat during a coffee break but have little time to exchange business cards. Two attendees may put their

tagged badges together. The backend system will identify the two tags, exchange their e-business cards, and record the time and location. At the end of day, all related records will be emailed to each attendee, from which she can easily recall the conversations.

Our vision of such *interaction identification* is to use cheap passive tags as **ubiquitous cyber-physical connection interfaces**, which form transient links between the physical world and the cyber world. It provides an alternative tool to build future IoT systems with seamless interactions between users and smart devices.

ChopTags module (Fig.3.2) is implemented by a simple hardware modification of COTS passive tags motivated by recent proposals of physical altering of tag hardware (47; 62). Each tag is disassembled into two parts: one includes the chip and the other includes the antennas. They are put into specific positions of an acrylic plate, which serves as a ChopTags module. With separation of antenna and chip, the tag cannot be powered by a reader. When two modules are snapped together, possibly with the help of neodymium magnets, both tags are connected and can be read. Tag interactions can thus be accurately identified.

ChopTags offers the following key features that make it ideal for ubiquitous applications:

- **Low-cost hardware and high scalability**: ChopTags can be made from cheap COTS tags and hence is suitable for large-scale deployment. The cost of adding more users and applications is minimal: only passive tags.

- **High accuracy**: ChopTags achieve 100% accuracy for single interaction events

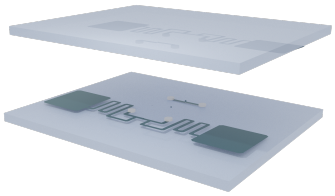


Figure 3.2: A pair of ChopTags modules

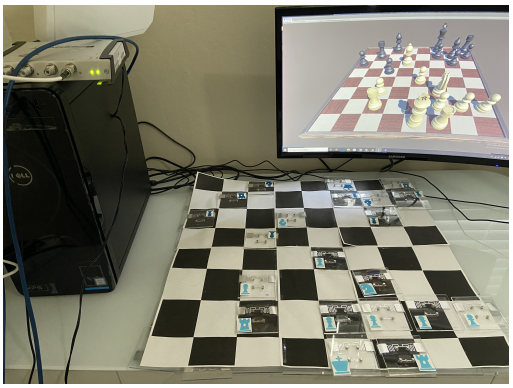


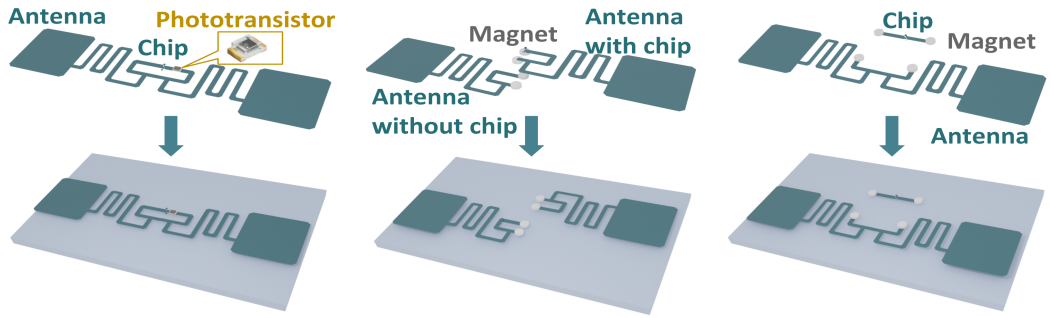
Figure 3.3: The chess notation/visualization system

and $> 95\%$ accuracy for 4 pairs of concurrent interactions.

- **User-friendly**: Children, senior citizens, people with medical conditions, and others who have difficulty of using other smart devices can easily use ChopTags. ChopTags can be made into badges and bracelets.

We implement two ChopTags applications for case studies: one chess notation and visualization system for item-item interactions and one exhibition room for user-item interactions. We address the following main challenges in designing and developing ChopTags and its applications. 1) We present a detailed study of three different tag-modification approaches to validate the reliability of ChopTags modules. 2) We design optimized protocols that are compatible to the current EPC standard for high-accuracy and low-latency identification in complex indoor environments with many tags. 3) We present further optimization to allow ChopTags to identify concurrent interactions.

The remaining paper is organized as follows. We introduce the detailed design of the ChopTags module in Section 3.2 and show the ChopTags-based applications in



(a) Approach 1 (b) Approach 2 (c) Approach 3

Figure 3.4: Three tag modification approaches

Section 3.3. We present the system implementation and evaluation results in Section 5.3 and discuss some related work in Section 3.5. We conclude this work in Section 4.6.

3.2 TagMii Module Design

3.2.1 Design Requirements and Challenges

As a ubiquitous interface for tag interaction identification, ChopTags should be able to satisfy the following design requirements: **R1**) reliability, **R2**) scalability, **R3**) low latency, and **R4**) ease of making. As ChopTags modules could be densely deployed in applications such as board games, which requires fine-grained detection accuracy, tag-tag interactions should be correctly identified with no error or noticeable delays. In the case of chess, the ChopTags system should be able to track the statuses of at least 96 modules, including those for 64 squares and 32 pieces. The communication and computation latency of ChopTags should be small, especially for large-scale deployments. The system should be easily deployed with little effort for installation and users should

be able to easily activate the module for portable uses.

We first argue that **existing RFID sensing and localization techniques are extremely difficult, if not impossible, to be used as a ubiquitous interaction identification method**. First, signal analysis of COTS passive tags is not inherently suitable for dense-deployed applications like chess: previous studies have reported that inductive coupling between closely placed tags (8; 45) has a significant impact on the backscatter signals. When multiple RFID tags are deployed close together on the same surface, the backscatter phase and RSS features are both affected by the physical contact with the RFID antenna, thereby having a substantial negative impact on the accuracy of interaction identification. Second, existing tag localization methods have errors of 10cm to 20cm (9**?** **?**), which are unacceptable for densely deployed tags. Furthermore, since the system has to keep records of all interactions, a large amount of data has to be handled. In many existing methods (29**?**), detection of *each single event* may need to collect a significant amount of physical signal data (e.g., $>$ 100MB).

3.2.2 Modifying passive tags

Different from the physical signal features used by the existing RFID sensing methods, which are not reliable in complex environments, we choose the most robust feature – the on/off and ID information of passive tags.

Our idea is to modify RFID tags to create switches that can automatically trigger on/off state changes during tag interactions. Note that in Fig. 3.4, each tag has a chip with two antennas. We studied three different approaches proposed by existing

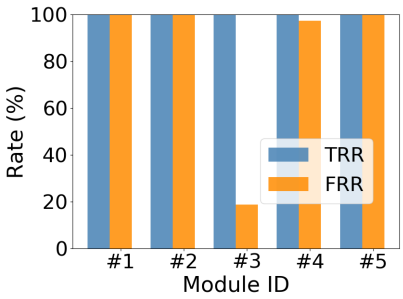


Figure 3.5: Performance of light switch

works, namely 1) cutting away a small part of a tag's antenna and replace the cut section with a phototransistor (Fig. 3.4(a)) (62); 2) cutting a tag from the middle and disassembling it into two parts: an antenna and the chip with the another antenna (Fig. 3.4(b)); 3) disassembling a tag into two parts: the chip and the two antennas (Fig. 3.4(c)) (47). The prior studies (62; 47) only consider a single user-item interaction at a time. Their design cannot be applied directly to a densely-deployed setting with many tag modules and possibly many concurrent interations. Such a new setting presents unique challenges as we will establish in the paper. Building on the prior hardware techniques (47), we focus on the system issues that arise from multiple desensely-deployed tag modules and multiple concurrent interactions. Solutions to these challenges will greatly expand the application scope as we have discussed in the introduction, which cannot be done by the prior art under a single-interaction design between a pair of tags.

3.2.2.1 Approach 1.

A phototransistor is an electronic switching and current amplification component that relies on exposure to light to operate. When light falls on the junction of a phototransistor, reverse current flows are proportional to the luminance. Therefore,

phototransistor can be used as a switch or a variable resistor where its value changes depending on the amount of light they get exposed to as suggested in (62). While the light changes, the resistance of a phototransistor changes accordingly. This change in resistance can be used to progressively detune the tag's antenna.

We build a module in Fig. 3.4(a) to achieve tag interaction identification: when one ChopTags module is stacked face to face towards another one, the light falls on the phototransistor will change, which will further affect its resistance. Hence the antenna's properties will also be influenced. Such dynamics should be observed by the reader in terms of a change in the RSS or the phase of received RFID signal, and even make the tag 'silent' to the reader. By monitoring these signal changes, we should be able to tell which two ChopTags modules are interacting with each other.

We build five modules and conduct 200 interaction events of these modules. The reader operates in rounds. In each round it queries all tags to reply. We show the *true reply rate (TRR)* of each module, defined as the number of replies from this tag module while it does not interact with another module, over the total number of queries (rounds) by the reader. We also define the *false reply rate (FRR)* for each module, defined as the number of replies from this module while it interacts with another module, over the total number of interactions involving this module.. We expect TRR to be 100% and FRR to be 0%. Fig. 3.5 illustrates the evaluation results. Take module #5 as an example, it has TRR and FRR of 100%, demonstrating that it will reply its ID to the reader regardless of being covered by another Module A. It can be noted that most of the modules have high FRR, which is undesirable since we require a module to

reply only if it is not covered by another module.

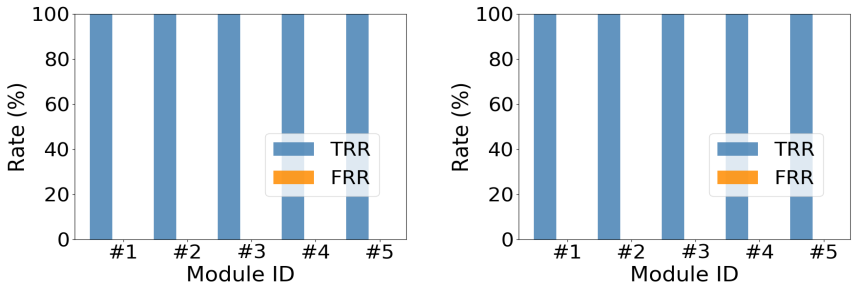
The high FRR could be caused by the following reasons: The change in resistance of a phototransistor may not be sufficient to break the tag's circuit and make the module completely unreadable. Also the required minimum transmission power of the reader to activate each modified tag is different. Module #5 may have minimum response power less than 30 dBm (transmission power in our experiments), which enables them to continuously reply the reader's queries even if they are covered by another module.

Hence we do not use this approach in ChopTags.

3.2.2.2 Approaches 2 and 3.

We explore the possibility of transforming a tag into a "contact switch" by separating the chip from the antennas. We modify a tag in two ways as shown in Approach 2 (Fig. 3.4(b)) suggested in (44) and Approach 3 (Fig. 3.4(c)) suggested in (47).

Approach 2 modifies the tag into a half-half structure and disassembles it into two parts: one with an antenna and the tag chip, and the other with an antenna. Each part has two terminals affixed by neodymium magnet thereby the two parts can connect to each other. The part with the chip will act as an ungrounded monopole antenna, which performs poorly at harvesting RF energy thus making the tag not readable to the reader. When a module is interrogated by the reader, it will keep silent since the monopole antenna cannot build up enough energy. However, if another tag module is



(a) Approach 2 (b) Approach 3

Figure 3.6: Performance of two modification approaches for contact switch

stacking on it, both tags will have complete circuits and can harvest enough power. Thus a pair of tags will be activated simultaneously.

In Approach 3 the chip and the antennas of each passive tag are completely separated. Generally, neither components are readable to the reader as the chip is unable to harvest energy by itself and the antennas will not reply any ID information to the reader without an IC. When two modules are snapped together and the terminals of the chip connect with the terminals of the antennas at the proper alignment, the two tag modules can act as normal RFID tags and reply to the reader.

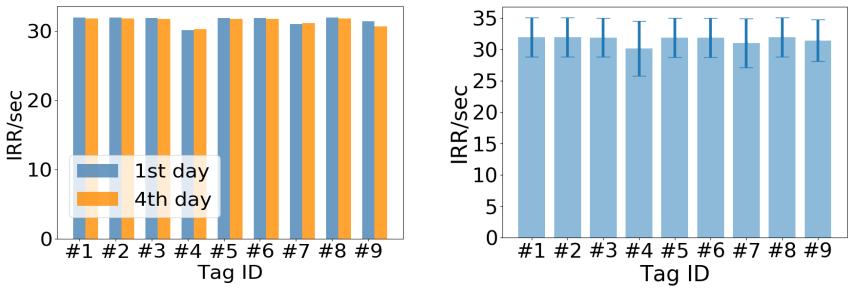
We first build 5 modules using Approach 2 (Fig.3.4(b)). The experimental environment and settings are similar to that for light switch in Section 3.2.2.1. In each trial, a volunteer picks up one of the modules and stacks it on another module. At the same time the reader keeps querying all tags. We still adopt TRR and FRR for performance evaluation but with difference in definition. In this experiment TRR is the probability a tag module replies its ID to reader while it is interacting with another module, and FRR refers to the possibility a item module replies its ID to reader while

it is actually not activated by other module. Fig. 3.6(a) shows the collected results for this design. We fine all modules have TRRs of 100% and FRRs of 0%, meaning that a module will not reply its ID to the reader unless it is activated by another module. The results fully satisfy the requirement of this module.

We build 5 modules using Approach 3 (Fig.3.4(c)). The experimental settings are the same as the previous one. Fig. 3.6(b) shows the collected results from this study. Similar to those of Approach 2, all modules have TRRs of 100% and FRRs of 0%, indicating that each module will reply its ID to the reader only if it is covered by another module. Compared to the results in Fig.3.6, the ChopTags modules built with either Approach 2 or 3 can be used for tag interaction identifications.

3.2.2.3 Stability of Contact Switches.

We conduct an experiment to test the stability of ChopTags modules over a long time period. We run a 4-day experiment to keep tracking the *individual reading rate* (IRR) of each tag module, defined as the number of replies received from this tag per second. The COTS RFID reader used in our experiments support a relatively constant number of successful read operations per second (about 400). 4 modules (ID #2 $\sim$ #5) built by Approach 2 and 4 modules (ID #6 $\sim$ #9) built by Approach 3 are used in this experiment. These 8 ChopTags modules are split into 4 pairs of interacting modules (#2 with #3, #4 with #5, #6 with #7, and #8 with #9). One un-modified passive tag (ID #1) serves as the reference object for comparison. All modules are placed on a desk in an office environment. Other experimental setups are the same as those in Section



(a) IRR at 1st and 4th day

(b) IRR for 9 tags over 4 hour

Figure 3.7: Stability performance of nine modules

3.2.2.1. A one-hour measurement is conducted every 24 hours. During the experiments, people move and work as usually in the office. The whole measurement lasts for 4 days.

Our first finding is that the IRRs of all tags on day 1 have no noticeable difference to those on day 4 (Fig. 3.7(a)). Fig. 3.7(b) presents the average IRR results collected during the four days. The average IRR of the 9 tags is about 32 replies per second with a deviation of 3.38 replies per second. All nine tags have similar IRRs, regardless of whether or not the tag has been modified. Another notable finding is that average IRR of nine tags don't change much in the four days (as shown in Fig. 3.7(a)). These result are important as they validate that the modified tag can stably work without performance degradation.

We eventually choose Approach 3 because we find that a module made by Approach 2 may occasionally reply ID if a user hand touches the magnet. Approach 3 has no such problem.

3.2.3 Remaining challenges.

Simply using the ChopTags modules for user/item interaction applications is not enough. For example, in our evaluation, if there are more than 100 modules in the environment, like the case of chess, more than 30% of the interaction events cannot be correctly identified. Also it is difficult to detect concurrent interactions of multiple module pairs. These challenges are addressed in the next section.

3.3 Application-specific Designs

In this section we show the designs of two application prototypes based on the ChopTags modules by resolving the unique challenges that arise when using ChopTags. One application prototype is a chess notation and visualization system called TagChess for item-item interactions, and the other application prototype is an on-wall tag array that can simulate a museum, retailing shelf, library shelf, or smart home control panel. The challenges that need to be resolved in these two application prototypes are different. The chess prototype has to deal with a large number of ChopTags modules (256 tags in this case). Hence reducing the interaction identification latency is essential. The on-wall tag array prototype must support multiple users. Hence it is necessary to differentiate multiple pairs of modules that are simultaneously interacted.

3.3.1 TagChess for item-item interactions

Benefits of TagChess. In this TagChess prototype, all chess pieces and squares on the chessboard are augmented with ChopTags modules, which we refer as

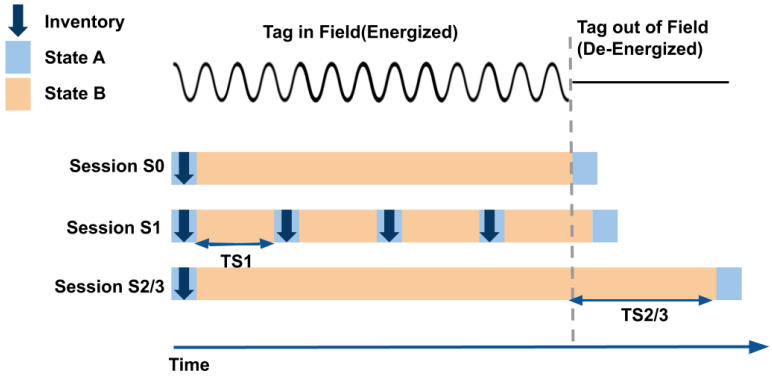


Figure 3.8: Tag behaviors with different sessions

piece modules and *square modules* respectively. Using TagChess, all moves will be recorded in a database and visualized on a computer screen. Hence notations are automatically taken. It is well-known that manual chess notations by the players are necessary but time-consuming and distracting. In addition, it provides an interface that allows one player to use a real chess board to play with an Internet opponent.

In the initial position, 64 modules (32 piece modules and 32 square modules) are activated and 32 other modules are not. The system needs to monitor their status changes throughout the game. We define the interaction event of moving a piece module from one square module to another one as a **PieceMove** event. Each piece also has two statuses: *'on-board'* and *'off-board'*, indicating whether the piece module is being activated or not. An on-board piece is a piece that is steady on a square, and an off-board piece can either be a captured piece or one held by the player in the air during a move.

