Enabling Novel Sensing and Interaction with Everyday Objects using Commercial RFID Systems

Hanchuan Li

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

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Shwetak N. Patel, Chair

Alanson P. Sample

Joshua R. Smith

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Enabling Novel Sensing and Interaction with Everyday Objects using Commercial RFID Systems

Hanchuan Li

Chair of the Supervisory Committee:
Dr. Shwetak N. Patel
Computer Science & Engineering

The Internet of Things (IoT) promises an interconnected network of smart devices that will revolutionize the way people interact with their surrounding environments. This distributed network of physical devices will open up tremendous opportunities for Human-Computer Interaction and Ubiquitous Computing, creating novel user-centered and context-aware sensing applications.

The advancement of IoT has been heavily focused on creating new and smart electronic devices, while the vast majority of everyday non-smart objects are left unchecked. Techniques based on active sensors are limited by their high deployment cost and the requirement for battery replacement. There currently exists a huge gap between the collection of devices integrated to the IoT and the remaining massive number of everyday physical objects.

Radio-frequency identification (RFID) has been widely adopted in the IoT industry as a standard inventory management infrastructure. In this thesis, I apply signal processing and machine learning techniques on low-level channel parameters of commercially available RFID tags to enable novel sensing and interaction with everyday objects. These new sensing capabilities allow for the system to understand daily activities, create tangible user interfaces, and enhance user experiences related to human-robot interactions and object interactions in augmented reality.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank all the people who have helped me to reach this significant milestone of my life.

First I would like to express my sincere gratitude to my family and friends. Your love and support motivate me to chase my passion and make nothing seem impossible.

Second, I would like to thank two people who are instrumental to my PhD study: Shwetak Patel and Alanson Sample. Shwetak, thank you for being an amazingly talented and supportive advisor. Your creativity, diligence, and optimism will always be inspirational to me. And thank you for allowing me to be part of the UbiComp Lab. I feel extremely lucky to be a part of this amazing group of students for the last five years. Alanson, thank you for your incredible advice and support during my internships, PhD and professional career. Your passion and contributions as a researcher have motivated me to pursue an industry research career. Shwetak and Alanson, thank you for your instrumental mentorship throughout my PhD. And I am aspired to be an impactful researcher like you guys in my future career.

I am extremely lucky to have spent the last five years of my life with an amazing group of people at the Ubicomp Lab. I will start by thanking the senior folks in the lab. Gabe Cohn and Sidhant Gupta, thank you for your help and guidance over the years. I am honored to follow your path to become a researcher at Microsoft and will continue looking up to you guys. Mayank Goel, you are an incredible researcher, and I was so happy to see you become Prof. Goel, keep rocking! Alex Mariakakis, you are a fantastic friend who I can always count on. Thank you for putting up with many of my random ideas over the years, and I can't wait to see the awesome Prof. Mariakakis version of you next year! Edward Wang, thank you for your friendship. I have always been amazed by your ability to look at the bigger picture in research. I am sure you will be very successful in your academic career. Eric Whitmire you are a great researcher, a fantastic engineer and an even better friend! Lilian de Greef, your talent in HCI and art never cease to amaze me. Elliot Saba and

Ruth Ravichandran, thank you for running the EE signal processing corner, where I learned a lot from chatting with you guys :). Josh Fromm, I admire your versatile technical skills, and it was a great pleasure working with you. Farshid Salemi and Morelle Arian, I am inspired by your passion and I really enjoyed working with you both. I am sure you will continue to be successful in PhD and future career! Tien-jui Lee and Jake Garrison, it was a real pleasure knowing you guys, and I was excited to learn you are staying in Seattle after graduation. Mohit Jain, thank you for many inspiring conversations and I will definitely find you when I visit India. And Chujong Park, Manuja Sharma, you guys are talented first-year students, and I am sure you will achieve a lot in the following years.

My special thanks go to Koji Yatani who motivated me to pursue my PhD in HCI and Ubicomp. Koji, thank you for mentoring me at MSR Asia and introducing me to the field. I am truly grateful. Tobias Grosse-Puppendahl and James Scott, thank you for being my mentors at MSR Cambridge where I learned a great deal. And thank you for supporting my professional career. Ed Lazowska and Greg Gottesman, thank you for introducing me to the entrepreneurship world. Your guidance has genuinely broadened our vision to look beyond just technology and thinking more about how to bridge gaps in the industry by solving practical, real-world problems. Elise DeGoede and Lindsay Michimoto, thank you for being such wonderful grad program advisors who were always there for me!