Intuitively, if a piece module is moved, its status switches from 'on-board' to

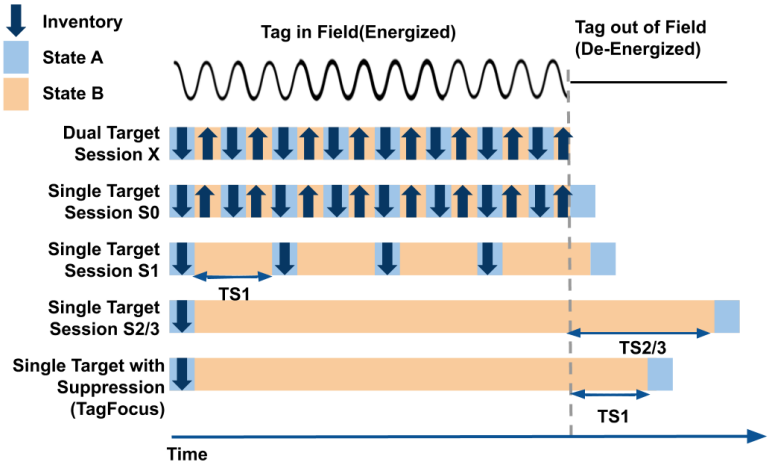


Figure 3.9: Tag behaviors of different reader setups

'off-board', then back to 'on-board'. Current RFID protocols uses slotted ALOHA as the MAC layer solution, which requires tags to compete for time slots to reply to the reader. The commodity RFID reader used in our experiments support at most 400 read operations per second, regardless of the number of tags and the communication distance. When the number of tags is 64, the IRR of each on-board piece module is about $5 \sim 7$ reads/sec. The IRR of an off-board piece module will maintain 0 reads/sec until this piece module is placed back to a square module. However, if we try to leverage the IRR change as an indicator to infer the status change of a piece module, we will receive a detection result with large error, as signal blockage and environmental dynamics sometimes may cause the decrease in a piece's IRR. In order to achieve high accuracy, status changes of a piece module must be continuously and precisely detected.

3.3.1.1 EPC Protocol-based Method Design.

In TagChess we want a piece module to reply its ID if and only if it has been moved, and keep silent until the next move. Hence we explore the inventorying features of passive RFID tags and try to make the tags 'behave' as we expect. COTS RFID readers can be configured using a number of settings to handle different use cases. In our design, we mainly focus on two important configuration settings: *Session* and *Search Mode*, to improve the identification accuracy.

The EPCglobal C1G2 air protocol specification (13) provides a mechanism called "Session" to allow multiple readers to communicate with a single tag. According to the standard, a reader can send out queries to the tags in one of four sessions (denoted as $S0$, $S1$, $S2$, and $S3$). Hence at most four readers can queries the same tag in parallel. For each session, a tag may be in two states represented by an *inventoried flag*, namely state A and state B. If a tag's initial state is A in a session and it is queried and replies to the reader querying this session, its state switches to B and vice versa. A reader can select to query only state A tags or state B tags. By allowing the reader to only query state A tags, we can avoid a tag to reply to the reader for multiple times. However in practical situations a tag with state B may also switch back to A, and the situations differ for the four different sessions. For the rest of this paper, we say a tag "in field" if it is within the interrogation range of the reader and its chip and antennas can be powered, thus it can receive the signals from the reader. And we say a tag "out of field" if it cannot receive the reader signal due to either being out of the interrogation range

or being a single ChopTags module.

The initial value of the inventoried flag depends on the session used by the reader as well as the previous value of the inventoried flag. Each session has an associated *persistence time*, which is the time duration that the inventoried flag is maintained after the reader stops energizing the tag. Detailed tag behaviors are discussed as followed (shown in Fig. 3.8):

- $S0$: When a tag is in field, the inventoried flag of $S0$ is always set to A. It has zero time for persistence, indicating that when a tag is out of field and not energized by the reader, its inventoried flag will immediately return to A.

- $S1$: $S1$ inventoried flag can have initial value of either A or B, depending on its previous state. The inventoried flag is set to previous value (either A or B) if the tag is energized before the persistence time expires, otherwise the inventoried flag will default to A. For instance, when it is in state A and is queried by the reader, the state switches to B. State B will last for a persistence time $T1$ from 0.5 to 5 seconds. After the persistence time expires, the inventoried flag will revert to A, no matter whether the tag is in field or out of field.

- $S2$ and $S3$: The initial value of $S2$ or $S3$ inventoried flag is set in the same way as session $S1$. If a tag is in state A, it can be queried by the reader and can reply to the reader. Then the state switches to B. Its state will keep in B as long as it is in field. Even if it is out of field, the state will still be B until a persistent time $T2/3$ expires. Note $T2/3$ can be very long(with a minimum of 2 seconds).

Our design requirement is that if a ChopTags is snapped with another module and being in field, it should reply once and become silent. When it is separated with another module (out of field) and snapped with another new module (in field again), it should be able to reply again. From that requirement, none of the above sessions is ideal. $S0$ definitely does not meet the requirement because tags cannot be silent. $S1$ will allow the tags to reply again after $T1$. For $S2/S3$, it is nice that a tag only replies once when it is in field. However, when it becomes out of field, we need to wait a long time $T2$ to make it able to reply again. We cannot assume a player should wait that long time until a move is identified. We further investigate other settings of RFID system.

We share the following *experimental observations* from our tests with a ImpinJ Speedway modeled R420 RFID reader, which is one of the most commonly used COTS RFID reader devices in research(8; 63; 48; 47), as shown in Fig. 3.9. Besides "Session", Impinj RFID readers also provide another important mechanism called "Search Mode" that will also influence a tag's overall behavior, hence can be configured to handle various use-cases. Impinj software and libraries allow users to change these settings. There are five search modes available on the Impinj reader used by us: (1) Dual Target; (2) Single Target; (3) Single Target with Suppression (TagFocus); (4) Dual Target Select $B \rightarrow A$; (5) Single Target Reset. Note that "Target" refers to whether the reader will only query tags that are in a single inventoried flag, either A or B (Single Target), or it will interrogate tags regardless of their inventoried flags (Dual Target). Note that due to channel hopping in actual implementation, for Single Target with session $S0$, the tag

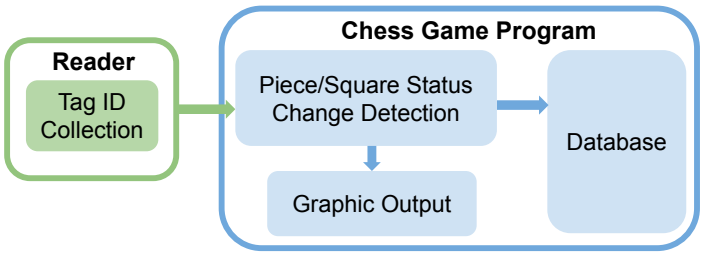


Figure 3.10: Overview of the chess game system

will behave similarly to Dual Target, as shown in Fig. 3.8.

We find that we set TagFocus with session $S1$, a tag only replies when it is in field, and the persistence time is $T1$, which is much shorter than $T2$ in practical situation. Hence it satisfies our requirement of ChopTags.

Therefore, we adopt the TagFocus setting in ChopTags, by which the tag will only reply for one time after being activated, and remain quiet until it has been unpowered. A piece module and its interacting square module will reply only once respectively and then keep silent, until the piece is moved. When this piece module is placed on a new square module, the newly formed piece-square interacting pair will be activated at the same time, thus again reply their IDs for once and remain silent.

3.3.1.2 Algorithmic optimization

The overview design of TagChess is shown in Fig. 3.10. It uses the data collected from the RFID reader to generate high-level interaction events (e.g., PieceMove in a chess game) by querying the backend database, and the final result will be shown as the graphic output in the computer screen. Further algorithmic optimizations are proposed to handle the status change detection process.

Two sets of modules, namely the piece module set (P) and the square module set (S), are pre-registered in ChopTags system. P includes all 34 tag IDs of the pieces – assuming each player has one extra backup queen – and S includes all 64 IDs of the squares. For each module in S, its corresponding row and column numbers are stored in the database.

When the RFID reader begins querying, by adopting the aforementioned TagFocus configuration, all the activated squares and pieces will remain quiet after replying their IDs once. After that, each piece will reply its ID only if a new PieceMove event is triggered, implying that the piece has been moved from the original square and reconnect to another square. The system obtain a list H of **moved** tag ID, which refers to the module who has just experienced a state change. Reader is set to report interaction events once a second. The total number of moved tags is recorded as $|H_n|$. In TagFocus configuration, if piece p_1 is moved from square s_1 to square s_2, then only modules of p_1 and s_2 will report status changes and s_1 will be unpowered. A *capture* move is composed of two sub-steps: 1) the captured piece module is removed from its square and 2) the capturer is moved from its original square to the square of the captured piece. Step 1) will not generate status changes and step 2) will report two status changes. All special cases, including *castling*, *en passant*, and *promotion*, are properly handled and can be successfully identified by TagChess. All other updates are regarded as invalid ones and a warning sign will be shown on the screen.

The $|H| > 2$ cases are mainly caused by moving multiple pieces and stacking them on different squares at the same time, which is invalid in the chess game – even

castling and *en passant* should be moved in an order. Although this case seldom occurs in item-item interactions, it may frequently occur in other application scenarios when users' synchronous interactions are allowed. We will address this problem and propose an effective solution in the following section.

3.3.2 Tag array for user-item interactions

In the applications of the museum auto-commentary system discussed in Section 3.1, smart retailing shelf, library shelf, and smart home control panel, the interaction pattern is that multiple tag modules are placed in the environment to represent different objects (called the *item module*) and each user holds a tag module (called the *user module*) to interact with the an item module. The backend system should identify the interacting modules, record these interactions, and provide further feedback (e.g., sending the museum commentary to the user's headphones or turning on a certain appliance.)

The major challenges of this type of applications is that multiple users may interacting with different item modules at the same time. Hence the system may receive four (or more) simultaneous status changes from four tags and needs to solve the module pairing problem, i.e., deciding which module is interacting with which. Intuitively, two interacting tags should be in close-proximity and hence have similar RF channel features.Thus we leverage the **channel similarity** in RFID communication to resolve the module pairing problem. Channel similarity is an important observation from practical RFID communication: If two tags are put in a physical proximity (e.g., 2cm), they will

show high similarity on their received phase information (9).

3.3.2.1 Channel Similarity

Channel similarity is based on the fact that nearby RFID tags experience a similar multipath environment (e.g., reflectors in the environment) and hence exhibit similar multipath profiles. Existing works (9; 57; 8) demonstrate that the channel conditions are extremely similar for two tags that are in close proximity We propose to use the received tag phase information to determine the similarity. For simplicity, we use tag T_i to refer to the ChopTags module built with tag T_i in the following content.

In an RF environment, the tag T_i's phase changes over the line-of-sight (LOS) distance can be calculated as:

$$\phi_{d_{ij}} = 2\pi(\frac{2d_{ij}}{\lambda}) \mod (2\pi), \tag{3.1}$$

where d_{ij} is the LOS distance between tag T_i and antenna A_j. Besides the dominated phase changes over distance, the received phase ϕ_{ij} of tag T_i at antenna A_j also comprises of the initial phase of the tag and the antenna, i.e., ϕ_{T_i} and ϕ_{A_j}, respectively. We can represent ϕ_{ij} as follows:

$$\phi_{ij} = (\phi_{d_{ij}} + \phi_{m_{ij}} + \phi_{T_i} + \phi_{A_j}) \mod 2\pi, \tag{3.2}$$

where $\phi_{m_{ij}}$ represents the phase changes introduced by the multi-path effects. However, it can be noted from Eq. 3.2 that even if two different tags are placed in close physical proximity, i.e., they have similar ϕ_d and ϕ_m, their measured phases at the same antenna are still highly likely to be different since the received phases are also impacted by the initial phases ϕ_{T_i} and ϕ_{A_j}, which varies by hardware diversity.

To deal with the phase deviations introduced by device-dependent features, we use **differential sensing** in (8), i.e., deploy multiple antennas in our system and measure the difference between tag's phase values received by multiple antennas, instead of measuring a tag's absolute phase value from a single antenna. The intuition underlying this design is that even though different tags have ambiguous and diverse initial phases, such difference can be canceled using the measurement from two antennas.

After adopting 2 antennas in the system, the difference of the phases of tag T_i and T_j that collected at antennas A_1 and A_2 can be calculated as:

$$\begin{aligned}
&\Delta\phi_i = \phi_{i1} - \phi_{i2} = (\phi_{d_{i1}} - \phi_{d_{i2}}) + (\phi_{m_{i1}} - \phi_{m_{i2}}) + \\
&(\phi_{A_1} - \phi_{A_2}). \\
&\Delta\phi_j = \phi_{j1} - \phi_{j2} = (\phi_{d_{j1}} - \phi_{d_{j2}}) + (\phi_{m_{j1}} - \phi_{m_{j2}}) + \\
&(\phi_{A_1} - \phi_{A_2}).
\end{aligned} \tag{3.3}$$

where $\phi_{A_1} - \phi_{A_2}$ is a constant. Therefore when tag T_j and tag T_i are close, they experience quite similar LOS path (as the path difference between them is much less than the LOS propagation path) as well as multipath changes, implying that they should have similar values between $\phi_{d_{i1}} - \phi_{d_{i2}}$ and $\phi_{d_{j1}} - \phi_{d_{j2}}$, as well as $\phi_{m_{i1}} - \phi_{m_{i2}}$ and $\phi_{m_{j1}} - \phi_{m_{j2}}$. Hence $\Delta\phi_i - \Delta\phi_j$ should always be a small value for samples at any time point. As a result, the phase difference $\Delta\phi_i$ of tag T_i should be close to that of tag T_j, $\Delta\phi_j$. While for those tags that are placed far away from tag T_i, they are likely to have different LOS distances as well as multipath environments with tag T_i, and hence their phase difference would have a much larger gap with that of $\Delta\phi_i$. We define the **phase difference profile** D_i for tag T_i as a vector where each element is $\Delta\phi_i(t)$ at

time t. By leveraging the phase differences, we can further determine the interacting tag pair among all activated tags.

3.3.2.2 Interaction Pairing Details

If tag T_i and T_j are the pair of modules interacting with each other, D_i and D_j should have similar values, i.e., the Euclidean distance between these two vectors should be small. We use the distance metric $Dist(D_i, D_u)$ to refer to the Euclidean distance between D_i and D_j and use this metric to evaluate the similarity between phase difference profiles.

Suppose n user modules and m item modules are simultaneously activated (sometimes $n \neq m$ as the module may not be correctly activated). When the reader has finished collecting the data for a time duration T, it immediately begins measuring the similarities between phase difference profiles over these $n + m$ modules. The choice of T determines the latency and processing overhead of computation. In our final implementation of ChopTags, $T = 2.5$s as we have empirically determined it to be suitable for a practical deployment. We have also evaluated the performance of ChopTags with other values of T and the results will be presented in Section 3.4.3.

The basic approach is to match the measured phase difference profiles of all item modules against the profiles of the user modules, aiming to find user-item pairs that have the smallest Euclidean distance. ChopTags first gets phase difference profiles $D_U^1, D_U^2 ... D_U^n$ for n user modules as well as $D_I^1, D_I^2 ... D_I^m$ for m item modules, These $n+m$ vectors should have equal length L after interpolation. ChopTags then iteratively

calculates the Euclidean distance between the phase difference profile vectors of one specific user module u and m item modules,

$$E_u^i = \sqrt{\sum_{k=1}^{L}(d_k^i - d_k^u)^2} \tag{3.4}$$

where d_k^i and d_k^u are the k-th elements of the phase difference profile D_I^i and D_U^u, respectively. ChopTags performs calculations between user module u and n item modules, and then finds the item module i who has the least E^i values that satisfy $|E_u^i| \leq E_h$, which are most likely to be paired with user module u. $E_h = \lambda\sqrt{L}$ where λ is a empirically pre-defined threshold for comparison. Notice that under some circumstances, for a specific user tag u, none of the item tags satisfy the $|E_u^i| \leq E_h$ requirement due to mis-manipulation. Hence no matching result is provided for user module u. In practice the user can easily solve this by retrying interacting with the item module.

3.4 implementation and Evaluation

3.4.1 ChopTags Module Implementation

The system prototypes are implemented based on COTS UHF RFID devices: an ImpinJ Speedway modeled R420 RFID reader, two Laird S9028-PCL directional antennas, and ChopTags modules built with three types of UHF RFID tags: AZ-9662, ImpinJ E41C/B, and Alien 964X. The chip and antenna of each tag was disassembled into two components, as shown in Fig. 3.4(c). All these prototypes are compatible to the standard EPCglobal Class 1 Generation 2 protocols (C1G2). In fact, even if two ChopTags modules are built with different model of tags, interaction event can

still be identified by the reader and the backend system. In the following experiments, we only present the evaluation results of ChopTags modules built with AZ-9662 tags. One COTS RFID reader is sufficient to cover a large indoor area as it can connect to multiple antennas and the antennas do not have to be placed at the same location of the reader. A wire cable can be used to connect an antenna to a reader. The transmission and receiving gains are both set to be 32.5dB. We run the software components of the prototypes at a Dell desktop with Intel Core i7-7700 CPU at 3.6GHz and 16G memory.

This work focuses on the applications in indoor environments such as museums, schools, and homes. No special considerations aregiven to building materials or furniture placement. These environments usually contain various multipath reflectors, moving objects, electrical devices, as well as wireless signals. We evaluate the prototypes of ChopTags in an typical office room (as shown in Fig. 3.11). Such environment can be used to simulate a museum where tagged exhibitions are densely deployed or a chess play room.

For each ChopTags module, the tag components are manually placed on a 70 (W) $\times$ 70 (L) $\times$ 20 (T) mm^3 laser-cut acrylic plate (Fig. 3.13). Two neodymium magnets with 4 mm radius and 2 mm thick are fixed to the terminals of each component as the magnetic connectors. Magnets attached on IC component and antenna part have exact opposite polarity facing upward to ensure snapping connection. The manual installation of each unit can be fabricated in 3-5 minutes by a graduate student. Note we do not intend to choose the best material or size of the module, which for sure have a big room to improve.

Section between Applications	Data Collection	Signal Processing	Graph Processing	Message Exchang
Method 1	1 sec	5.04 ms	11.20 ms	2.82 ms
Method 2	2 sec	7.43 ms	11.20 ms	2.90 ms

Table 3.1: Latency of TagChess

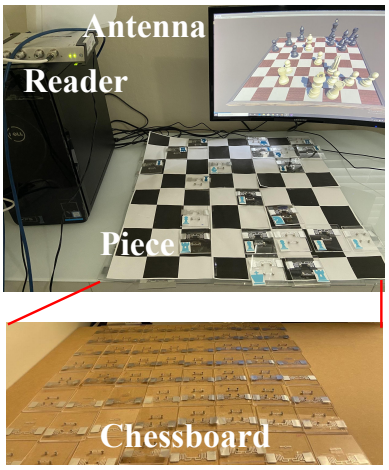


Figure 3.11: Prototype environment

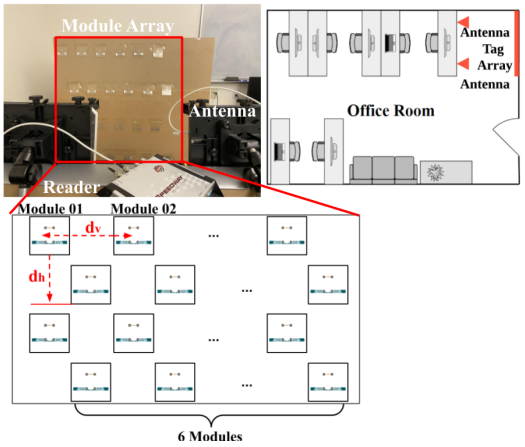


Figure 3.12: Deployment of the tag array in an office

3.4.2 TagChess Evaluation

3.4.2.1 Methodology

Fig. 3.11 shows an appearance of the TagChess prototype. The experiments are conducted on a platform in the aforementioned office room. A 8×8 grid of square modules are deployed on the platform as a chess board, while piece modules are stacking on the chess board that allow users to play with. A reader antenna is hanging above the

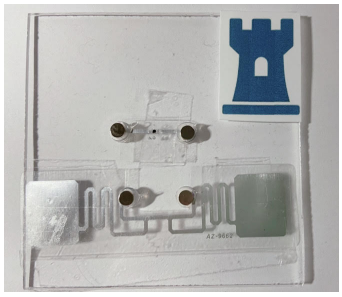


Figure 3.13: A ChopTags module of rook

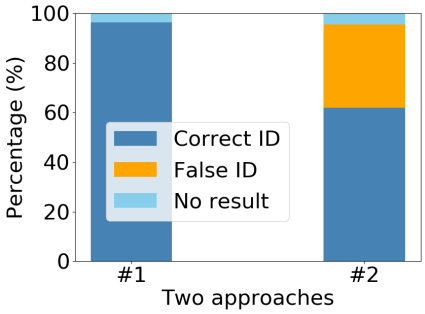


Figure 3.14: Accuracy of two approaches

platform to monitor the interaction events. The distance between the reader antenna and the chess board is 0.8m. The graphic program of TagChess is developed in Unity3D.

To better evaluate the performance, we compare our EPC protocol-based design (which is referred as 'Method 1') with the vanilla design ('Method 2') leveraging IRR changes as the indicator to infer the status change of a module. By using Method 2, the system should regard a module (piece/square) as experiencing a PieceMove event if the IRR increment of this module in two consecutive seconds is larger than a threshold T, while at the same time its IRR in the first second is close to 0 reads/sec. System will update and report the ID every 2s. The threshold T is dynamically refreshed to be half of the average IRR of all piece modules in the last update.

3.4.2.2 Performance.

We first show the latency of TagChess. The response time mainly consists of four parts, namely time for data (replies) collection, time for signal processing, time for generating graphic output in Unity3D, and the message exchanging duration between

reader and server programs. We start by placing 32 piece modules on the chess board, and then move the these pieces for 50 moves, and obtain the average refresh time for these four sections. The results are present in Table 3.1. As can be seen from the result, the latency is mainly contributed by data collection, which is set to 1 second (Method 1) and 2 seconds (Method 2). Method 2 requires significantly long time. Later we show that even with 2x data collection time, Method 2 still cannot identify all moves correctly. W

We define three types of identification for each PieceMove event, *correct reports*, *false report*, and *no result*, to evaluate the accuracy of TagChess. Correct reports are those events identified correctly by TagChess. False report are those events identified incorrectly by TagChess. Sometimes user may not be able to see the correct move she just operated due to some mis-manipulation. She can retry connecting the piece and square modules. However, if no ID is reported after trying for three times, then we consider this PieceMove event receive a 'No result' feedback from the system .

We conduct 250 moves and compute the accuracy of TagChess. As shown in Fig. 3.14, the protocol-based design (Method 1) is very reliable. All events are correctly identified. While for method 2, it will produce 30%+ wrong pairing results. One reason is that the environmental dynamics may also cause the increase of a module's IRR.

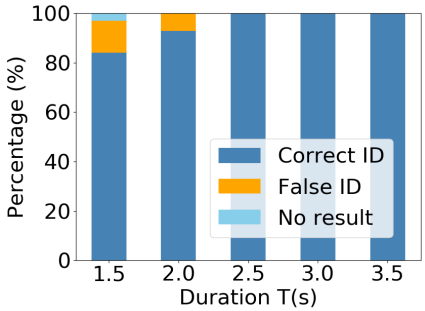


Figure 3.15: Impact of signal collection time T

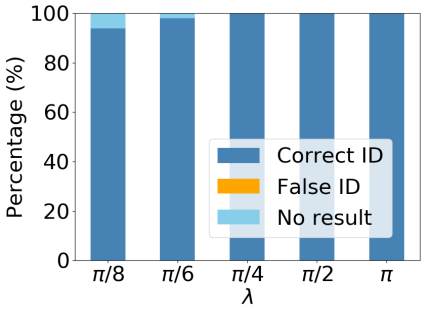


Figure 3.16: Impact of threshold λ

3.4.3 Tag Array Evaluation

3.4.3.1 Methodology

We evaluate the performance of the ChopTags based tag array system in a complex office environment. We use a module array with 4 rows and 7 columns as shown Fig. 3.12 to simulate the modules attached on a wall of a museum or items on a shelf. The distance between the reader antennas and ChopTags's module array is 1.2m in the office room. Before the experiment, the horizontal distance d_h between two adjacent module is initially set to be 25cm, while the vertical distance d_v between them are set to be 30cm. In our experiments, we invite 4 volunteers with heights varying from 158cm to 177cm. We let the volunteers randomly move in the testing space and use their ChopTags modules to interact with the item module in the space. Every accuracy value shown in this section is the average of 100 interaction experiments. We compare the results with the recently proposed TagMii (8). Note the results of TagMii are directly got from the original paper (8).

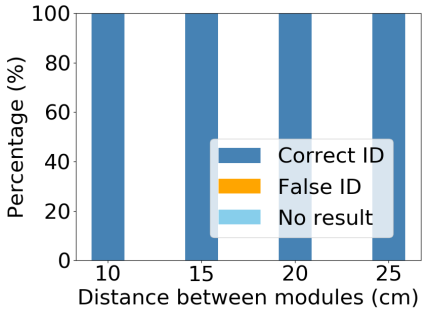


Figure 3.17: Impact of distance bewteen tags

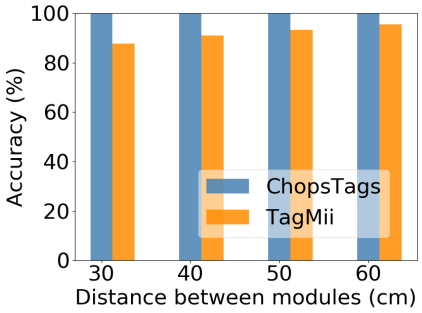


Figure 3.18: Impact of distance bewteen modules

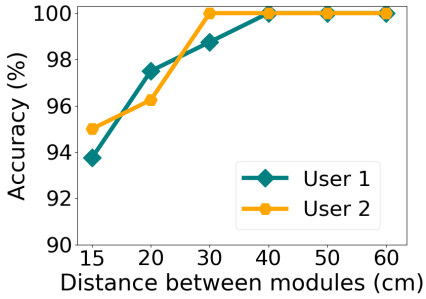


Figure 3.19: Impact of distance between modules (concurrent case)

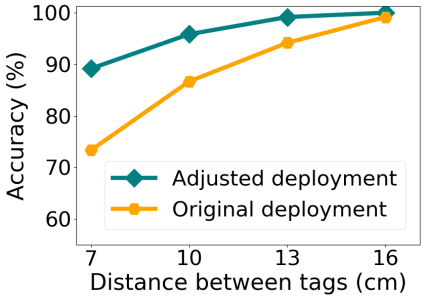


Figure 3.20: Impact of more concurrent interactions

We first evaluate the performance of ChopTags in the situation where there is only one user interacting with the an item at a same time, in the dynamic environment where multiple users move around. Accuracy μ is defined as $\mu = n_c/n_0$, where n_0 is the total number of interaction events operated by the user, while n_c is the number of events where the IDs of involved pair of ChopTags modules are both correctly identified. The false report rate ϵ is defined as $\epsilon = n_e/n_0$, where n_e is the number of interaction events reported, which can be either reporting a wrong user module ID or a wrong item module ID.

Impact of parameter λ: The comparison threshold $E_h = \lambda\sqrt{L}$ in Section

3.3.2.2 plays an important role in judging whether an item module can be considered as an interacting target with the user module. A small λ may cause the miss of target item module. We vary λ from $\frac{\pi}{8}$ to π, and show the accuracy μ of ChopTags in Fig. 3.16. ChopTags can maintain an accuracy of about 100% when $\lambda = \frac{\pi}{4}$ and this value is consistent across different setups, hence we set the λ to be $\frac{\pi}{4}$.

Accuracy for non-concurrent interactions: We conduct the experiments by varying the horizontal distance between adjacent item modules from 10cm to 60cm in the office. To better evaluate the performance of ChopTags as well as compare with the previous work, we divide the experiments into two categories, namely dense deployment where 10cm$\leq d_h \leq$ 30cm, and sparse deployment where 30cm$< d_h \leq$ 60cm. For sparse deployment we compare the evaluation results with those in TagMii which have similar deployment settings of Fig. 3.18. TagMii cannot deal with dense deployment with distance $<$ 30cm. As illustrated in Fig. 3.17 and Fig. 3.18 , for the single user case, even if the deployment of module array is much denser ($d_h = 10cm$) compared to TagMii, the accuracy is always 100%.

Latency: The latency of ChopTags to determine the interacting module pair is shown in Table 3.2. The results show that the processing time of ChopTags is very short ($\leq$ 10ms) during the experiments.

Section	Data Collect.	Processing	Comm.
Time (ms)	2.5s	9.50ms	1.41ms

Table 3.2: Latency for user-item interactions

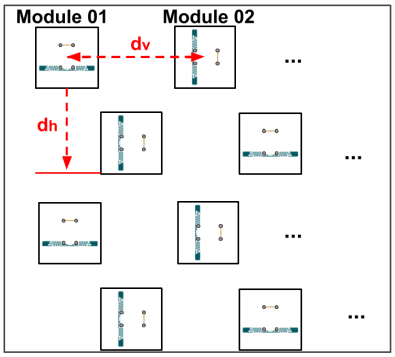


Figure 3.21: Adjusted arrangement for module array

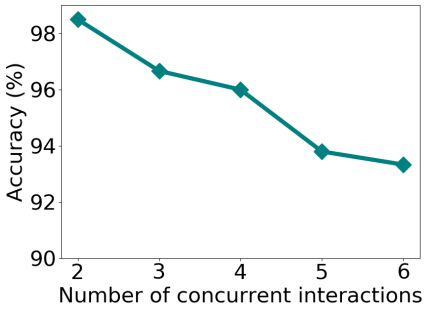


Figure 3.22: Accuracy of concurrent interactions

A main contribution of ChopTags is that it can support and resolve multiple concurrent interactions. We first study the case when two users interact with their target item modules at the same time, and gradually increase the number of users to demonstrate how the number of users will influence the system performance.

Signal collection duration T: As mentioned in Section 3.3.2.2, ChopTags estimates the similarities after every time interval T. Intuitively, a longer data collection duration will result in higher accuracy. However, the user experience will be greatly downgraded if the duration lasts for too long. We vary the time duration T from 1.5s to 3.5s with an interval of 0.5s and evaluate the accuracy in Fig. 3.15. From the results, when $T = 2.5s$, ChopTags can achieve an accuracy of 100% of pairing and identification. Thus, we require each user to put the user tag for at least 2.5s in the experiment before receiving the feedback.

Impact of distance between item modules: In the experiments, we ask 2 volunteers to interact with 2 adjacent modules (deployed in the same horizontal line) at the same time. We vary the d_h between the adjacent item modules from 10cm to

60cm in the office and record the evaluation results. As shown in Fig. 3.19, different from the results in the single user case, the accuracy will decrease as the item modules are deployed in a denser way. The main reason behind this is that the pairing modules are determined based on the similarities between their phase difference profiles. As the d_h gets shorter, adjacent item modules are more likely to experience a similar multi-path environment and hence exhibit similar phase profiles, which leads to ambiguity in resolving the pairing problem.

Impact of different deployment settings: When the module array is densely deployed (with $d_h \leq 25$cm), the performance will degrade due to the coupling effect between tags. Hence, in order to ensure the accuracy of concurrent interactions in dense deployment, we explore different deployment settings. Instead of deployment of Fig. 3.12, we re-arrange the modules in a way shown in Fig. 3.21. We show the corresponding accuracy in Fig. 3.22. Compared with the previous deployment, the new arrangement can achieve high accuracy and can still maintain a accuracy around 90% even if the modules are deployed seamlessly (with horizontal interval $d_h \approx 7$cm).

Impact of number of concurrent interaction events per antenna: We further study the case when more than two users concurrently interact with their target item modules. Note that the d_h is still initially set to be 25cm, and the vertical distance d_v is set to be 30cm. During the experiments, we ask the volunteers to arbitrarily interact with their target modules at the same time without caring about the distance between two adjacent activated item modules. Fig. 3.20 presents the experiment accuracy as we gradually increase the number of concurrent interactions. When less than

4 users per antenna interact with the item modules at the same time, ChopTags system can still maintain a high accuracy above 96% with generating a small fraction of wrong results, as the number of concurrent events per antenna grows to 6, the accuracy slightly decreases to around 93%, suggesting that ChopTags is still able to resolve up to 6-user concurrent interactions in dense deployment. In practical applications, we do not expect that there would be as many as 6 concurrent interactions per antenna.

3.5 Related Works

RFID sensing uses physical signal features from tags to detect their status, which is a promising approach for passive tag based applications such as gesture-based inputs (41; 42). IDSense (43) enables coarse-grained touch events recognition of objects using real-time classification of PHY-layer signal features. PaperID (44) is a similar work that uses supervised machine learning to detect different types of on-tag and free-air interactions with custom-designed RFID tags. Pradhan *et al.* (45) show how changes in the received signal phase caused by touching on RFID tag can be leveraged to detect the finger swipe or touch gesture without any pre-training stage. To our knowledge, the only work that addresses a similar problem to ChopTags is the recently proposed TagMii (8). However, TagMii relies on similar signal features to identify tag interactions. When the number of background tags increases, the identification accuracy might be lower than 80%. All tags are required to be at least 30cm apart, which significantly limits the applications. In addition TagMii requires around 10s to collect enough signals

for each interaction. All these solutions require a considerable time and computation for signal analysis, while their performance is also affected by the environmental dynamics.

Recent studies have proposed to physically modify the tag circuitry and detect the correlated state changes of electromechanical sensors in the tag circuitry caused by human behaviors (47; 48). RFIBricks (47) is a building-block system which symmetric 2D patterns of RFID contact switches are deployed on the top and bottom of each block. Such design enables the backend system to recognize which block is stacked on which, as well as the stacking orientation. RFIMatch (48) detects finger-touching on a tag based on the correlated state change between the tag and an RFIMatch fingerstall worn by user. Different from ChopTags, these methods are not a generalized approach to identify tag interactions and they do not solve the problems such as simultaneous interactions and long query time of large-scale tags.

3.6 Conclusion

ChopTags is a novel solution for tag interaction identification developed with COTS passive RFID tags, which may enable many ubiquitous computing applications that require the recognition of user/item interactions. ChopTags is the first to combine the information of both tag ID presence and physical signal features to infer interactions. ChopTags achieves near 100% accuracy and only requires every user or item to carry a passive tag. We implement two application prototypes based on ChopTags and evaluate the prototypes in complex environments.The results how that ChopTags

is highly accurate and reliable with low latency.

3.7 Acknowledgement

H. Cai, J. Leyva, I. Phan, and C. Qian were partially supported by National Science Foundation Grants 1750704, 1932447, and 2114113. G. Wang is supported by National Natural Science Foundation of China Grant 62002284 and J. Han is supported by National Natural Science Foundation of China Grants U21A20462, 61872285. We thank the anonymous reviewers for their suggestions and comments.

Chapter 4

RV-Track

4.1 Introduction

With the advent of Internet of Things (IoT), the inexpensive, network-enabled cameras have being widely deployed in public places like office buildings, shopping malls, and airports, enabling a wide range of real-world applications such as visual surveillance, suspicious activity detection, and crowd behavior analysis. These cameras record every event in their view, making their videos valuable evidence. Identifying and tracking persons in videos is an essential task. Consider the following problem: sharing of video evidences, even for the common good, creates some threats to the privacy of individuals in the footage. Though people are often willing to release their video evidences to the public, especially when having captured unusual scenes such as traffic accidents, crime scenes, etc, sometimes they may also want to be blurred in the surveillance cameras for privacy purposes. Special personals such as law-enforcement

agents, high-level government officials, administration of a company, and legitimate employees in an office building should be able to request to hide their trajectories in the video within their authority regions. For example, as shown in Fig. 4.1, suppose that in an office building where surveillance cameras are pervasively deployed and one person wants to remove her trajectory in the video. In such situation, accurately identifying and tracking the identifies of persons in camera videos is important.

The state-of-the-art face and identify recognition systems (64) all require a cumbersome appearance-collection process, which consists of photographing and manually annotating each user for training purposes. Hence for scenarios that include a lot of personal dynamics, such as law-enforcement agents visiting a building or travelers in an airport, this appearance-collection and training process is impractical. Since the unavailability of connection between persons and cameras, it is challenging for a vision system to automatically associate the identity of a person with little efforts on the training process (65).