Finally I would like to thank many other faculty members, students, and industry collaborators who had made my PhD journey truly special including: Gregory Abowd, Ravin Balakrishnan, Danielle Bragg, Eric Brockmeyer, Elizabeth Carter, Safiye Celik, Victor Chan, Tianqi Chen, Shumo Chu, Elise Dorough, Daniel Epstein, Hao Fang, Earlence Fernandes, Anhong Guo, Liang He, Luheng He, Scott Hudson, Yuchen Jin, Ravi Karkar, Yoshi Kohno, James Landay, Anthony LaMarca, Gierad Laput, Vincent Lee, Shu Liang, Xi Lin, Yao Lu, Sandra Matz, Amrita Mazumdar, Lindsay Michimoto, Meredith Morris, Kenneth Morand, Samer Moubayed, Sandrine Mu, Stefanie Mueller, Emerson Park, George Robinson, Joan Sanders, Aditya Sankar, Rahul Shah, Haichen Shen, Faiz Sherman, Joshua Smith, Ha Trinh, Chieh-yih Wan, Yuntao Wang, Chenglong Wang, Ham Wongsuphasawat, Tongshuang Wu, Haijun Xia, Robert Xiao, Eric Xie, Cong Yan, Chouchang Yang, Hao Zhang, Qiao Zhang, Xiaoyi Zhang, Yang Zhang, Yi Zhao, Tianyi Zhou, and Danyang

Zhuo. I would also like to thank various agencies who generously supported my PhD study including Disney Research, Intel Labs, Qualcomm Research and Procter & Gamble.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ADLS	Activities of Daily Living
AI	Artificial Intelligence
ALOHA	Random access network communications protocol
AoA	Angle of Arrival
AR	Augmented Reality
COTS	Commercial Off-The-Shelf
EM	Electromagnetic
EMI	Electromagnetic Interference
EPC Gen2	Electronic Product Code Generation 2
FCC	Federal Communication Commission
FFT	Fast Fourier Transform
HCI	Human Computer Interaction
HF	High frequency radio spectrum
HMM	Hidden Markov Model
IC	Integrated Circuit
IEEE	The Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Unit
IoT	Internet of Things
ISM band	The industrial, scientific and medical radio band
NFC	Near Field Communication
PCC	Pearson Correlation Coefficient
PMU	Phase Measurement Unit

QR code	Quick Response Code
RBF	Radial Basis Function
RF	Radio Frequency
RFID	Radio-Frequency Identification
RG	Research Goal
RSSI	Received Signal Strength Indicator
SVM	Support Vector Machine
SD	Standard Deviation
SAR	Synthetic Aperture Radar
SNR	Signal to Noise Ratio
SDK	Software Development Kit
TC	Thesis Contribution
UHF	Ultra High Frequency Radio Spectrum
USRP	Universal Software Radio Peripheral
VR	Virtual Reality
WSN	Wireless Sensor Network
3D	Three-dimensional

Chapter 1

INTRODUCTION

The last 50 years have witnessed significant advancements in computation. The increased density of electronic circuits has enabled computers to be smaller in form factor yet better in computational power. In 1965, Gordon Moore predicted that the number of transistors in integrated circuits (IC) would double about every year. Later in 1975, he revised his prediction to every two years. This is known as Moore's law, one of the most impactful principles in the computation industry. This exponential growth trend is illustrated in Figure 1.1 by Roser et al. [136].

Moore's Law has allowed computers to evolve dramatically in form and application. A Morgan Stanley study [147] describes the evolution of computing platforms from mainframe machines to mobile devices, growing the number of deployed computers from millions to tens of billions (Figure 1.2). The physical space that computers occupy has also changed drastically, going from a time where these machines needed to sit in a dedicated room owned by large organizations to today's world where portable mobile devices are ubiquitous in people's everyday life. There have been recent debates on whether the exponential growth of Moore's law is sustainable; however, it is still believed that industry will follow the trend of increasing computational power on a smaller footprint in the foreseeable future. This trend has allowed for the natural distribution of computer systems in everyday physical environments, and it is predicted that computers would eventually be embedded into everyday objects, creating truly smart and immersive physical environments. This phenomenon is described as the Internet of Things (IoT).

Recent advances in IoT and the ubiquitous usage of smart devices have dramatically changed the way people interact with their living environment, enabling smart homes, smart cities, wearable electronics, and personal health applications. According to Gartner [39], 8.4 billion devices were added to the Internet of Things by 2017. To connect this large network of devices, industry standards employ wireless communication modules such as WiFi, Bluetooth, and Zigbee. These wireless platforms provide real-time communication between devices and seamless sharing of sen-