We raise an interesting research question: *is it possible for cameras to recognize and track the identities of persons in the fields of view (FoV) with least training efforts on their appearances?* In other words, we aim to (i) identify a moving person in a camera with her ID and (ii) track the movement of this person as long as she appears in the cameras. We use the term *'target'* to refer to such person and **assume all targets are legitimate users**. We do not consider to identify intruders because 1) intruders will be captured by cameras and 2) identifying intruders needs the database of the appearance of all possible intruders which is out of the scope of this work.

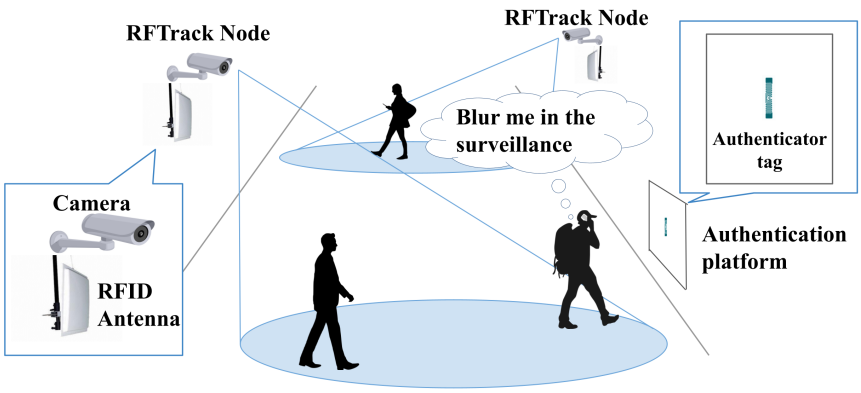


Figure 4.1: Application scenario

Recently, RF-based person identification systems (66; 67) have attracted a lot of research interests due to the appealing sensor-free and contact-free (non-intrusive) nature. Specifically, radio-frequency identification (RFID) sensing technique can tell the identities (IDs) of targets in an area. However connecting these IDs with the persons in cameras still remains unsolved. Therefore in this paper, we propose RV-Track, a novel identifying and tracking scheme for targets carrying RFID tags, which fuses RFID signals and video cameras together using commodity off-the-shelf (COTS) RFID devices and depth cameras. The extra cost is very small because each tag only costs several cents. To achieve these objectives, RV-Track needs to resolve the following challenges.

- **How to automatically associate the identities of persons with the objects in videos?** To address this challenge, we introduce a simple **user authentication (UA)** phase: at the entrance of the camera-covered space, a target needs to perform tags inductive coupling (8) at an authentication platform and get her tag being recognized by RV-Track. Once an UA action has been observed, the RFID subsystem of RV-Track then calculates the moving trajectories of all user

tags based on their phase signal snapshots. Meanwhile for the CV subsystem of RV-Track, given consecutive recorded video frames, RV-Track will estimate the moving trajectory of each object in these frames. We apply a pose estimator AlphaPose (68) to infer poses of different persons in video frames and then leverage right and left wrists as anchors to measure the hand moving trajectory of each object. The object-ID pairing module in RV-Track then calculates the similarity score between RF and visual traces, and tries to match each RF trace to the target's visual trajectory.The UA phase is quite necessary to match each user to their ID in the system. Due to the inaccuracy of recovered RF trajectories, when multiple users carrying tags with different IDs enter the FoV, it is challenging to pair the multiple RF trajectories with the vision tracklets. Hence UA allows the system to be aware of the newly arrived users and improve the accuracy of pairing.

- **How to capture persistent features suitable for long-time ID connections?** Essentially, the similarity scoring-based object-ID pairing is to associate the person id (PID) assigned by CV system with the tag ID (TID), and such pairing could be unstable in real application when PID-switch problem (69) often happens, since the CV system will randomly assign a new PID to the objects in videos, breaking the mapping between old PID and its paired TID. We exploit user appearance information and use a lightweight Re-Identification (*Re-ID*) model to extract the user feature. Such user feature will be associated with its corresponding TID and stored in a dynamic dictionary. If a PID-switch event occurs, we will

repair the new PID with TID by determining whether new and old PIDs belongs to the same user.

- **How to use limited amount of data to train a learning model for both UA and person identification?** To deal with this challenge, we include a data augmentation method to expand the size of the training data set, which dramatically reduces the data collection effort. We further develop a transfer learning method to adapt the trained network for new user identification.

We implement RV-Track using COTS RFID devices and Microsoft Kinect cameras. All implementation and experiments are conducted in various complex environments with rich multi-path, mobility, wireless signals, electrical devices, and magnetic fields, in order to valid RV-Track for practical applications. Even in these environments, RV-Track provides high accuracy. To summarize, this paper makes two key contributions:

- It fills the gap for direct communication between persons and cameras by automatically associating the person identities to objects in videos with few training process on their appearances.
- It proposes a novel model that achieves long-term person Re-ID with the help of RFID signals, which is robust to occlusion and poor lighting.

Recently TagAttention (**?**) also uses RFID-vision fusion to track labeled objects. However, TagAttention cannot be used for applications discussed in this work, because it

cannot differentiate objects with similar appearances. The studied objects in TagAttention are with completely difference shapes and colors. It cannot be used to identify persons especially those with similar clothes.

The remaining paper is organized as follows. Section. 4.2 presents the related work. We present the detailed design of RV-Track in Section. 3 . We show the system implementation and evaluation results in Section. 4 and Section. 4.5. We finally conclude this work in Section. 4.6.

4.2 Related Work

4.2.1 Vision-based Person Tracking

Person detection and tracking is an important field in the computer vision research area (70). Recent approaches rely on deep learning and can be divided into two categories: offline and online. For online framework, main challenges lie in how to learn robust associating metric of linking the detections to tracks, when to create new tracks by distinguishing the true detecting results from false positive ones, and when to terminate the lost tracks (71; 72). DeepSort (73) is a representative work which learns the appearance features from person re-identification task to associate with detections. While for the offline approach, the main challenges become how to construct the graph and network, and how to optimise the global labelling problem of them. SiameseCNN (74) proposes to leverage detection-to-detection pairwise similarity and those between short tracks to handle the problems. However, all these aforementioned tracking schemes

only pay attention to tracking the mobile persons in the frames, without noticing the real physical identities of persons.

4.2.2 Vision-RF Fusion

Recent studies have exploited the possibilities of fusing RF techniques with vision for tracking and identifying mobile targets by matching the information from both channels (75; 76). Mandeljc (75) detects and tracks anonymous human objects from videos by matching the IDs from RF-channel to the detected human instances based on the location information, though this work achieves relatively accurate identification, it requires dedicated and expensive Ultra-Wideband (UWB) devices. ID-Match (67) is a vision-RFID fusion system for human identification from a group through a depth camera and an RFID sensor, however, it fails to provide effective solution for PID-switch problem: the CV system will match the person IDs to tag IDs incorrectly when two persons are very similar in appearance, because it is based on similarity scoring, and the performance will also degrade in dynamic environment where persons walk around in all directions. This problem will be more serious when multiple persons walk in the area. TagVision(77) and RC6D(78) focus more on estimating the object/human pose, instead of discussing the possibility of recognizing and tracking the IDs of persons in the fields of view. Compared with prior work, RV-Track not only fills the gap for direct communication between persons and cameras by automatically associating the person identities to objects in videos with few training process on their appearances, but also proposes a novel model that achieves long-term person Re-ID with the help of RFID

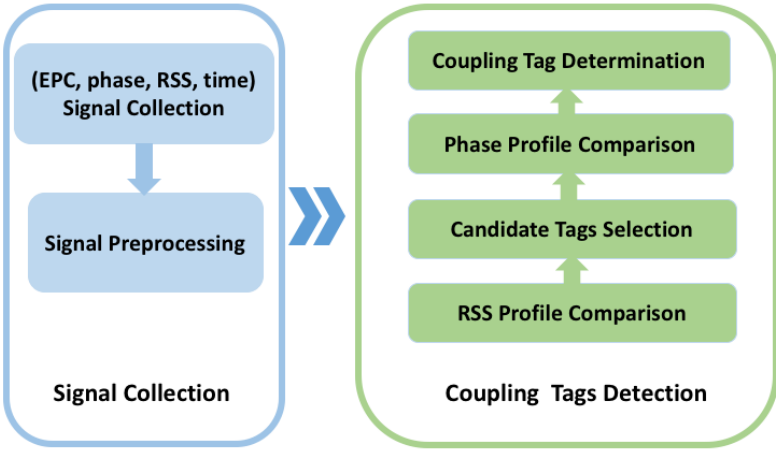


Figure 4.2: System Overview

signals, which is robust to occlusion and poor lighting

4.3 System Design

4.3.1 System Overview

Consider a scene in a office building where RV-Track nodes are pervasively deployed and people come and go in all directions. Basically, each RV-Track is prototyping with an omnidirectional antenna and a depth camera (e.g., Microsoft Kinect) as Fig. 4.1 shown. Each person carries a RFID tag with her and once the person enter the camera's view, she needs to perform UA at the authentication node. The data collected from both camera and RFID antenna will be sent to a server, which then pairs a human object in the video with a tag ID to identify the target. We show the overview of our CV-RFID fusion system in Fig. 4.2, which mainly goes through the following steps to

recognize and track the human targets:

- **Data Collection and Pre-processing:** The depth camera records the image frames of the scene and in the meantime the reader interrogates all tags and collects a set of phase/RSS readings. Afterwards, both image data and RF signal snapshots are sent to the server. The server performs RSS profile smoothing (**?**) and phase deperiodicity (8) over the received RF signal to obtain clean phase and RSS data.

- **User Authentication (UA):** Similar to TagMii (8), we adopt occurrence of the inductive coupling phenomenon as an indicator of user-tag interaction. By monitoring the RSS variations of authenticator tag, we can determine the possibility that authenticator may be coupling with the target's tag in a specific time duration, and start ID pairing if a user authentication action has been observed.

- **User Identification and Tracking (UIT):** The UIT tasks can be divided into two steps: *short-term object-ID pairing* and *long-term pairing*. Short-term pairing module performs two main tasks: similarity scoring and pairing. The similarity scoring sub-module calculates a similarity matrix, in which each element is the similarity between a pair of RF trace and visual trace. According to these similarity scores, the pairing sub-module then leverages some statistical measures to identify possible pairs. A long-term pairing module is also designed to deal with the PID-switch issue when short-term pairing result is broken due to either lack of lighting or object occlusion. Here, we exploit long-term features of human object

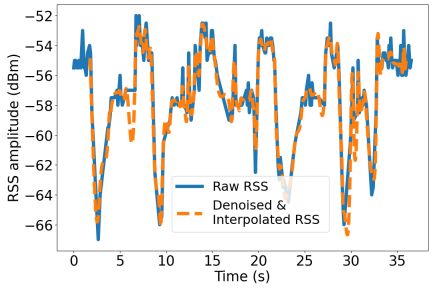


Figure 4.3: RSS signal noise elimination

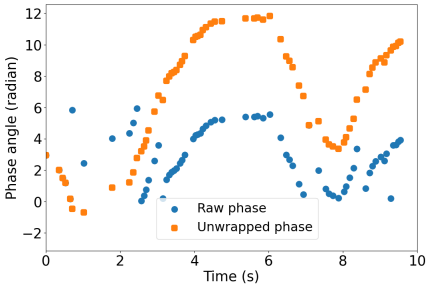


Figure 4.4: Phase signal deperiodicity

to enhance pairing reliability.

- **Data Augmentation** RV-Track also designs a data augmentation method to reduce the amount of human effort in data collection.

4.3.2 Data Collection and RF Signal Pre-processing

After collecting RF data and visual data, RV-Track performs a set of pre-processing steps to obtain clean phase and RSS readings. There are three steps in the signal preprocessing module, namely RSS denosing, phase unwarpping, resampling and windowing.

Signal Denoising: Corrupted by the environmental noise and hardware imperfection, the collected RSS signals are fluctuated and noisy (Fig. 4.3), which is error-prone for the following signal processing. To address this problem, we first smooth the raw RSS profile by employing a low-pass filter (8).

Phase Unwrapping: The raw phase data of RFID signal received by reader is a periodic data stream, gradually changing from 0 to 2π, which is referred as a wrapped

phase. This ambiguity of wrapped phase will increase the complexity of subsequent algorithm. Moreover, while commercial RFID reader processes the received signal,it also introduces an ambiguity of π radian. Therefore, we adopt phase profile smoothing (**?**) and two-threshold based phase de-periodicity method (8) to unwarpped the original phase data. The unwrapped phase profile is shown in Fig. 4.4.

Resampling: The fusion and pairing of the RFID and visual channels requires the synchronization of two-channel data readings. However, current RFID protocol uses slotted ALOHA as the MAC layer solution, *i.e.*, RFID tags randomly respond to the reader, hence the received RSS/phase sequence is not uniform in the time domain (8). In order to synchronize the two channels, we calibrate the reader's and the camera's reading timestamps according to the system's clock before data collection, and choose to use the camera's (Kinect) timestamps as the standard timestamps. The denoised RSS and unwarpped phase sequences are interpolated at a frequency of 30Hz (the same sampling rate as Kinect) based on the camera's standard timestamps.

Windowing: We then split the interpolated RSS stream into a set of subsequences using sliding window whose length is m with 2/3 overlapping. In our final implementation of RV-Track, m is set to 90 as we have empirically determined it to be suitable for a practical deployment. We use the following three methods, namely **mean subtraction**, **standardization** and **normalization** (79) to transform the RSS subsequences into the data form needed by the classification model in Section. 4.3.3 and make sure the values in each window are then normalized to have zero mean and unit variance.

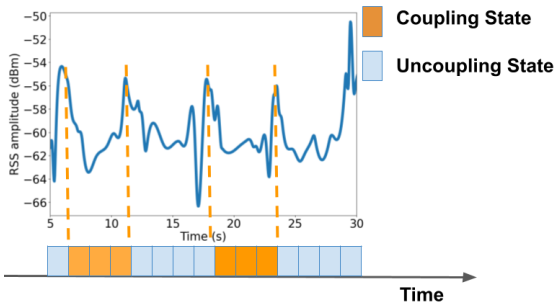


Figure 4.5: Sequences of UAD Results

4.3.3 User Authentication

UA is proposed based on the inductive coupling process, which is explained in details in TagMii (8). During our experiments, we require all participants to perform UA (tags inductive coupling) at a pre-deployed authentication platform at the entrance of the camera-covered space, to get her/his tag being recognized by RVTrack. An extra RFID tag is fixed at the platform serving as the 'item tag' which has the same functionality as the item tag in TagMii. While a user puts the user tag at a close distance ($\leq$ 2 cm) to the item tag, the signal from either tag would be different from the signal when a tag is put alone. A sudden change in the backscattered signal strength from both tags will occur at the beginning of inductive coupling. The system can track such signal strength changes to find out the potential user tags that may be in the coupling state among all user tags in the environment, and further select the pair of tags that are truly in the coupling state. One of the tags can be mobile and the user puts it close to the static one for 2 seconds.

However, different from TagMii (8), which uses an abrupt change in tag's

RSS caused by a deliberate user hand wave as the starting time indicator for coupling interaction, RV-Track regards the problem of recognizing coupling state from RSS signal stream as a binary **time series classification** (TSC) problem, where the overarching objective is to determine whether current RSS signal is at coupling state.

Let $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{|\mathcal{D}|}$ represent the collection of all labeled RSS signal, where $\mathbf{x}^{(i)}$ is the i^{th} RSS sequence and $\mathbf{y}^{(i)}$ indicates its corresponding label. $\mathbf{x}^{(i)} = (x_1^{(i)}, ..., x_m^{(i)})$ is a RSS sequence composed of m amplitude values $x_t^{(i)}$, and $\mathbf{x}^{(i)}$ can be assigned to either the 'no-coupling state' or 'coupling state' class. Intuitively, we label $\mathbf{x}^{(i)}$ with $\mathbf{y}^{(i)} = 0$ if it belongs to the no-coupling class, and with $\mathbf{y}^{(i)} = 1$ to the coupling class.

We propose to learn a mapping from the RSS sequence $\mathbf{x}^{(i)}$ to $\mathbf{y}^{(i)}$ in a supervised way by using Recurrent Neural Networks (RNNs), which explicitly analyze RSS sequence in temporal domain. Specifically, we use Long Short-Term Memory (LSTM) (80; 81) based networks, to effectively learn long-term dependencies in the data. It has been observed that traditional RNN has the problem of handling long-term temporal dependencies due to gradient vanishing (82). As an extended model of RNN, LSTM can remember information for long periods of time (83), which makes it a natural choice for this application.

We choose to use LSTM-based network,*i.e.* LSTM-FCN (83) for this TSC task and build the sequence classification model with it. We adopt a similar network architecture as the one in (83), while remove the dimension shuffle module. The overall network is a two-stream network that combines a fully convolutional (FCN) stream and

a LSTM stream. where the FCN stream has three 1D convolutional layers, each with batch normalization and ReLU. For the LSTM stream, there is one LSTM layer with 128 units and an dropout rate of 0.8. The output of the LSTM is concatenated with the output of global pooling (GAP) layer from the FCN stream.

In general, the prediction probability vector $\tilde{\mathbf{y}}^{(i)}$ of all gestures can be represented as:

$$\tilde{\mathbf{y}}^{(i)} = \mathbb{LSTM} - \mathbb{FCN}(X; \theta) \tag{4.1}$$

where θ is a set of all parameters in LSTM-FCN sequence classifier. where $\mathbb{LSTM} - \mathbb{FCN}$ is the classifier , and $\tilde{\mathbf{y}}^{(i)}$ is the predicted label. We naturally choose to calculate binary cross-entropy between $\tilde{\mathbf{y}}^{(i)}$ and $\mathbf{y}^{(i)}$ over all training samples $\mathcal{D}_{train} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N}$ as the loss function, and optimization criterion for the network is to minimize the negative log-likelihood:

$$\mathcal{L}(\tilde{\mathbf{y}}, \mathbf{y}) = -\frac{1}{N}\sum_{i-1}^{N}[\mathbf{y}^{(i)} \log \tilde{\mathbf{y}}^{(i)} + (1 - \mathbf{y}^{(i)}) \log(1 - \tilde{\mathbf{y}}^{(i)})] \tag{4.2}$$

where $N = |\mathcal{D}_{train}|$, is the training dataset. The training and evaluation details are illustrated in Section. 4.5.2. The UA action is captured when the authenticator tag as well as at least one person's tag are in the 'coupling state' at the same time. We regard these person's tags as 'candidate tags' (8), *It is noteworthy that in real experiments, there will be no more than 2 persons selected as 'candidate' in most of time*, and we will detail the matching process for these candidates in the following sections.

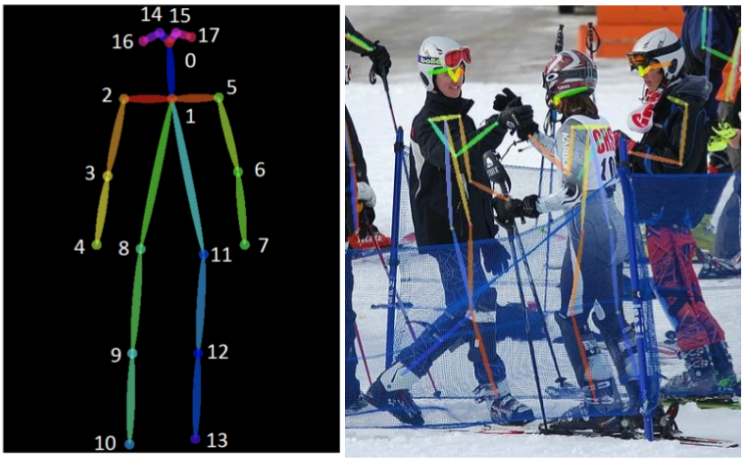


Figure 4.6: Alphapose landmarks

4.3.4 User Identification and Tracking

After a user authentication action is detected, the data collected from both camera and candidate RFID tags are sent to a server for User Identification and Tracking (UIT) purpose. Our UIT system has four modules, namely visual trajectory extraction, RF trajectory extraction, short-term object-ID pairing and long-term feature pairing, as shown in Fig. 4.2. The visual trajectory extraction module retrieves human objects and hand moving trajectories from a sequence of video frames, meanwhile, the RF trajectory extraction module retrieves user tag moving information from received RFID phase data. The short-term object-ID pairing module first compares the moving trajectories from both data sources and assigns them similarity scores, and then couples human objects with tags based on the similarity scores.

4.3.4.1 Visual trajectory extraction

Consider a time duration $[t_0, t_k]$, the consecutive video frames $V_{t_0}, V_{t_1}, ..., V_{t_k}$ collected from the depth camera are first fed into AlphaPose (68) to retrieve the skeletons $S_1^t, S_2^t, .., S_J^t$ of J persons in video frame V_t. AlphaPose is able to connect skeletons across consecutive images belonging to the same human object. When estimating the pose of a single person, it generates n three-element predictions in the format of $(x_i; y_i; c_i)$, where n is the number of keypoints (landmarks) to be estimated, x_i and y_i are the pixel coordinates of the i-th keypoint, and c_i is the confidence of the above coordinates. In this work, landmark and keypoint are used interchangeably and we adopt the COCO keypoints setting (68) ($N = 17$). The outputs of AlphaPose are shown in Fig. 4.6. Therefore, for each extracted skeleton S_j^t in a video frame V_t, AlpphaPose generates 17 three-element predictions in the format of (x_i^t, y_i^t, c_i^t), and the perdictions for a specific landmark over all video frames will also be a time series data. In RV-Track, we retrieve the positions of both right wrist la $W_j^{(R)t}$ and left wrist keypoint $W_j^{(L)t}$ from each skeleton S_j^t in a image, and denote the keypoint series over all frames as $W_j^{(R)t_0,t_1,...,t_k}$ and $W_j^{(L)t_0,t_1,...,t_k}$ respectively. There two series will be further analyzed to recover a human object's hand movements in 3D real-world space.

Currently, the trajectories of each person's hands that we get are represented as pixel values in the camera image plane. To estimate the real-world level trajectory, we map the coordinates series $W_j^{(R)t_0,t_1,...,t_k}$ to the original depth images $V_{t_0}, V_{t_1}, ..., V_{t_k}$ frame by frame and obtain the distance trajectory of corresponding voxel in 3D world

space (**?** 77). We denote the distance trajectory from the first frame V_{t_0} to the last frame V_{t_k} as $d_j^{(R)t_0,t_1,...,t_k}$ for right wrist. Similarly, we can retrieve the 3D motion trajectory for left wrist $d_j^{(L)t_0,t_1,...,t_k}$. And the visual trajectories set in this time duration will be $\mathcal{T}_{visual} = \{d_j^{(R)t_0,t_1,...,t_k}, d_j^{(L)t_0,t_1,...,t_k}\}_{j=1}^{J}$.

4.3.4.2 RF trajectory extraction

Ideally in an RF environment, the distance L from the reader antenna to the tag T_i at time t can be calculated as follows:

$$L(t) = \frac{\phi_i(t) \times c}{4\pi f} \tag{4.3}$$

where $\phi_i(t)$ refers to the corresponding phase change over the signal travel distance and c is the speed of light. However in realistic deployment, we can not explicitly calculate the exact distance $L(t)$, since that except for the phase change $\phi_i(t)$ over distance, the real measured phase $\phi_i^r(t)$ at antenna also contains some additional phase changes introduced by reader and tag's circuit, i.e., $\phi_i^r(t) = (\phi_i(t) + \phi_{T_i} + \phi_A) \mod (2\pi)$, here ϕ_A and ϕ_{T_i} are the additional phases introduced by reader antenna and tag respectively (8**?**). Instead of recovering the accurate location changes, we can only obtain the relative distance changes of tag T_i in a time duration from t_x to t_y as followed:

$$\Delta L = L(t_y) - L(t_x) = \frac{(\Delta n\pi + (\phi_i^r(t_y) - \phi_i^r(t_x))) \times c}{4\pi f} \tag{4.4}$$

where $L(t_x)$ can be used as a reference distance which can be set as the first calculation. Therefore, by calculating the relative distance changes sample by sample in time slot $[t_0, t_k]$,we can recover a relative moving trajectory of the tag T_i, which we

denote as $d_{id_i}^{t_0,...,t_k}$. And the RF trajectories set recovered from all candidate tags will will be $\mathcal{T}_{RF} = \{d_{id_i}^{t_0,...,t_k}\}_{i=1}^{I}$. As aforementioned in Section. 4.3.3, in most time, $I \leq 2$ and the RF trajectories set wouldn't contain many RF traces waiting for pairing.

4.3.4.3 Short-Term Object-ID pairing

After retrieving I RF trajectories $\mathcal{T}_{RF} = \{d_{id_i}^{t_0,...,t_k}\}_{i=1}^{I}$ and $2J$ visual trajectories $\mathcal{T}_{visual} = \{d_j^{(R)t_0,t_1,...,t_k}, d_j^{(L)t_0,t_1,...,t_k}\}_{j=1}^{J}$, RV-Track first compares the moving trajectories from both data sources and calculates a similarity matrix Sim, in which each element $sim_{ji}, j \in \{1, ..., J\}$ and $i \in \{1, ..., I\}$, is the similarity score of $d_j^{(R)t_0,...,t_k}$ (or $d_j^{(L)t_0,...,t_k}$) and $d_{id_i}^{t_0,...,t_k}$, and then couples human objects with tags based on the similarity scores.

We propose to use the Dynamic Time Warping (DTW) algorithm (84) to evaluate the similarity between the two temporal sequences. DTW is an algorithm designed to deal with different timing or moving speeds among the two time series data, it captures the similarity by dynamically warping the two sequences and finding the best matching. A larger DTW value implies a larger difference and hence a lower similarity. Therefore, we can define the similarity score as follows:

$$sim_{j,i}^{key} = \frac{1}{1 + DTW(d_j^{(key)t_0,...,t_k}, d_{id_i}^{t_0,...,t_k})}, key \in \{L, R\} \tag{4.5}$$

and normalize the score range between 0 and 1. The closer the similarity score $sim_{j,i}$ is to 1, the higher probability that the two trajectories should be paired.

After similarity scoring, we get Sim, which is a two-dimensional array storing all the similarity scores between $2J$ visual trajectories and I RF trajectories. Intu-

itively, we can match the pair that produces the highest similarity score. However, in real experiment, it is observed that the highest similarity score does not always imply the highest matching probability, since sometimes for the a RF trajectory $d_{id_i}^{t_0,\ldots,t_k}$, its similarity scores with other visual trajectories may all be small and quite similar, and it will be meaningless to match the pair with the highest score in this case.

To deal with such matching ambiguity, we further take 'kurtosis' (84), the confidence of the similarity scores into consideration. Kurtosis (85) is a measure of the 'tailedness' of the probability distribution of a real-valued random variable. Generally, data sets with high kurtosis tend to have heavy tails (or outliers) while data sets with low kurtosis tend to have light tails (or lack of outliers). For a RF trajectory $d_{id_i}^{t_0,\ldots,t_k}$, if it has a similarity score $sim_{j,i}$ with a visual trajectory which is much larger than all the other pairs, we will obtain a higher kurtosis value, which implies a higher confidence of matching the visual trace with the highest score. To realize this, we calculate the kurtosis values of each RF trajectory $d_{id_i}^{t_0,\ldots,t_k}$ as followed:

$$Kurt(sim_{1:J,i}^{key}) = E[(\frac{sim_{j,i}^{key} - \mu}{\sigma})^4], key \in \{L, R\} \tag{4.6}$$

where $sim_{1:J,i}^{key}$ are $d_{id_i}^{t_0,\ldots,t_k}$'s similarity scores with all visual traces, μ is the mean of $sim_{1:J,i}^{key}$ and σ is the standard deviation of $sim_{1:J,i}^{key}$.

During the pair matching, we first sort the RF trajectories according to their $Kurt(sim_{1:J,i}^{key})$ in a descending order. The RF trajectory with the highest confidence will determined as the target's trace, and will first be matched with its most similar visual trajectory. The matched pair $(d_{id_i}^{t_0,\ldots,t_k}, d_j^{(key)t_0,t_1,\ldots,t_k})$ is then added to the result

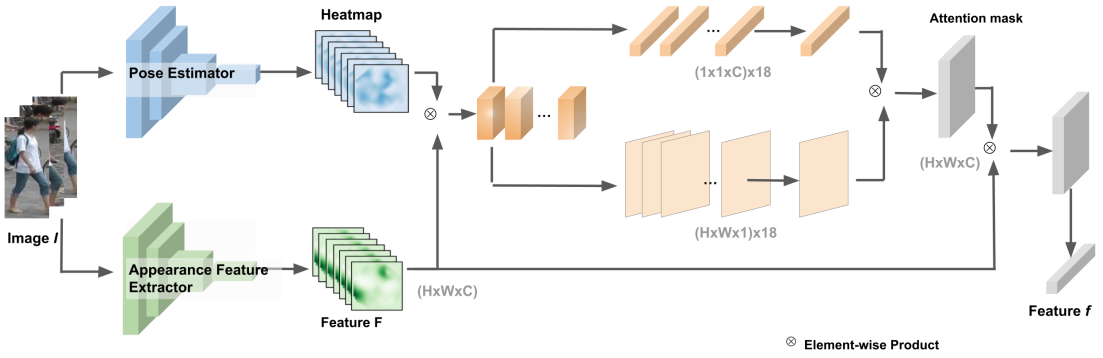


Figure 4.7: Overview of Re-ID system

set $\mathcal{P}$ and excluded from the remaining pairing procedure. The pairing procedure will keep repeating every time when new UA action occurs until all RF trajectories in FoV have been paired. The final result is the pairing set $\mathcal{P} = \{(d_{id_i}^{t_0,\ldots,t_k}, d_j^{(key)t_0,t_1,\ldots,t_k})\}_{i=1}^{I}$.

4.3.4.4 Long-Term Feature Pairing

Essentially, object-ID pairing in Section. 4.3.4.3 is to associate the PID assigned by CV system with the TID, and such pairing could be unstable in dynamic environment. It is observed in real experiment that when a target walks close to the edge of FoV, PID-switch problem could happens, a new PID will be randomly assigned to the this target, breaking the mapping between old PID and its paired TID. Therefore in this section, we further leverage long-term target feature information to resolve the PID-switch issue.

After we obtain the pairing result $\mathcal{P}$, we first construct a dynamic dictionary $\mathcal{M}$ with the Re-ID features $\boldsymbol{f}$ of all paired targets in $\mathcal{P}$, here $\mathcal{M} = \{id_i : f_i, \}_{i=1}^{I}$. Such Re-ID feature is extracted at the specific frame when the target first gets matched. Basically when new PIDs appear in the FoV, the status of TIDs can be classified into

two different cases:

- When there's no new TID arriving or leaving the FoV, implying that some TIDs in $\mathcal{P}$ are longer being matched, hence new PIDs should be able to repaired with these TIDs. If only one new PID has been caught, we directly pair this PID to the unmatched TID. For the case when more than one PID has changed, we need to first build a candidate set $\mathcal{S}_C$ with the features of unmatched TIDs from $\mathcal{M}$, and then construct a query set $\mathcal{S}Z_Q$ by extracting the features of persons marked with new PIDs. The similarity score between representative feature in $\mathcal{S}_Q$ and those from $\mathcal{S}_C$ is computed using Euclidean distance measurement. A low score means that the two features are correlated and vice-versa. We then sort the results from low to high and determine the matched pairs at the end.

- When new TID arrives, the feature of new person has not been updated in $\mathcal{M}$, hence RV-Track will only wait for UA action of this new target.

We extract the long-term target feature $\boldsymbol{f}$ using a lightweight Re-ID network. Fig. 4.7 shows the simple diagram of our proposed lightweight Re-ID model. This model consists of two basic branches, namely **appearance branch** and **pose branch**, and an **attention-based aggregation module**. It receives an image I as an input and outputs a feature representation f as illustrated.

Two Branches The appearance branch is designed to extract user's global feature, using Omni-Scale Network (OSNet) (86) as backbone. OSNet is a lightweight Re-ID model that is more than one order of magnitude smaller than the popular

ResNet50-based models, while still achieves state-of-the-art performance on 6 person Re-ID datasets and beats much larger networks (86). We remove the final classification layer (linear FC + softmax) mounted on the top of original OSNet which classifies the class of each input to retain the global feature. Formally, we denote the feature extracted from appearance backbone as $\mathbf{F}$, $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$, where H, W, and C denote the height, width, and the number of channels of the feature, respectively.

The pose branch is designed for the generation of heatmap using pose extractor AlphaPose. Based on the predicted landmark coordinate and the coordinate confidence, Alphapose outputs the heatmap of each landmark. Inspired by (87), we set a confidence threshold γ to remove the landmark heatmap information if the output confidence is less than γ, since the landmark with small confidence score is highly likely to be occluded in original image. Generally, the landmark coordinate can be expressed as follows:

$$LM_i = \begin{cases} (x_i, y_i) & \text{if } c_i \geq \gamma \\ 0 & \text{others} \end{cases} \quad (i = 1, ..., N) \tag{4.7}$$

Where LM_i denotes the i-th landmark and (x_i, y_i) are the corresponding pixel coordinates values, and γ is threshold. N is the number of landmarks ($N = 17$). These landmarks are then utilized to generate heatmaps consisting of a 2D Gussian centered on the ground truth location (87). The value of the corresponding heatmap is set to 0 if $LM_i = 0$. We denote the generated heatmap as $\mathbf{m}_i (i = 1, ..., N)$.

Later we combine the heatmaps and features $\mathbf{F}$ and feed the fused features into an attention module to improve the distinctiveness of features.

Attention Module. Given the feature map $\mathbf{F}$ from OSNet and heatmap of the i-th body part $\mathbf{m}_i$ from AlphaPose model, we first perform multiplication operation between them to get the fused features which reflect the potential characteristics of different body parts:

$$f^i = \mathbf{F} \bigotimes \mathbf{m}_i (i = 1, 2, ..., N) \tag{4.8}$$

where f^i represents the feature of the i-th body part. The proposed attention mechanism A_{mask} is comprised of two attention components: 1) **channel attention** A_{Ch} which emphasizes interdependent channel maps by integrating associated features among all channel maps, and 2) **position attention** A_{Posi} which selectively aggregates the feature at each position by a weighted sum of the features from all positions (88).

Channel Attention. Since the fused body part-level feature f^i can be formulated as $f^i = \{ch^i_j\}^M_{j=1}$, where ch^i_j denotes the feature vector at the j-th position of the i-th body part whose dimension is $\mathbb{R}^{1\times1\times C}$, and $M = H \times W$. We can aggregate the position information as follows:

$$Ch^i_{avg} = Relu(\frac{1}{M}\sum_{i=1}^{M} ch^i_j w_c) \tag{4.9}$$

and w_c are the weight parameters which can be learned during training, and $Ch^i_{avg} \in \mathbb{R}^{1\times1\times C}$ is the average value of the weighted sum from each channel. And the channel attention mask is the average of these N masks:

$$A_{Ch} = \frac{1}{N}\sum_{i=1}^{N} Ch^i_{avg} \tag{4.10}$$

Position Attention. Feature f^i can also be expressed as $f^i = \{posi^i_j\}^C_{j=1}$, in which we first obtain the i-th part-level feature map set $\{posi^i_j\}$ represents the feature

at the j-th channel with size of $\mathbb{R}^{H\times W\times 1}$ (C is the number of channels). Similar to channel attention, we aggregate the position information as follows

$$Posi^i_{avg} = Relu(\frac{1}{C}\sum_{i=1}^{C} posi^i_j w_p) \tag{4.11}$$

where w_p are the weight parameters needed to be learned, and $Posi^i_{avg} \in \mathbb{R}^{H\times W\times 1}$. Finally, we directly compute the averaged position attention mask of N parts as:

$$A_{Posi} = \frac{1}{N}\sum_{a=1}^{N} Posi^i_{avg} \tag{4.12}$$

Aggregated Attention. After we learn the two attention masks, we integrate them together to obtain a complementary multi-dimensional attention mask. We conduct an element-wise product to accomplish the attention feature fusion. The final attention mask $\mathbf{A}_{mask}$ is a matrix with the size of $H \times W \times C$.

$$\mathbf{A}_{mask} = Relu((A_{Posi} \bigotimes A_{Ch})W_a) \tag{4.13}$$

where W_a is the weight matrix. Finally, we perform an element-wise multiplication between the $\mathbf{A}_{mask}$ and the global feature from our backbone OSNet and output the body part-level attention feature map $\mathbf{F}_{mask}$.

$$\mathbf{F}_{mask} = \mathbf{F} \bigotimes \mathbf{A}_{mask} \tag{4.14}$$

Note that feature map $\mathbf{F}_{mask}$ is fed into a FC layer and converted into the user feature $\boldsymbol{f}$ which has a dimension of ζ. We discuss the choice of ζ in Section. 4.5.3. A classification layer is adopted at the end to classify the class of each input. We leverage the widely-used triplet loss function (89) during the training phase, the detailed training process is described in Section. 4.5.3.