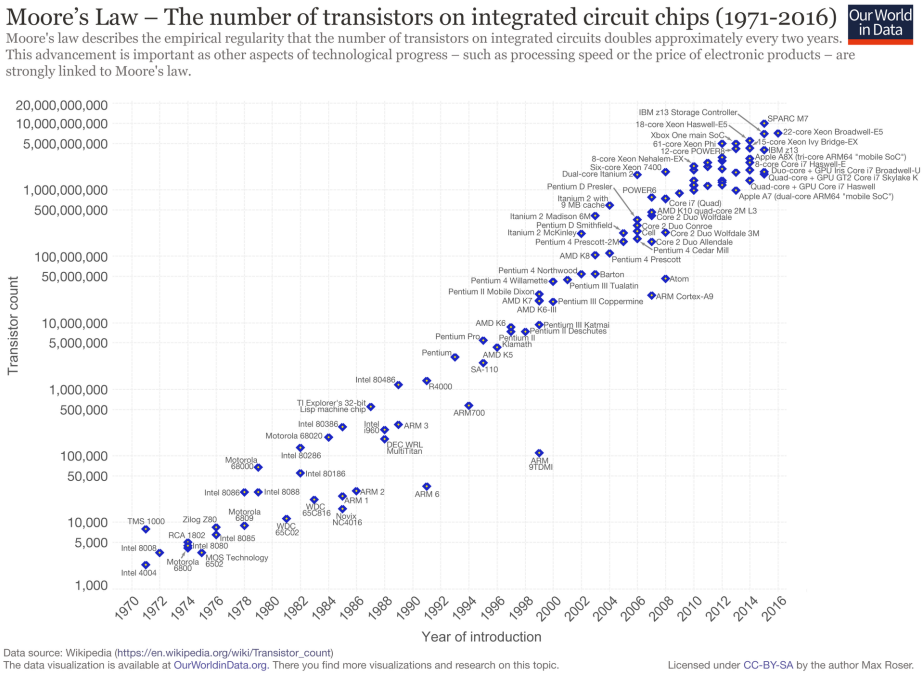


Figure 1.1: The number of transistors on integrated circuits chips [136].

sory data to facilitate end user applications. Given the power consumption of these wireless modules and sensors, IoT devices today require onboard batteries. The unit cost of each sensor, the physical size of these sensors, the bandwidth used by wireless modules, and the requirement of battery replacement all add up to the cost of maintaining this network. For everyday objects, the size of these electronics can be obtrusive, and the cost of maintenance is still too high to be scalable. As a result, there currently exists a big gap between the electronic devices that are now integrated into the IoT framework and the vast majority of non-electronic objects people are interacting with on a daily basis, leaving out a tremendous opportunity for creating compelling user experiences. Being able to integrate sensing and interaction capabilities into these objects will dramatically improve the ubiquity of the current IoT framework and the capacity of creating versatile applications.

Understanding interactions between users and everyday objects could fill this gap in the current IoT framework and enable many compelling sensing applications and novel user experiences. The objects we use define and reflect our daily activities. Leveraging the sequence of object usage information, computer systems could infer and recognize user's daily activities. Imagine a use case

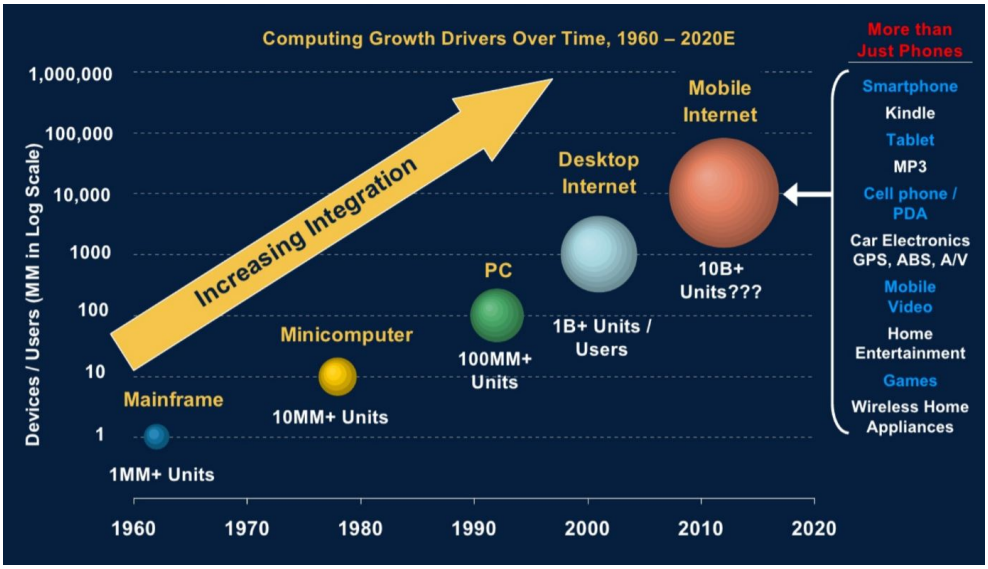


Figure 1.2: Computer growth over time [147].

where a smart speaker provides a person with recipe recommendations according to the groceries they purchased, a coffee mug that is aware of a persons hydration level, or a pill bottle that is aware of a persons medication intake. This ubiquitous sensing capability of everyday objects can enable immersive user experiences across many different application domains. These experiences can be augmented by tangible user interfaces for user input sensing. Today, many electronic devices utilize capacitive touchscreen for user interaction sensing, yet touchscreens are not scalable with everyday objects given their size, cost, and power consumption. By giving everyday objects the ability to sense gesture inputs, traditional interfaces for controlling devices would disappear; any object or surfaces could be repurposed as a control interface. This could also help create links between physical objects and digital media. A pop-up book embedded with tangible interfaces could trigger sounds or animation effects to facilitate interactive and creative storytelling experiences. Going one step further, recognizing the precise identity and physical location of objects and users involved in these interactions could help fuse other platforms such as virtual reality or augmented reality wearables to further blur the barrier between what is physical and what is digital. For example, a VR device could identify and locate a chair and then render it in the virtual world as an iron throne, allowing users to experience natural interactions with the virtual world through objects that

exist in reality. In augmented reality, precise item identification with online media can create a retail experience where a user can feel a piece of clothing in the real world while having access to customer reviews online. Product information videos could also be projected onto the object in real-time to enrich the shopping experience.

These compelling applications outlined above can be enabled by understanding how the user is interacting with everyday objects. More specifically, the sensing capability can be broken into 3 different directions: (1) sensing the state of the objects (e.g., motion, rotation or interference from ambient objects or users) (2) sensing expressive user input (e.g., different hand gestures), and (3) understanding the identity and location of users and objects. Combing these sensing capabilities could enabling new and immersive user experiences in high impact spaces such as health and wellness, remote collaboration, retail, and education. The challenge lies in how these sensing capabilities can be applied in an unobtrusive, scalable, and user-friendly way. In the later sections of this chapter, I further discuss these application domains and propose research goals aiming to address these problem spaces.

Table 1.1 outlines a few commonly used wireless communication platform that are currently deployed in the IoT space. Protocols like UHF RFID and NFC employ completely passive tags that are small, flexible, and do not require a battery. In these protocols, each RF tag has a small memory block to store a unique digital identification value. The low-cost nature of the tags makes them ideal to use on everyday objects in scale. When under operation, each RF tag harvests from ambient signals to back-scatter its identity to the reader. These protocols are designed as replacements for barcode systems. As such, they are often limited to identification purposes in industry. Given the limited sensing range of NFC, in this thesis, I focus on the long-range UHF-RFID system. I leverage channel parameters reported by the reader to enhance the sensing and interaction capabilities of the RFID system through signal processing and machine learning. By attaching UHF tags onto everyday objects, I enable interactive applications that are similar to those achieved by battery-powered wireless sensor network (WSN), such as understanding daily activities or creating tangible user interfaces. My approach allows commercial UHF RFID systems to understand the state of everyday objects, their users, and the interaction between the two parties. In the following sections, I describe how I have tailored these techniques to enable novel sensing in a few different application spaces.

Platform	**WiFi**	**Bluetooth**	**Zigbee**	**UHF RFID**	**NFC**
Range	30-100m	10m	10-20m	10m	0.1m
Power	High	Low-Mid	Low	Zero (tag)	Zero (tag)
Privacy	Low	Mid	Mid	Mid	High
Cost	$10	$10	$10	$0.1 (tag)	$0.3 (tag)

Table 1.1: Wireless communication platform comparison

1.1 Sensing Object Interactions and Understanding Daily Activities

One cornerstone of Ubiquitous Computing is the ability to sense and understand human behavior. An important goal in this space is to enable computing systems to automatically recognize humans daily activities, leading to context-aware applications in smart homes, personal wellness, and assisted living. Ubiquitous sensing capabilities in everyday objects can enable immersive user experiences, driving applications in health, wellness, and user interaction domains. The challenge is to enable these sensing capabilities in an easy and scalable way.

Past approaches to activity recognition include vision-based systems [110, 122], which leverage visual clues to infer daily activities. Video streams provide rich contextual information; however, vision-based systems are limited by their requirement of line-of-sight, significant computation power, and privacy concerns. Researchers have also explored sensing object usage by instrumenting users with a near-field RFID reader glove [120]. The presence or absence states of RFID tags attached to objects in close proximity indicate the activity being conducted. However, this approach requires users to wear battery-powered devices, which raises user burden and limits its feasibility in real-world applications. To make such systems feasible for real-world application, scalability must be improved without increasing user burden. The research goal here is:

RG1: Sensing object interactions and recognizing daily activities in a scalable and privacy-preserving manner without requiring user instrumentation