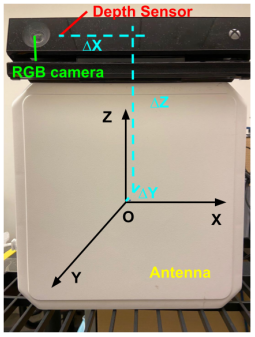


Figure 4.8: A RFTrack node prototype

4.3.5 Data Augmentation

When training with smaller dataset, deep neural networks tend to overfit. To avoid overfiting and achieve better training effectiveness, RV-Track conducts the data augmentation to enrich the training data using two signal transformation approaches, namely **jittering** and **time-shifting-and-warpping** (90). Jittering is a simple yet effective method to add random noise to the original data without changing the length of original data. Time-shifting-and-warpping is another useful data augmentation scheme. It is common that the user walks through RV-Track node with different speeds. Such differences will cause the collected RSS/phase samples scale various ranges. To mitigate above impact, we use shifting-and-warping-based augmentation to generate such data. Note that network in Fig. 4.2 takes the augmented data and the original data together as inputs for training.

4.4 System Deployment

4.4.1 Experimental Setup

We utilize a sensor setting similar to (**?**) for our experiments. Each RV-Track node consists of a Kinect depth camera and Laird S9028-PCL directional antenna. The Kinect camera is deployed on the top of the antenna, which is connected to an ImpinJ Speedway modeled R420 RFID reader. (Note that we have also tried implementing the system using dual RGB cameras, and estimating the depth of the moving target to recover her/his trajectory based on classical stereo matching algorithms such as semi-global matching, as well as some deep learning-based end-to-end approaches. Compared with the RGB-D camera, the dual-RGB system tends to have higher calibration overhead and lower matching accuracy, hence we chose to adopt the RGB-D camera in our system.) The center of the antenna is chosen as the origin O of 3D localization reference system, and we measure the coordination $[\Delta X, \Delta Y, \Delta Z]^T$ of Kinect depth sensor in this reference space. Then we could translate the 3D point cloud in Kinect reference system into the RF reference system. The tags employed are model Alien 964X, which we have observe inductive coupling in (8). While conducting the experiment, the tag carried by each person is attached on a piece of cardboard and each person needs to hold this cardboard to interact with the item tag, in order to eliminate the extra influence brought by human hand. The transmission and receiving gains are both 32dB. All experiments are conducted in an office room with various multi-path reflectors, moving objects, wireless signals (WiFi, LTE, and Bluetooth), electrical devices (servers, work-

stations, printers, refrigerator), and magnetic fields (whiteboard). The deployed ImpinJ Speedway modeled R420 RFID reader is sufficient to cover such a large indoor area (6m $\times$ 9m).

Softwares: The RFID reader and Kinect data collection program is implemented with *C#* code on a Dell desktop, equipped with Intel Core i7-7700 CPU at 3.6GHz and 16G memory. The software is designed using the Low-level Reader Protocol (LLRP) and Kinect Toolkit and outputs phase and RSS readings as well as video streams. The proposed Re-ID model are implemented in Python, based on Pytorch framework, and are trained and tested at a desktop (served as remote server) with one NVIDIA RTX 3070.

4.5 Experiments

4.5.1 Datasets

In our default experiments, we conduct two rounds of experiments and collect two datasets, namely $\mathcal{D}_{au}$ and $\mathcal{D}_{reid}$. $\mathcal{D}_{au}$ refers to an RFID RSS dataset collected for detecting the user authentication action in the first round (video frames are also collected and serve as the ground truth), which we recruit 11 volunteers (5 males and 6 females) for experiments. During the collection, we ask each volunteer to perform user authentication for 150 times, and collect RFID RSS measurements throughout the process. We ensure that only one volunteer is in the sensing area and no other person moves around. The experiments are conduct in two different locations in the same office.

Model	Accuracy(%)
Feature-based	87.07 ± 0.07
2-layer LSTM	90.25 ± 0.09
LSTM-FCN	93.27 ± 0.13

Table 4.1: Prediction accuracy on the collected dataset D_{au}

Each RSS sequence will late be segmented and manually labeled with 'no-coupling state' or 'coupling state'.

In the second round, we recruit 7 volunteers (3 males and 4 females) and collect $\mathcal{D}_{reid}$ by asking the volunteers to perform user authentication action one by one and then randomly walks around the sensing area. Note that $\mathcal{D}_{reid}$ contains RFID RSS, phase data as well as video frames from Kinect. These experiments are still conduct in the same locations as those in first round. In total, 700 trials are conduct for 2-person scenario and 200 trails for 3-person and 4-person case, respectively. The identity of the users for each experiment are manually labeled as well. Note that this work focuses on common indoor environments such as home and office with no special considerations given to building materials, room selection or furniture placement.

4.5.2 User Authentication

While evaluating the performance of LSTM-FCN classifier, the dataset $\mathcal{D}_{au}$ is randomly split into training $\mathcal{D}_{au}^{train} = 0.75\mathcal{D}_{au}$ and testing $\mathcal{D}_{au}^{test} = 0.25\mathcal{D}_{au}$ sets. Furthermore, 0.8 $\mathcal{D}_{au}^{train}$ of the training set is used to train the model, and the remaining

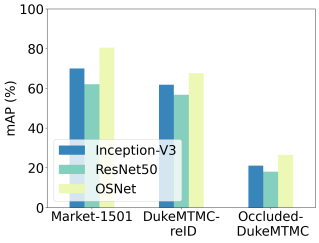


Figure 4.9: Backbone selection

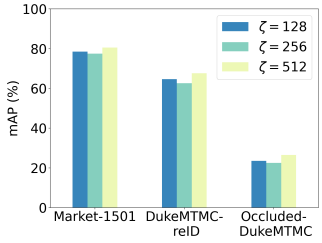


Figure 4.10: Feature dimension comparison

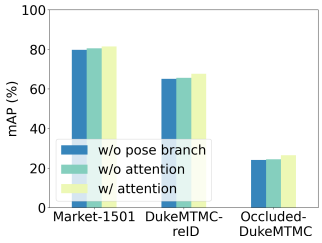


Figure 4.11: Effect of pose branch and attention module

$\mathcal{D}_{au}^{val} = 0.2\mathcal{D}_{au}^{train}$ is used as a validation set.

4.5.2.1 Evaluation Metric

We evaluate the performance of UA module using the accuracy measure, which is defined as the percentage of correctly predicted coupling events. For a set of test RSS samples $X = \{x_1, x_2, ..., x_i, ..., x_n\}$ with corresponding true labels $Y = \{y_1, y_2, ..., y_i, ..., y_n\}$, the accuracy is defined as:

$$accuracy = 100 \times \frac{1}{n} \sum_{i=1}^{n} I_{y_i}(\tilde{y_i}) \quad \% \tag{4.15}$$

where $\tilde{y_i}$ is the system prediction for input x_i, and $I_{y_i}(\tilde{y_i}) = 1$ if $\tilde{y_i} = y_i$ and $I_{y_i}(\tilde{y_i}) = 0$ otherwise.

4.5.2.2 Implementations & Performance

We compare the results with two different models, namely feature-based classifier (84) and vanilla LSTM classifier (91), which are both widely used for TSC tasks. Both two classifiers are trained and evaluated with the dataset $\mathcal{D}_{au}$. For feature-based

method, we first extract 7 low-level features, including *mean, variance, skewness, kurtosis, median, zero crossing rate, max-min difference*, from each sequence sample in dataset. Later the handcrafted features of each sample are feed to a binary classifier to determine whether this sample belongs to a 'coupling state' or not. We adopt several different training algorithms, namely K Nearest Neighbor (KNN), Decision Tree, Random Forest and Support Vector Machine (SVM), and use an ensemble manner to combine their classification results. And the vanilla LSTM classifier is basically a 2-layer LSTM classifier, where each layer has the same parameters setting as the LSTM branch in LSTM-FCN. As can be seen from the result (Table. 4.1), LSTM-FCN classifier outperforms the rest two in detecting 'coupling state'.

4.5.3 User Re-ID Model

4.5.3.1 Dataset

To demonstrate the efficacy of the proposed lightweight Re-ID model, we first evaluate the proposed model on two holistic datasets: Market-1501 (92), DukeMTMC-reID (93) and one partial dataset: Occluede-DukeMTMC (87).

Market-1501 is a large-scale person re-ID dataset collected from six cameras. There are 32,668 DPM-detected pedestrian image boxes of 1,501 identities: 750 identities are utilized for training and the remaining 751 identities are used for testing. There are 3,368 query images and 19,732 gallery images with 2,793 distractors.

DukeMTMC-reID is a subset of DukeMTMC (94) for person ReID. The images are cropped by hand-drawn bounding boxes. The data was taken from 8 cameras

of 1,404 identities with respective 16,522, 2,228 and 17,661 images in the training, query and gallery sets.

Occluded-DukeMTMC is currently the largest occluded person re-ID dataset derived from the original DukeMTMC dataset. It has 15,618 pedestrian images in the training set that contains 702 identities. And its testing set covers 17,661 gallery images and 2,210 query images including 1,110 identities.

4.5.3.2 Evaluation Metric

We use mean average precision (mAP) to evaluate the the quality of proposed re-identification model. In the following results we will report the mAP scores for Market-1501, DukeMTMC-reID, and Occluded-DukeMTMC where more than one ground truth images are in the gallery.

4.5.3.3 Implementation details.

We adopt OSNet as our appearance branch backbone with removing the final classification layers. The OSNet backbone weights are initialized using weights from the same OSNet architecture trained on Market-1501 dataset (86). The batch size is set to 40 and the training process is done for 100 epochs with the learning rate initialize at 0.015 and is decayed by 0.1 when reaching 40 and 80 epoch. We use AlphaPose pre-trained on COCO dataset(68) to obtain the landmarks from input images. The confidence score threshold γ is set to 0.2 (87).

4.5.3.4 Superiority of Proposed Network

We evaluate the proposed method in the following three aspects and investigate the superiority of two-branch network as well as the attention module on various datasets. Note that all the models do not use dilation.

Effect of backbone selection. Since the appearance branch leverages CNN-backbone for extracting features, the selection of CNN-backbone is essential for building a good feature extraction network. We implant multiple state-of-the-art CNN-based backbones with various model complexity into our Re-ID model and evaluate the overall performance, including Inception(95)and 50 layers ResNet (resnet50) (96) . As shown in Fig. 4.9 , our choice of OSNet achieves the best overall accuracy over three datasets compared with other backbones.

Effect of feature dimension size. As described in Section. 4.3.4.4, we use a variable f with size of ζ to represent user feature in Re-ID network. If ζ is too small, the feature may not be able to capture enough user information, while a feature with a large dimension will incur high computation overhead. Therefore, we need to determine a good dimension size. We separately explore three settings $\zeta = 128$, $\zeta = 256$, and $\zeta = 512$, and show their corresponding accuracy in Fig. 4.10. Since ζ =512 outperforms the other two on three datasets, hence we set ζ to 512 for the overall experiments.

Effect of pose branch and attention module We first compare our model with a baseline that does not explicitly use any body parts information, by directly adopting OSNet as the appearance feature extractor, and feeding the outputting 512-

dimensional image descriptor into a classifier to performance Re-ID task. Then we also compare with the baseline without any attention module to evaluate the impacts attention module imposed on the recognition performance. Fig. 4.11 compare the proposed method with the baseline, As the results illustrate, using part maps consistently improves the accuracy on all the three datasets from the baseline, and the integration of attention module will also remarkably enhance the effectiveness of our framework in terms of mAP.

4.5.4 Overall Performance

After demonstrating the efficacy of the proposed Re-ID model, we use samples in $\mathcal{D}_{reid}$ to evaluate the overall performance of RV-Track.

4.5.4.1 Evaluation Metric.

We still leverage the accuracy measure to evaluate the performance of short-term pairing module, while with its definition slightly different. For a set of persons with TID $T = \{x_1, x_2, ..., x_i, ..., x_k\}$ with corresponding true PID $P = \{y_1, y_2, ..., y_i, ..., y_k\}$, the accuracy is defined as the percentage of correctly pairing events:

$$accuracy = 100 \times \frac{1}{k} \sum_{i=1}^{k} I_{y_i}(\tilde{y_i}) \quad \% \tag{4.16}$$

where $\tilde{y_i}$ is the system matched PID for person with TID x_i, and $I_{y_i}(\tilde{y_i}) = 1$ if $\tilde{y_i} = y_i$ and $I_{y_i}(\tilde{y_i}) = 0$ otherwise.

Later, we check the improvement brought by long-term pairing module. 34 pid-switch events have been observed in $\mathcal{D}_{reid}$ (6 in 2-person trails, 12 in 3-person case

and 16 in 4-person case respectively), and we repair those PIDs using the long-term pairing method. In the matching process, the features for the retrieval process are taken from the Re-ID network in Section. 4.5.3. As can be seen in Table. 4.2, RV-Track achieves an average accuracy of 90.39% in short-term person identification and tracking. We correct 31 pid-switch events among 34, hence only a slight decrement is observed in long-term accuracy of 90.14% .

Trial Setting	Short-term Accuracy(%)	Long-term Accuracy(%)
2-person	92.92	92.92
3-person	91.00	91.00
4-person	87.25	86.50

Table 4.2: Overall accuracy

4.6 Conclusion

In this paper, we present RV-Track, a vision-RFID fusion system which can enable cameras to recognize the physical IDs of persons in their views and track the targets instantly when they appear in the video frames. By asking the users to perform a simple authentication, the system will be aware of the targets' IDs in its FoV. Later by exploiting the target features extracted from a light-weight Re-ID model, it can handle the PID-switch problems which often occur in vision tracking systems. Experiments

show that RV-Track can actively identify and track the RFID-tagged target with a accuracy of 90.14%.

Chapter 5

RF-HGR and the future research

5.1 Introduction

Recent wireless sensing techniques such as WiFi (97), RFID (81), and millimeter radar (98) have provided promising solutions for human gesture recognition (HGR) due to their characteristics of privacy-protection, device-free and ubiquity. Among them, RFID has attracted much attention given its low-cost and light-weight superiorities. A typical RFID-based HGR system generally contains three steps, namely, gesture class pre-definition, data collection, as well as recognition model training and testing. Users first need to pre-define a group of gestures classes (which are termed as *base classes*) to recognize, and then collect sufficient signal samples for each base class to build a training set (*base dataset*). A recognition model will be trained based on this training set, which can provide accurate recognition on these base gestures once well trained.

However, most of existing approaches mainly focus on coarse-grained gestures

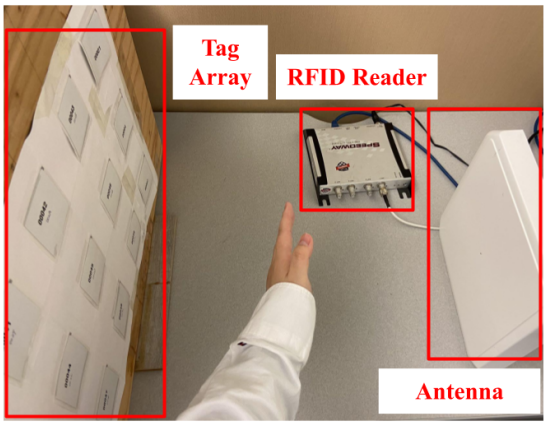


Figure 5.1: Experiment setup

like 'pull' or 'push', while many gesture based applications need to sense more complicated and fine-grained gestures like finger movements and sign gestures. When it comes to unseen gesture classes that are not included in the base dataset, the recognition performance will degrade dramatically (99), such limited scalability results in two kinds of overhead when trying to infer unseen gesture classes. For one thing, sufficient signal samples have to be re-collected for each unseen class to meet the demand of recognizing novel gestures, causing heavy **data collection overhead** and bringing massive inconvenience to users. Meanwhile, the recognition model has to be re-trained using all training data as long as a single unseen class is introduced. Such **training overhead** is non-trivial as it usually involves a complex model structure (such as RNN or LSTM (81)) to infer gestures information in RFID signals. Moreover, the recognition model trained with RFID signal samples collected from one specific environment also suffers from significant performance degradation when applied in a new dataset collected from a different place, as the collected wireless signal not only contains information of the performed gesture but also carries substantial information of environment, user,

location and even the orientation of user (100). We denote the factors uncorrelated to the gestures can as a *domain* (100) (*e.g.*, the environment or the user). Consequently, labor-intensive data collection and labeling efforts as well as model re-training efforts are required for each new domain, which further impedes the generalization of RFID-based HGR techniques.

Hence, there is an urgent demand for a lightweight cross-domain sensing approach in which users only need to collect few signal samples (*i.e.*, few shots) for any unseen class, and the model only needs to be fine-tuned using these samples. In this paper, we explore the possibility of employing domain-adaptive few-shot learning (DA-FSL) to address the scalability issue and domain shift challenge, and present RF-HGR, a novel RFID-based HGR system to recognize unseen gestures in different domain settings. Our preliminary experiments evaluate RF-HGR with the dataset collected from real-world experiments and the results show that the system is capable of producing high accuracy models for the HAR task using As far as we know, this is one of the first attempts addressing both the few-shot DA and few-shot recognition problems jointly in device-free HGR.