Towards this goal, I designed and implemented IDSense [84] (Chapter 3), a new human-object

interaction sensing system which uses a commercially available UHF RFID to detect human-object interactions. Differing from traditional approaches that utilize RFID tags' binary presence states, I leverage RF channel parameters to detect the state changes of individually tagged objects. The object usage indicated by these state changes contains rich contextual information about the users interaction with the surrounding environment, which allows the system to understand fine-grain daily activities. With IDSense, physical interactions can be linked with digital media to enable compelling user experiences. In addition, IDSenses activity recognition capabilities can be combined with assistive technology to improve the life quality in assisted living scenarios [23]. Compared to existing wearable solutions, IDSense does not require user instrumentation, which significantly improves the user experience;in turn, this helps with technology acceptance and compliance issues reported in previous literature [74]. Unlike previous approaches that require users to wear active RFID readers or customized RFID tags, IDSense is capable of sensing user object interactions and recognizing fine-granular daily activities using an off-the-shelf RFID system without user instrumentation and minimal instrumentation to the objects.

1.2 Tangible User Interfaces for Everyday Objects

Today, smart devices are designed to accommodate touch-sensitive input methods that can enable interactive user experiences. Everyday objects, on the other hand, do not have embedded electronics and thus can not support such interactivity. Bringing seamless interactivity to everyday objects can obviate the need for traditional user control interfaces (e.g., switches, remotes), allowing any object or surface to be repurposed as a control interface. Taking this one step further, wireless tangible user interfaces would support connections between physical objects and digital media, creating a truly immersive IoT user experience. The challenge here is how to lower the barrier of creating wireless, battery-free, input interfaces for everyday objects.

Traditionally, physical input interfaces (e.g., buttons, sliders, knobs) are limited by physical constraints like the size of electronic circuits, wiring requirements, and the need to supply or replenish power. These limitations make it challenging to embed physical interfaces into everyday objects, especially objects that are lightweight and flexible (e.g., a piece of paper). The research goal here is:

RG2: Creating tangible and customizable user input interfaces for everyday objects while maintaining their passive and flexible nature.

To tackle this challenge, I proposed PaperID [83] (Chapter 4), a technique that allows off-the-shelf RFID tags to be turned into a variety of simple mechanisms to fabricate passive, thin, and flexible input devices. PaperID customizes RFID tags with hand-drawn conductive ink traces to sense a variety of input gestures on and around tags, which enables seven different interaction methods (e.g., buttons, sliders, knobs). Due to the low cost, wireless, and lightweight nature of the technique, PaperID is especially suitable for rapidly prototyping input interface for everyday objects.

I demonstrated and evaluated PaperID in applications like a customized remote controller for IoT devices and interactive paper crafted objects, such as pop-up books. PaperID expands the functionality of RFID tags to quickly and easily create interactive interfaces on paper and other mediums. The application space for PaperID is multifold. In this thesis, I implemented and evaluated applications such as a real-time polling questionnaire embedded in an exam sheet and a gesture-based controller for lighting, as well as several examples of interactive toys that are connected to digital media.

1.3 Localization and Identification for Objects and Individuals

The identification and localization of objects and their users will allow computer systems to facilitate multi-user sensing and interactive applications. Imagine a retail scenario where the location and identity of each object can be precisely measured. Online media could be seamlessly merged with physical environments, allowing users to simultaneously look up a piece of clothing in the real world, feel the fabric, access online customer reviews, and watch product introduction videos. All of this information can be projected onto the object in real-time using augmented reality to enrich shopping experiences.

However, it is challenging for existing technology to simultaneously identify and locate objects or individuals with enough speed to enable seamless tracking for real-time applications. Recognition systems based on computer vision can robustly determine where objects and individuals are in their

field-of-view [25], but it is still challenging to identify what the objects are and who are the individuals interacting with the objects. State-of-the-art deep learning vision systems have shown promising results when given large amounts of still photos for training [43, 153]. However, extensive training is not feasible in many real-world scenarios. The uncertainty of illumination, orientation, and occlusions also add to the challenges of vision-based approach. Alternatively, recognition approaches based on wireless signals usually require the active participation of the user by carrying a wireless transponder [120]. However, it is still very difficult for such systems to determine precisely where the person is due to the multi-path nature of propagating electromagnetic waves. The research goal here is:

RG3: Enabling accurate identification and localization of individuals and objects.

Towards this goal, I developed two hybrid systems merging camera-based computer vision and RF-based wireless communication. In particular, IDMatch [85] (Chapter 5) uses a novel reverse synthetic aperture technique to recover the relative motion paths of RFID tags stuck on objects or worn by individuals. The motion paths are correlated to physical motion paths measured with a 3D depth camera to achieve accurate user identification. IDCam (Chapter 6) leverages the correlation between the motion of user's hand and objects to identify and enhance user interactions under an augmented reality scenario.

I demonstrated the capability of the system through a robotic host Q&A application. The ability of simultaneous recognition and localization of individuals allows the robot to greet users by their names and to make eye contact during conversations to support natural human-robot interactions. In addition, precise identification and localization of objects and individuals can combine objects in the real world with virtual content generated by AR and VR technologies. For example, a VR device can identify and locate a chair and then render it in the virtual world as an iron throne. One can also overlay virtual content on top of real objects using AR technology. In IDCam, the system utilize the part number stored in a products RFID tag to look up its price and reviews online; that data is shown in an AR headset overlay that follows with the product as the user moves their head.

1.4 Dissertation Goals and Contributions

The research goal for this dissertation is to support novel sensing and interaction techniques with everyday objects using commercially available RFID systems. I propose techniques to enhance the sensing capabilities of commercial RFID by mining contextual information from its communication channel using signal processing and machine learning. In addition, I combine RFID with other sensing modalities to serve the purposes of supporting natural human-robot interactions, enhancing AR experience, and creating tangible interfaces. Finally, I evaluate a variety of novel IoT applications enabled with the above proposed techniques. Throughout this dissertation, I provide support for my thesis statement:

Thesis Statement: Applying signal processing and machine learning techniques to low-level channel parameters of commercially available RFID tags can enable novel sensing and interaction applications with everyday objects.

Here I summarize the hypotheses and artifact thesis contributions (TC) that I will experimentally prove and deliver, respectively, in support of my thesis statement:

TC1: Signal processing and machine learning techniques leveraging RFID channel parameters to detect user interactions with object enhanced with RFID tags

TC2: Demonstrate that daily activities with different level of details can be recognized by sensing the status of RFID tags attached to everyday objects.

TC3: Create wireless, passive, and customizable input devices combining RFID tags with conductive ink printing to enable sensing of user inputs on everyday objects

TC4: Signal processing and machine learning techniques for fusing computer vision with RFID to enable simultaneous identification and localization for objects and individuals.

TC5: Benchmark the sensor fusion approach proposed through enhanced human-robot interaction and augmented reality retail applications.

1.5 Thesis Organization

Research Goals	Contributions	Addressed In
RG1: Unobtrusive and scalable way of sensing object interactions and daily activities	**TC1, TC2**	**Chapter 3:** Through the implementation and evaluation of *IDSense*: A human object interaction sensing system based on commercial UHF RFID.
RG2: Passive and customizable user input interface for everyday objects	**TC3**	**Chapter 4:** Through the implementation and evaluation of *PaperID*: A technique combining conductive inkjet printing with RFID chip to create customizable touch sensitive interfaces
RG3: Achieve accurate localization and recognition of individuals and objects	**TC4, TC5**	**Chapter 5 & Chapter 6:** Through the implementation and evaluation of *IDMatch* and *IDCam*: Techniques for fusing computer vision with RFID to enhance human robot interactions and object manipulations in augmented reality.

Table 1.2: Dissertation organization

Chapter 2

RELATED WORK

This thesis presents a new class of sensing techniques leveraging low-level channel parameters of commercial UHF RFID system, addressing three research goals: (1) unobtrusive ways of sensing people interactions with objects and inferring daily activities, (2) creating passive, thin and flexible tangible user interfaces, and (3) achieving accurate localization and recognition of individuals and everyday objects. This chapter is structured around these research goals, providing an in-depth survey of existing technologies within each of these application domains to better position my thesis work. I survey prior work related to each research goal with particular focus on wearable and distributed sensing systems. More specifically, I first give an overview of previous work addressing activity recognition, tangible user interfaces, and user/object localization via a variety of on-body or distributed sensing mechanisms. Then, I outline prior RFID sensing techniques and discuss how my thesis work is different when compared to these earlier approaches.

2.1 Understanding Daily Activities

Automatic means of activity recognition are one of the key building blocks needed to enable context-aware ubiquitous computing applications. There is a variety of prior work that aims to accomplish this goal using user-centered sensing and distributed sensing, leveraging active and passive sensors each with their unique strengths and weaknesses.