5.2 Methodology

RF-HGR system consists of three major modules, namely, data collection, signal preprocessing and deep learning module, as shown in Fig. 5.2. RF-HGR first collects raw phase and RSS data streams of tag array and passes them to the preprocessing module. Note that a data augmentation submodule is adopted to further improve the

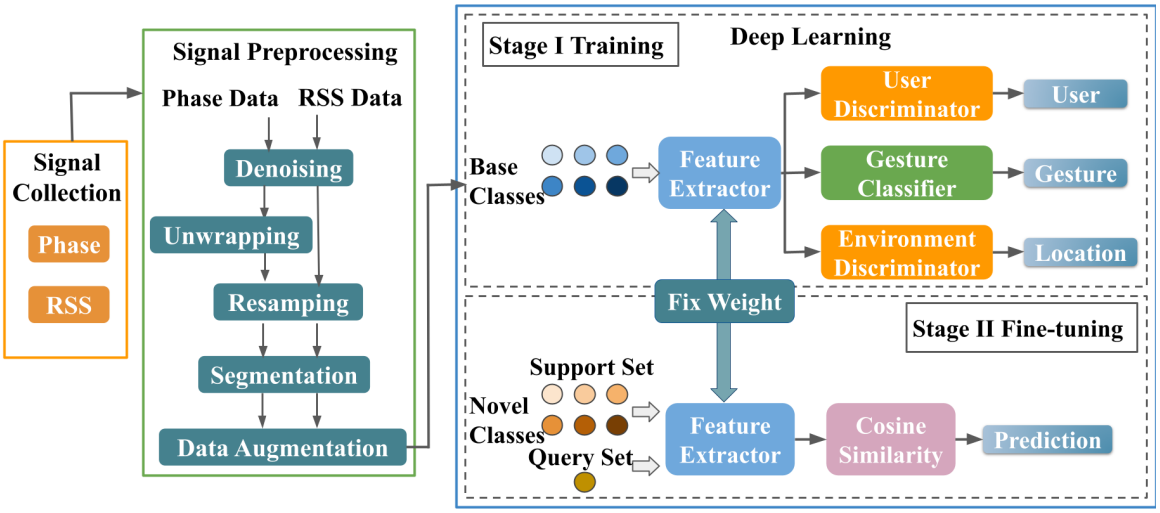


Figure 5.2: System Overview

expression ability and diversity of data. After preprocessing, the data streams are fed into the deep learning module for gesture recognition.

We employ metric-based FSL in deep learning module to recognize novel classes in cross domain scenarios, using only a few samples. Specifically, a two-stage approach is designed, in which we will first train a feature extractor using the base dataset and then fine-tuning a feature classifier using few-shot data. We formulate the few-shot classification problem as an N-way-K-shot classification in which the training set contains $J = K \times N$ samples from N classes, each with K samples. Note that the training set and test set are called support set $\mathcal{D}_{support}$ and query set $\mathcal{D}_{query}$ respectively.

Let (x^b, y^b) denote an input base RFID data vector and its corresponding ground-truth label. Then, the base dataset $\mathcal{D}_{base}$ is constructed leveraging the RF data samples collected for different gesture classes, where $\mathcal{D}_{base} = \{(x_1^b, y_1^b), (x_2^b, y_2^b), ..., (x_N^b, y_N^b)\}$.

In the first stage, we use a standard supervised learning manner to train a feature extractor f_θ together with a gesture classifier $C(\cdot|W, b)$ using the data in $\mathcal{D}_{base}$. The ample number of classes in the base dataset enables the feature extractor to learn unique latent features from base classes so that the distances between instances of the same class are closer while instances of different classes are further apart. Specifically, we design a *RFID Transformer* as the feature extractor to learn high-level representation of RFID input. To further deal with domain-independent issues, we adopt two domain discriminators, namely user discriminator D_u, and environment discriminator D_e, to discard extraneous features specific to users or environments, while remain all features related with the classification task. Once the training is completed, the classifier and discriminators will be removed to obtain the trained feature extractor.

In the second stage, we fine-tune a new classifier. The sample in the query set $\mathcal{D}_{query}$ will be classified based on the cosine similarity of the extracted features between the query sample and the labeled support samples from $\mathcal{D}_{support}$. A classifier with the trainable weight matrix $W \in \mathbb{R}^{d\times c}$ is added after the pre-trained feature extractor f_θ. We fine-tune this weight matrix W using support samples while keeping the parameters of feature extractor f_θ fixed. To classify a query sample x^q, we first extract an embedded feature vector $f_\theta(x^q)$ and then compute its cosine similarity pair-wisely with each weight vector in W to obtain the similarity scores for all classes. The softmax function σ maps the scores to a probability distribution, and the query sample will be classified to the class with the highest probability.

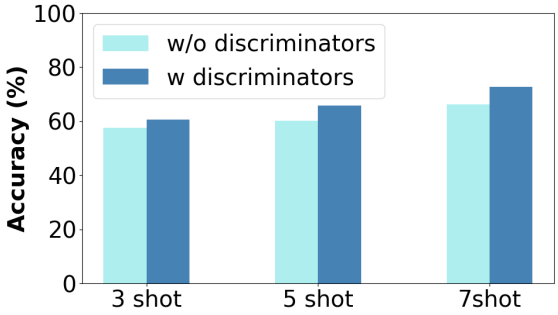


Figure 5.3: Overall accuracy of RF-HGR

5.3 Evaluation

Evaluation setup The prototype of RF-HGR is built with COTS UHF RFID devices. We utilize a ImpinJ Speedway modeled R420 RFID reader with a Laird S9028-PCL directional antenna to capture signals from the 4×4 model Alien-9629 RFID tag array with tags spaced 4cm apart. The proposed FSL model is trained and tested at a Dell desktop equipped with Intel Core i7-7700 CPU at 3.6GHz and one NVIDIA RTX 3070. We conduct our preliminary study in a laboratory environment as shown in Fig. 5.2 and recruit 2 female volunteers to participate in our experiments. The participants are asked to perform the gestures in the sensing area. In total, more than 2500 signal samples of 25 different gesture classes are collected, in which 5 gestures ('push and pull', 'sweep', 'slide', 'clap', and 'draw triangle') are defined as the unseen gestures while the base dataset involves the same 20 gesture classes as the ones in training dataset introduced by (99). Note that in the unseen dataset, the support set and query set of the same gesture class are collected from the same domain.

Preliminary results We leverage *accuracy*, which shows the probability that

an unseen signal sample can be correctly recognized, to evaluate the performance of RF-HGR. We calculate the accuracy to recognize unseen gestures in the three-, five-, and seven-shot settings and show the results in Fig. 5.3. It can be observed that in these settings, the recognition accuracy of RF-HGR is 50.6%, 65.7% and 72.8% respectively, demonstrating that RF-HGR can achieve a relatively high accuracy even if just learning from a few samples.

5.4 Further Approach

The experiment of RF-HGR is still at the early stage. In the future study, we will explore RF-HGR in the following directions:

- Expand the system into sequence recognition.
- Extend the system in various environments.
- Explore the possibility of leveraging the tag array and the antenna array to localize and track the hand movements of the user.

Chapter 6

Conclusion

In this thesis, we first design and implement TagMii, a noval approach to enable user-item interactions using passive RFID tags. Compared to other solutions that require a user to carry non-trivial hardware, TagMii only requires every user to carry a passive tag. The reader deployed in the environment monitors the interaction events and pair the user tag and the corresponding item tag. The key advantage of TagMii is that it is cost-efficient and especially convenient for children, some seniors, people with certain disabilities, and others who do not operate smartphones. We evaluate TagMii in complex environments with rich multipath, mobility, wireless signals, and magnetic fields and and TagMii to be accurate in recognizing user-item interactions in various setups.

Then we take a step further and propose ChopTags to deal with the inherent scalability limitation of TagMii. ChopTags achieves near 100% accuracy and the reader deployed in the environment can simultaneously monitor the interaction events of a large

number of tag modules. ChopTags provides low cost, high accuracy, high scalability, user-friendliness, and some level of privacy. We propose to implement two application prototypes based on ChopTags and evaluate the prototypes in complex environments to show that ChopTags is highly accurate and reliable with low latency. In future work, we aim to explore the possibility of enabling cameras in public areas to recognize the physical-identity(ID) of a person in the fields of view and track the persons with specific IDs with least or even no training effort.

We also notice that state-of-the-art computer vision methods can effectively detect and track objects/persons only if their appearances are pre-registered, while the appearance-collection process is quite time-consuming and cumbersome. Based on this issue, we present RVTrack, a vision-RFID fusion system which can enable cameras to recognize the physical IDs of persons in their views and track the targets instantly when they appear in the video frames. By asking the users to perform a simple authentication, the system will be aware of the targets' IDs in its FoV. Later by exploiting the target features extracted from a light-weight Re-ID model, it can handle the PID-switch problems which often occur in vision tracking systems. Experiments show that RVTrack can actively identify and track the RFID-tagged target with an accuracy of 90.14%.

We further explore the possibility of employing domain- adaptive few-shot learning (DA-FSL) to address the scalability issue and domain shift challenge in RFID-based human gesture recognition tasks, and present RF-HGR, a novel RFID-based HGR system to recognize unseen gestures in different domain settings. RF-HGR is one of the first attempts to jointly address the few-shot DA and few-shot recognition problems in

device-free HGR, in which users only need to collect a few signal samples for any unseen class, and the model only needs to be fine-tuned using these samples. Our experiments evaluate RF-HGR with the dataset collected from real-world experiments. The results show that the system is capable of producing high accuracy models for the HAR task.

6.1 Contributions

The thesis contributed the following:

- A survey of state-of-the-art identification and sensing applications for UHF RFID systems.
- Design and implement a novel approach to enable user–item interactions using passive RFID tags.
- Propose a new solution for high-accuracy and scalable user–item interactions using passive RFID tags, which is the first work to combine the information of both tag ID presence and physical signal features to infer interactions.
- Design a vision-RFID fusion system which can enable cameras to recognize the physical IDs of persons in their views and track the targets instantly when they appear in the video frames.
- A demonstration of all application prototypes and evaluate the prototypes in complex environments.

Bibliography

[1] H. Landaluce, L. Arjona, A. Perallos, F. Falcone, I. Angulo, and F. Muralter, "A review of iot sensing applications and challenges using rfid and wireless sensor networks," *Sensors*, vol. 20, no. 9, p. 2495, 2020.

[2] J. Su, Z. Sheng, V. C. Leung, and Y. Chen, "Energy efficient tag identification algorithms for rfid: survey, motivation and new design," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 118–124, 2019.

[3] W. Yao, C.-H. Chu, and Z. Li, "The adoption and implementation of rfid technologies in healthcare: a literature review," *Journal of medical systems*, vol. 36, pp. 3507–3525, 2012.

[4] M. Ahmad, T. M. Ghazal, and N. Aziz, "A survey on animal identification techniques past and present," *International Journal of Computational and Innovative Sciences*, vol. 1, no. 2, pp. 1–7, 2022.

[5] W. C. Tan and M. S. Sidhu, "Review of rfid and iot integration in supply chain management," *Operations Research Perspectives*, vol. 9, p. 100229, 2022.

[6] F. Chetouane, "An overview on rfid technology instruction and application," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 382–387, 2015.

[7] R. M. Ferdous, A. W. Reza, and M. F. Siddiqui, "Renewable energy harvesting for wireless sensors using passive rfid tag technology: A review," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 1114–1128, 2016.

[8] H. Cai, G. Wang, X. Shi, J. Xie, M. Wang, and C. Qian, "When tags 'read'each other: Enabling low-cost and convenient tag mutual identification," in *2019 IEEE 27th International Conference on Network Protocols (ICNP)*. IEEE, 2019, pp. 1–11.

[9] J. Wang and D. Katabi, "Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight," in *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4. ACM, 2013, pp. 51–62.

[10] H. Cai, G. Wang, J. Leyva, I. Pham, J. Han, S. Chen, and C. Qian, "Choptags: An accurate and low-cost interface to identify user/item interactions," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2022, pp. 244–252.

[11] H. Cai and C. Qian, "Enabling identity-aware tracking by vision-rfid fusion," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 675–676.

[12] ——, "Poster: Rf-hgr: Domain-independent few-shot recognition for unseen ges-

tures with rfid," in *2022 IEEE 30th International Conference on Network Protocols (ICNP)*. IEEE, 2022, pp. 1–2.

[13] E. Global, "Specification for RFID Air Interface EPC™ Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz-960 MHz," Technical report, GS1, Tech. Rep., 2008.

[14] M. Kodialam and T. Nandagopal, "Fast and Reliable Estimation Schemes in RFID Systems," in *Proc. of ACM MOBICOM*, 2006.

[15] C. Qian, H. Ngan, Y. Liu, and L. M. Ni, "Cardinality Estimation for Large-scale RFID Systems," *IEEE Transactions on Parallel and Distributed Systems*, 2011.

[16] Z. Zhou, B. Chen, and H. Yu, "Understanding RFID Counting Protocols," *IEEE/ACM Transactions on Networking*, 2016.

[17] Y. Zhao, Y. Liu, and L. M. Ni, "VIRE: Active RFID-based Localization Using Vrtual Reference Elimination," in *Proceedings of IEEE ICPP*, 2007.

[18] X. Shi, M. Wang, G. Wang, B. Huang, H. Cai, J. Xie, and C. Qian, "TagAttention: Mobile Object Tracing without Object Appearance Information by Vision-RFID Fusion," in *Proceedings of IEEE ICNP*, 2019.

[19] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor Location Sensing Using Active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.

[20] C. Hekimian-Williams, B. Grant, X. Liu, Z. Zhang, and P. Kumar, "Accurate

Localization of RFID Tags Using Phase Difference," in *Proceedings of IEEE RFID*, 2010.

[21] T. Liu, L. Yang, Q. Lin, Y. Guo, and Y. Liu, "Anchor-free Backscatter Positioning for RFID Tags with High Accuracy," in *Proceedings of IEEE INFOCOM*, 2014.

[22] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proceedings of ACM MobiCom*. ACM, 2014, pp. 237–248.

[23] L. Shangguan, Z. Li, Z. Yang, M. Li, and Y. Liu, "OTrack: Order Tracking for Luggage in Mobile RFID Systems," in *Proceedings of IEEE INFOCOM*, 2013.

[24] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu, "Relative Localization of RFID Tags Using Spatial-Temporal Phase Profiling," in *Proceedings of USENIX NSDI*, 2015.

[25] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu, "See Through Walls with COTS RFID System!" in *Proceedings of ACM MOBICOM*, 2015.

[26] J. Han, C. Qian, X. Wang, D. Ma, J. Zhao, W. Xi, Z. Jiang, and Z. Wang, "Twins: Device-free Object Tracking Using Passive Tags," *IEEE/ACM Transactions on Networking (TON)*, vol. 24, no. 3, pp. 1605–1617, 2016.

[27] Y. Ma, N. Selby, and F. Adib, "Minding the Billions: Ultra-wideband Localization for Deployed RFID Tags," in *Proceedings of ACM Mobicom*, 2017.

[28] J. Han, H. Ding, C. Qian, D. Ma, W. Xi, Z. Wang, Z. Jiang, and L. Shangguan, "CBID: A Customer Behavior Identification System Using Passive Tags," in *Proceedings of IEEE ICNP*, 2014.

[29] L. Shangguan, Z. Zhou, X. Zheng, L. Yang, Y. Liu, and J. Han, "Shopminer: Mining Customer Shopping Behavior in Physical Clothing Stores with COTS RFID Devices," in *Proceedings of ACM SenSys*, 2015.

[30] H. Ding, C. Qian, J. Han, G. Wang, W. Xi, K. Zhao, and J. Zhao, "RFIPad: Enabling Cost-efficient and Device-free In-air Handwriting Using Passive Tags," in *Proceedings of IEEE ICDCS*, 2017.

[31] L. Yang, Y. Li, Q. Lin, X.-Y. Li, and Y. Liu, "Making Sense of Mechanical Vibration Period with Sub-millisecond Accuracy Using Backscatter Signals," in *Proceedings of ACM Mobicom*, 2016.

[32] "Phychips Inc." http://www.phychips.com/applications-main/.

[33] "ImpinJ Inc." http://www.impinj.com/.

[34] Z. An, Q. Lin, and L. Yang, "Cross-Frequency Communication: Near-Field Identification of UHF RFIDs with WiFi!" in *Proceedings of ACM MobiCom*. ACM, 2018, pp. 623–638.

[35] C. Turner, A. Casbard, and M. Murphy, "Barcode Technology: Its Role in Increasing the Safety of Blood Transfusion," *Transfusion*, vol. 43, no. 9, pp. 1200–1209, 2003.

[36] Y. Liu, J. Yang, and M. Liu, "Recognition of QR Code with Mobile Phones," in *2008 Chinese control and decision conference*. IEEE, 2008, pp. 203–206.

[37] Y. M. Kow, X. Gui, and W. Cheng, "Special Digital Monies: The Design of Alipay and Wechat Wallet for Mobile Payment Practices in China," in *IFIP Conference on Human-Computer Interaction*. Springer, 2017, pp. 136–155.

[38] M. Vazquez-Briseno, F. I. Hirata, J. de Dios Sanchez-Lopez, E. Jimenez-Garcia, C. Navarro-Cota, and J. I. Nieto-Hipolito, "Using RFID/NFC and QR-code in Mobile Phones to Link the Physical and the Digital World," in *Interactive Multimedia*. IntechOpen, 2012.

[39] M. K. Schultz, "A case study on the appropriateness of using quick response (qr) codes in libraries and museums," *Library & Information Science Research*, vol. 35, no. 3, pp. 207–215, 2013.

[40] T. Lotlikar, R. Kankapurkar, A. Parekar, and A. Mohite, "Comparative Study of Barcode, QR-code and RFID System," *Int. J. Comput. Technol. Appl.*, vol. 4, no. 5, pp. 817–821, 2013.

[41] P. Asadzadeh, L. Kulik, and E. Tanin, "Gesture recognition using rfid technology," *Personal and Ubiquitous Computing*, vol. 16, no. 3, pp. 225–234, 2012.

[42] J. Wang, D. Vasisht, and D. Katabi, "Rf-idraw: virtual touch screen in the air using rf signals," in *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4. ACM, 2014, pp. 235–246.

[43] H. Li, C. Ye, and A. P. Sample, "Idsense: A human object interaction detection system based on passive uhf rfid," in *Proceedings of ACM CHI*. ACM, 2015, pp. 2555–2564.

[44] H. Li, E. Brockmeyer, E. J. Carter, J. Fromm, S. E. Hudson, S. N. Patel, and A. Sample, "Paperid: A technique for drawing functional battery-free wireless interfaces on paper," in *Proceedings of ACM CHI*. ACM, 2016, pp. 5885–5896.