2.1.1 Activity Recognition via Wearable Devices

Wearable systems require users to wear devices that continuously monitor their interactions with their surroundings. Previous work has demonstrated promising results when tracking the state of the body by attaching IMUs onto the user [11, 48, 97, 111] (i.e., running, walking, sitting), but encounters difficulties when recognizing more complex activities (i.e., cooking, reading a book). The objects we utilize in daily living can provide a rich amount of contextual information given the

correlation between the objects used and the activities performed. To gain a better understanding of daily activities, Ren et al. [130] applied computer vision techniques towards body-worn cameras to capture hand-held objects. They achieved an accuracy of 64% with a clean background scheme; however, challenges like background clutter and hand occlusion dramatically reduced the accuracy. In recent years, deep learning techniques have demonstrated promising results in understanding user's daily activities using sensors carried by the user, such the IMU on the mobile phone [184, 91, 182]. Castro et al. [20] presented a deep learning activity recognition system based on egocentric cameras. This work achieved an accuracy of 83% classifying 19 daily activities, but lacked support for recognizing fine-grained activities (e.g., which dish is being prepared). In addition, privacy concerns of first-person visual recordings limit the usage of computer vision systems in living environments. Prior work has also explored identifying objects utilized in daily living by recognizing their unique electromagnetic interference (EMI) pattern [80, 164]. The approach by Laput et al. [80] used a watch-based software defined radio. They achieved 97.9% accuracy across 24 devices. However, the vast majority of objects in daily living do not generate EM noise, limiting the reach of such approach. The audio signal from wearable devices has also been explored for activity recognition. In particular, Yatani et al. [180] presented a wearable acoustic sensor that could listen for sounds produced in the user's throat area to classify user activities. In general, user-wearable solutions have the advantage of being mobile, extending the coverage of activity sensing compared to distributed solutions. However, the increased user burden due to body-mounted devices limits the technology acceptance and compliance for applications in living environments [74].

2.1.2 Activity Recognition via Distributed Sensors in Physical Environments

Researchers have also explored distributing sensors in different environments for activity recognition For example, binary state-change sensors, like piezoelectric switches, have been used to detect object usage [160, 154]. Each object is assigned a wireless sensor that is triggered when the object is moved. However, the form factor and unit cost is still not ideal for adoption on everyday objects. Infrastructure mediated sensing is a novel approach that leverages the existing home infrastructure to mediate the transaction of events and to understand the human activity [116]. Prior work leveraging this principle includes ElectriSense [47], which measures the electromagnetic interference (EMI)

generated by electric appliances to classify and detect appliance usage. The propagation of EMI signals along the powerline allows this system to detect multiple appliances from a single location, similar to the wearable solution by Laput et al., it is challenging to generalize EMI-based sensing to non-electronic devices. SNUPI [24] is a ultra-low-power sensing platform. SNUPI leverages powerline infrastructure for communication, which allows a single battery cell to operate for years for sensing applications. This work demonstrated great research and commercial impact through real-world applications and products. However, it is still challenging to scale these sensing applications towards everyday objects.

Computer vision is also a popular technique in distributed sensing. Researchers have explored leveraging video surveillance cameras and computer vision techniques to understand human activities [109, 88, 161]. To further improve the sensing accuracy, several techniques based on depth camera has been proposed and studied [110, 175]. Audio signal has also been proven effective for activity recognition. In particular, microphones has been deployed in living environments [51, 22] or onto robots [150] for sensing human activities in the ambient environments. Several prior work has also demonstrated using wireless signal reflection (e.g., WiFi backscattering) to enable activity-related sensing human motion detection [174, 2], body gestures [123, 77], or physiological processes like breathing [128, 1] or heart beats [185, 3].

A few research spaces are being designed and instrumented as home environments to facilitate studies of activity recognition, including the PlaceLab [67] in the MIT house_n [103] and the Georgia Tech Aware Home [78]. Researchers have investigated activity recognition tasks in these spaces using a variety of sensing modalities. In particular, several researchers have leveraged sensory data collected in the PlaceLab for daily activity recognition tasks [50, 90, 53]. These sensors include object state sensors, location beacons, microphones, cameras, current and water flow sensors, humidity and light sensors, object and person motion detectors, etc. Although combining multiple sensing modalities can improve activity recognition accuracy, it also increases the burden of deployment and maintenance, creating barriers for real-world adoption. In general, techniques based on sensor networks demonstrate promising results in activity recognition but also face challenges in form factor, high unit cost, and the requirement for battery replacement.

2.2 Input Devices for everyday objects

Electronic IoT devices usually come with dedicated physical interfaces. For example, smartphones and tablets come with the capacitive touchscreen to accept user inputs. Having touch-sensitive interfaces is crucial towards enabling interactive user experiences. However, the form factor and battery required for a touchscreen is not scalable with everyday objects. Consider one common object: paper. Paper is extremely thin and flexible, and yet its low-costnature makes it disposable. Enabling touch sensitivity on paper would greatly enrich user's experience with this common media. Previous research has discussed applications that could be enabled by paper-based input interfaces, including cross-media hyperlinking, documentation tagging, and interactions techniques connecting written text with digital media [6, 93, 46, 60, 148]. However, it is very challenging to enable unobtrusive input sensing while still preserving the lightweight, passive, and disposable nature of the paper.