[45] S. Pradhan, E. Chai, K. Sundaresan, L. Qiu, M. A. Khojastepour, and S. Rangarajan, "RIO: A Pervasive RFID-based Touch Gesture Interface," in *Proceedings of ACM MobiCom*. ACM, 2017, pp. 261–274.

[46] C. Gao, Y. Li, and X. Zhang, "Livetag: Sensing human-object interaction through passive chipless wifi tags," in *15th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 18)*, 2018, pp. 533–546.

[47] M.-J. Hsieh, R.-H. Liang, D.-Y. Huang, J.-Y. Ke, and B.-Y. Chen, "Rfibricks: interactive building blocks based on rfid," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 189.

[48] R.-H. Liang, M.-J. Hsieh, J.-Y. Ke, J.-L. Guo, and B.-Y. Chen, "Rfimatch: distributed batteryless near-field identification using rfid-tagged magnet-biased reed switches," in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, 2018, pp. 473–483.

[49] K. Katsuragawa, J. Wang, Z. Shan, N. Ouyang, O. Abari, and D. Vogel, "Tip-tap:

Battery-free discrete 2d fingertip input," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 2019, pp. 1045–1057.

[50] H. Ding, J. Han, C. Qian, F. Xiao, G. Wang, N. Yang, W. Xi, and J. Xiao, "Trio: Utilizing tag interference for refined localization of passive rfid," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 828–836.

[51] Y. Á. López, M. E. de Cos Gómez, and F. L.-H. Andrés, "A received signal strength rfid-based indoor location system," *Sensors and Actuators A: Physical*, vol. 255, pp. 118–133, 2017.

[52] K. Chawla, C. McFarland, G. Robins, and C. Shope, "Real-time rfid localization using rss," in *2013 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, 2013, pp. 1–6.

[53] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu, "Stpp: Spatial-temporal phase profiling-based method for relative rfid tag localization," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 596–609, 2016.

[54] C. Jiang, Y. He, X. Zheng, and Y. Liu, "Orientation-aware rfid tracking with centimeter-level accuracy," in *Proceedings of ACM/IEEE IPSN*. IEEE Press, 2018, pp. 290–301.

[55] J. S. Choi, H. Lee, D. W. Engels, and R. Elmasri, "Passive uhf rfid-based localization using detection of tag interference on smart shelf," *IEEE Transactions*

on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 2, pp. 268–275, 2011.

[56] G. Wang, H. Cai, C. Qian, J. Han, X. Li, H. Ding, and J. Zhao, "Towards replay-resilient rfid authentication," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 385–399.

[57] G. Wang, C. Qian, J. Han, W. Xi, H. Ding, Z. Jiang, and J. Zhao, "Verifiable smart packaging with passive RFID," in *Proceedings of ACM UBICOMP*, 2016.

[58] J. Impin, "Speedway revolution reader application note: Low level user data support," *Speedway Revolution Reader Application Note*, 2010.

[59] E. R. Standard, "Epct m radio-frequency identity protocols class-1 generation-2 uhf rfid protocol for communications at 860mhz-960mhz version 1.0. 9," 2005.

[60] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, "TagScan: Simultaneous Target Imaging and Material Identification with Commodity RFID Devices," in *Proceedings of ACM MobiCom*. ACM, 2017, pp. 288–300.

[61] Q. Lin, L. Yang, C. Duan, and Y. Liu, "Revisiting reading rate with mobility: Rate-adaptive reading of cots rfid systems," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1631–1646, 2018.

[62] J. Wang, O. Abari, and S. Keshav, "Challenge: RFID Hacking for Fun and Profit," in *Proceedings of ACM MobiCom*. ACM, 2018, pp. 461–470.

[63] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu, "Relative Localization of RFID Tags Using Spatial-Temporal Phase Profiling." in *NSDI*, 2015, pp. 251–263.

[64] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer vision and image understanding*, vol. 189, p. 102805, 2019.

[65] S. Cao and H. Wang, "Enabling public cameras to talk to the public," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, pp. 1–20, 2018.

[66] B. Korany, C. R. Karanam, H. Cai, and Y. Mostofi, "Xmodal-id: Using wifi for through-wall person identification from candidate video footage," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–15.

[67] H. Li, P. Zhang, S. Al Moubayed, S. N. Patel, and A. P. Sample, "Id-match: A hybrid computer vision and rfid system for recognizing individuals in groups," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4933–4944.

[68] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2334–2343.

[69] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," *arXiv preprint arXiv:1603.00831*, 2016.

[70] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey," *Neurocomputing*, vol. 300, pp. 17–33, 2018.

[71] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: Learning to track multiple cues with long-term dependencies," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 300–311.

[72] K. Fang, Y. Xiang, X. Li, and S. Savarese, "Recurrent autoregressive networks for online multi-object tracking," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 466–475.

[73] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3645–3649.

[74] L. Leal-Taixé, C. Canton-Ferrer, and K. Schindler, "Learning by tracking: Siamese cnn for robust target association," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 33–40.

[75]

[76] C.-J. Wu, S. Houben, and N. Marquardt, "Eaglesense: Tracking people and devices in interactive spaces using real-time top-view depth-sensing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 3929–3942.

[77] C. Duan, X. Rao, L. Yang, and Y. Liu, "Fusing rfid and computer vision for fine-grained object tracking," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[78] B. Zhang, M. Li, X. Xie, L. Fu, X. Tong, and X. Liu, "Rc6d: An rfid and cv fusion system for real-time 6d object pose estimation," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 690–699.

[79] M. A. A. Haseeb and R. Parasuraman, "Wisture: Rnn-based learning of wireless signals for gesture recognition in unmodified smartphones," *arXiv preprint arXiv:1707.08569*, 2017.

[80] H. X. Tan, N. N. Aung, J. Tian, M. C. H. Chua, and Y. O. Yang, "Time series classification using a modified lstm approach from accelerometer-based data: A comparative study for gait cycle detection," *Gait & posture*, vol. 74, pp. 128–134, 2019.

[81] Y. Yu, D. Wang, R. Zhao, and Q. Zhang, "Rfid based real-time recognition of ongoing gesture with adversarial learning," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 298–310.

[82] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[83] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "Lstm fully convolutional networks for time series classification," *IEEE access*, vol. 6, pp. 1662–1669, 2017.

[84] R. Y.-C. Tsai, H. T.-Y. Ke, K. C.-J. Lin, and Y.-C. Tseng, "Enabling identity-aware tracking via fusion of visual and inertial features," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2260–2266.

[85] Wikipedia contributors, "Kurtosis — Wikipedia, the free encyclopedia," 2021, [Online; accessed 12-November-2021]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Kurtosis&oldid=1054446498

[86] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3702–3712.

[87] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, "Pose-guided feature alignment for occluded person re-identification," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 542–551.

[88] Y. Zhai, X. Han, W. Ma, X. Gou, and G. Xiao, "Pgmanet: Pose-guided mixed attention network for occluded person re-identification," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[89] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[90] B. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks. arxiv 2020," *arXiv preprint arXiv:2007.15951*.

[91] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *Plos one*, vol. 16, no. 7, p. e0254841, 2021.

[92] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1116–1124.

[93] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3754–3762.

[94] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *European conference on computer vision*. Springer, 2016, pp. 17–35.

[95] Y. Chen, X. Zhu, and S. Gong, "Person re-identification by deep learning multi-scale representations," in *Proceedings of the IEEE international conference on computer vision workshops*, 2017, pp. 2590–2600.

[96] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[97] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using wifi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 401–413.

[98] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–19, 2016.

[99] R. Xiao, J. Liu, J. Han, and K. Ren, "Onefi: One-shot recognition for unseen gesture via cots wifi," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 206–219.

[100] H. Kang, Q. Zhang, and Q. Huang, "Context-aware wireless-based cross-domain gesture recognition," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 503–13 515, 2021.

[101] J. Hack and P. Ramakrishnan, "Cvchess: Computer vision chess analytics," 2014.

[102] C. Koray, E. Sumer, V. Štruc *et al.*, "A computer vision system for chess game tracking," in *21st Computer Vision Winter Workshop Luka Cehovin*, 2016.

[103] M. A. Czyzewski, A. Laskowski, and S. Wasik, "Chessboard and chess piece recognition with the support of neural networks," *arXiv preprint arXiv:1708.03898*, 2017.

[104] J. Ding, "Chessvision: Chess board and piece recognition," Technical Report. Stanford University. URL: https://web. stanford. edu/class, Tech. Rep., 2016.

[105] "DGT Projects," http://www.dgtprojects.com/.

[106] L. Yang, Y. Qi, J. Fang, X. Ding, T. Liu, and M. Li, "Frogeye: Perception of the Slightest Tag Motion," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 2670–2678.

[107] L. Shangguan, Z. Zhou, and K. Jamieson, "Enabling Gesture-based Interactions with Objects," in *Proceedings of the ACM MobiSys*. ACM, 2017, pp. 239–251.

[108] H. Ding, C. Qian, J. Han, G. Wang, Z. Jiang, J. Zhao, and W. Xi, "Device-free Detection of Approach and Departure Behaviors using Backscatter Communication," in *Proceedings of ACM UbiComp*, 2016.

[109] G. Wang, C. Qian, L. Shangguan, H. Ding, J. Han, N. Yang, W. Xi, and J. Zhao, "HMRL: Relative Localization of RFID Tags with Static Devices," in *Sensing, Communication, and Networking (SECON), 2017 14th Annual IEEE International Conference on*. IEEE, 2017, pp. 1–9.

[110] J. Han, H. Ding, C. Qian, W. Xi, Z. Wang, Z. Jiang, L. Shangguan, and J. Zhao, "Cbid: A Customer Behavior Identification System Using Passive Tags," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2885–2898, 2016.

[111] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual Touch Screen in the Air Using RF Signals," in *Proceedings of ACM SIGCOMM*, 2014.

[112] X. Chen, F. Lu, and T. Y. Terry, "The "weak spots" in stacked UHF RFID tags

in NFC applications," in *RFID, 2010 IEEE International Conference on.* IEEE, 2010, pp. 181–186.

[113] R. K. Wangsness, "Electromagnetic Fields," *Electromagnetic Fields, 2nd Edition, by Roald K. Wangsness, pp. 608. ISBN 0-471-81186-6. Wiley-VCH, July 1986.*, p. 608, 1986.

[114] D. Vasisht, J. Wang, and D. Katabi, "Rf-idraw: Virtual Touch Screen in the Air Using RF Signals," in *Proceedings of the 6th Annual Workshop on Wireless of the Students, by the Students, for the Students.* ACM, 2014, pp. 1–4.

[115] M. Meeker, "Internet Trends 2014-Code Conference," *Retrieved May*, vol. 28, p. 2014, 2014.

[116] S. Hong-Ying, "The Application of Barcode Technology in Logistics and Warehouse Management," in *2009 First International Workshop on Education Technology and Computer Science*, vol. 3. IEEE, 2009, pp. 732–735.

[117] S. Hinske and M. Langheinrich, "An rfid-based infrastructure for automatically determining the position and orientation of game objects in tabletop games."

[118] J. Tan and P.-L. P. Rau, "A design of augmented tabletop game based on rfid technology," *Procedia Manufacturing*, vol. 3, pp. 2142–2148, 2015.

[119] C. Floerkemeier and F. Mattern, "Smart playing cards–enhancing the gaming experience with rfid."

[120] C. Magerkurth, A. D. Cheok, R. L. Mandryk, and T. Nilsen, "Pervasive games: bringing computer entertainment back to the real world," *Computers in Entertainment (CIE)*, vol. 3, no. 3, pp. 4–4, 2005.

[121] "Separation of Multiple Passive RFID Signals Using Software Defined Radio, author=Shen, Dawei and Woo, Grace and Reed, David P and Lippman, Andrew B and Wang, Junyu, booktitle=2009 IEEE International Conference on RFID, pages=139–146, year=2009, organization=IEEE."

[122] J. Yu, K. Liu, and G. Yan, "A novel rfid anti-collision algorithm based on sdma," in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*. IEEE, 2008, pp. 1–4.

[123] S. Thys, W. Van Ranst, and T. Goedemé, "Fooling automated surveillance cameras: adversarial patches to attack person detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[124] Z. Zhu, T. Huang, B. Shi, M. Yu, B. Wang, and X. Bai, "Progressive pose attention transfer for person image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2347–2356.

[125] R. Quan, X. Dong, Y. Wu, L. Zhu, and Y. Yang, "Auto-reid: Searching for a part-aware convnet for person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3750–3759.

[126] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2138–2147.

[127] K. Han, Z. Wang, and Z. Chen, "Fingerprint image enhancement method based on adaptive median filter," in *2018 24th Asia-Pacific Conference on Communications (APCC)*. IEEE, 2018, pp. 40–44.

[128] Y. Zhu, T. Tan, and Y. Wang, "Biometric personal identification based on iris patterns," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 2. IEEE, 2000, pp. 801–804.

[129] S. Pan, T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, H. Y. Noh, and P. Zhang, "Footprintid: Indoor pedestrian identification through ambient structural vibration sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–31, 2017.

[130] C. Huang, H. Chen, L. Yang, and Q. Zhang, "Breathlive: Liveness detection for heart sound authentication with deep breathing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–25, 2018.

[131] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, "Toward unobtrusive in-home

gait analysis based on radar micro-doppler signatures," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2629–2640, 2019.

[132] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "A human gait classification method based on radar doppler spectrograms," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–12, 2010.

[133] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 1–13, 2015.

[134] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[135] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 267–281.

[136] C. Lin, J. Hu, Y. Sun, F. Ma, L. Wang, and G. Wu, "Wiau: An accurate device-free authentication system with resnet," in *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2018, pp. 1–9.

[137] M. Shahzad and S. Zhang, "Augmenting user identification with wifi based gesture recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–27, 2018.

[138] A. Huang, D. Wang, R. Zhao, and Q. Zhang, "Au-id: Automatic user identification and authentication through the motions captured from sequential human activities using rfid," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–26, 2019.

[139] C. Zhao, Z. Li, T. Liu, H. Ding, J. Han, W. Xi, and R. Gui, "Rf-mehndi: A fingertip profiled rf identifier," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1513–1521.

[140] A. Kalyanaraman, D. Hong, E. Soltanaghaei, and K. Whitehouse, "Forma track: tracking people based on body shape," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–21, 2017.

[141] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 363–373.

[142] C. Wang, L. Xie, W. Wang, Y. Chen, Y. Bu, and S. Lu, "Rf-ecg: Heart rate variability assessment based on cots rfid tag array," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, pp. 1–26, 2018.

[143] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2016, pp. 75–82.

[144] H. Xu, D. Wang, R. Zhao, and Q. Zhang, "Faho: deep learning enhanced holographic localization for rfid tags," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 351–363.

[145] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "Grfid: A device-free rfid-based gesture recognition system," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 381–393, 2016.

[146] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4282–4291.

[147] B. Xie, J. Xiong, X. Chen, E. Chai, L. Li, Z. Tang, and D. Fang, "Tagtag: material sensing with commodity rfid," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, 2019, pp. 338–350.

[148] D. Chung, K. Tahboub, and E. J. Delp, "A two stream siamese convolutional neural network for person re-identification," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1983–1991.

[149] @_CHINOSAUR, ""venue is too cold" #bingo #chi2014," Tweet, May 2014,

retrieved Febuary 2, 2015 from https://twitter.com/_CHINOSAUR/status/461864317415989248.

[150] Nintendo R&D1 and Intelligent Systems, "*Super Metroid*," Game [SNES], Kyoto, Japan, April 1994, nintendo, Kyoto, Japan. Played August 2011.

[151] S. R. Klemmer, M. Thomsen, E. Phelps-Goodman, R. Lee, and J. A. Landay, "Where do web sites come from?: Capturing and interacting with design history," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '02. New York, NY, USA: ACM, 2002, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/503376.503378

[152] B. D. Mather, "Making up titles for conference papers," in *Ext. Abstracts CHI 2000*. ACM Press, 2000, pp. 1–2.

[153] M. Schwartz, *Guidelines for Bias-Free Writing*. Bloomington, IN, USA: ERIC, 1995.

[154] P. T. Zellweger, N. O. Bouvin, H. Jehøj, and J. D. Mackinlay, "Fluid annotations in an open world," in *Proc. Hypertext 2001*. ACM Press, 2001, pp. 9–18.

[155] ACM, "How to classify works using acm's computing classification system," 1998, http://www.acm.org/class/how_to_use.html.

[156] A. Cavender, S. Trewin, and V. Hanson, "Accessible writing guide," 2014, http://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-writing-guide/.

[157] R. E. Anderson, "Social Impacts of Computing: Codes of Professional Ethics," *Social Science Computer Review December*, vol. 10, no. 4, pp. 453–469, 1992.

[158] M. L. Heilig, "Sensorama simulator," U.S. Patent 3,050,870, August 1962, filed Februrary 22, 1962.

[159] J. Kaye and P. Dourish, "Special issue on science fiction and ubiquitous computing," *Personal and Ubiquitous Computing*, vol. 18, no. 4, pp. 765–766, 2014. [Online]. Available: http://dx.doi.org/10.1007/s00779-014-0773-4

[160] Psy, "Gangnam style," Video, July 2012, retrieved August 22, 2014 from https://www.youtube.com/watch?v=9bZkp7q19f0.

[161] W. Feng, Z. Hu, W. Wu, J. Yan, and W. Ouyang, "Multi-object tracking with multiple cues and switcher-aware classification," *arXiv preprint arXiv:1901.06129*, 2019.

[162] N. Wang, J. Shi, D.-Y. Yeung, and J. Jia, "Understanding and diagnosing visual tracking systems," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3101–3109.

[163] Y. Li, C. Huang, and R. Nevatia, "Learning to associate: Hybridboosted multi-target tracker for crowded scene," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 2953–2960.

[164] B. Leibe, K. Schindler, and L. Van Gool, "Coupled detection and trajectory es-

timation for multi-object tracking," in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.

[165] Y. Xiu, J. Li, H. Wang, Y. Fang, and C. Lu, "Pose Flow: Efficient online pose tracking," in *BMVC*, 2018.

[166] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018.

[167] I. E. Sutherland, "Sketchpad, a man-machine graphical communication system," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1963.

[168] L. Winner, *The Social Shaping of Technology*, 2nd ed. UK: Open University Press, 1999, ch. Do artifacts have politics?, pp. 28–40.