Paper-based interfaces have been explored previously as a method for rapid prototyping. Qi and colleagues [125] introduced a paper circuitry approach where a user could build control interfaces on a sheet of paper, attaching electronic components with conductive ink and copper tape. Kawahara and colleagues [76] extended this work by creating inkjet-printed circuits on photo paper for rapid prototyping. They demonstrated easy and effective ways for printing connection circuits or even sensors to enable interactive scenarios. However, the utility of these paper-based interfaces is limited to on-paper devices, which in turn are constrained by the difficulty of augmenting paper with wireless communication components. Attaching wireless modules and power sources for them makes it difficult to retain the flexible and lightweight nature of paper. Additionally, using active components counteracts the benefit of using inexpensive paper as a medium.

There are other ways to integrate user interfaces onto everyday objects. For example, using computer vision-based sensing on a hand-held or wearable projector has been heavily studied for interface projection and user input sensing [104, 102, 12, 19, 171]. Even though such techniques allow user interfaces to be projected onto everyday objects, computer vision based input sensing is still limited by inconsistent environmental factors (e.g., lighting conditions, object orientation, and occlusion from the human body). Additionally, having users wear or carry a battery-powered projector system adds to user burden.

2.3 Simultaneous Localization and Identification of Users and Objects

The ability for computing systems to quickly recognize and localize individuals and objects is the key component for understanding the context of interactive applications and crucial towards creating immersive user experiences. However, this becomes especially challenging in dynamic environments that require fast, implicit, non-intrusive, and ubiquitous identification and localization.

User recognition systems based on computer vision (including stereo cameras, structured light, and LiDar) can robustly determine where people are in their field of view, but not who those individual people are. State-of-the-art face recognition systems have shown promising results when given large amounts of still photographs for training [153, 43, 113, 170, 126, 31]. Gathering those photos requires a cumbersome user registration process consisting of photographing and manually annotating each participant for training purposes, which is not feasible when scaling to a large number of participants or casual, one-time encounters. Face recognition is also limited by constraints that normally plague vision-based sensing, such as orientation, lighting conditions, and image resolution.

Advancements in computer vision have also brought real-time object recognition into reality. Large-scale open-source image databases like ImageNet [27] provide a variety of labeled images with which researchers can train and test their systems. State-of-the-art deep learning models have demonstrated promising results when given large amounts of pictures for training [55, 167, 151, 129, 186, 143]. However, the granularity of object labels in a dataset like ImageNet is categorical. ImageNet may contain labels for multiple brands of jeans, but the labels for all of them is the same. In a retail setting, for example, identifying an object by its brand and size is critical. Knowing the precise identity of an object is still very challenging for computer vision given the visual resemblance of similar products (e.g., the same jeans in slightly different sizes). In addition, machine learning algorithms require many labeled pictures taken from different angles and settings, incurring a substantial deployment cost. Since models must be retrained as new types of objects are added to the dataset, this approach does not scale well. Furthermore, certain contextual information is not embedded in the visual clue (e.g., the owner of a coffee mug), making it extremely challenging for computer vision to understand this information and create a unique identity for each object.

Alternatively, objects and users can be easily recognized if they are instrumented with visual identifiers. Prior work has also looked at retrieving item information through barcodes [118, 75].

However, this interaction is not as seamless and natural as it may sound. Taking a picture of a barcode requires that the barcode is easily accessible to the customer and that the customer can properly frame the code in their smartphone's camera. QR codes have been widely utilized for object identification in augmented reality or accessibility applications [73, 13]; scanning a QR code requires similar effort by the user.

Alternatively, recognition approaches based on wireless devices require active participation by the user, but significantly increase the reliability of identification. They typically require the user to activate an app on their cell phone for each interaction or to carry a wireless transponder [135, 61]. These devices can actively transmit data to identify the user or object involved in an interaction. In this case, it is challenging to determine the precise location of objects and users. This is primarily due to the nature of propagating electromagnetic waves, which makes it difficult to precisely locate a transmitter with an accuracy greater than 1-2 meter [89, 176, 106].

2.4 RFID Related Sensing Techniques for Interaction Detection

This section focuses on the body of prior RFID work similar to techniques presented in this thesis. I will break down this body of work into a few specific application domains including activity recognition, touch interfaces, and localization. I will also discuss opportunities of combining RFID with other sensing modalities to enhance RFIDs sensing capabilities. For each application domain, I will further discuss how my approach has advanced upon the prior work to improve scalability, usability, and sensing accuracy..

2.4.1 Activity Recognition using RFID

In the field of activity recognition, RFID systems have the advantage of being passive and low-cost when compared to other distributed sensing solutions. Philipose et al. developed a wearable near-field RFID reader that could identify tags attached to objects nearby [119, 37]. Objects held in hands are recognized by the wrist-worn reader; over time, the usage of these objects can be utilized to infer higher level activities. This work combined user-centric sensing with distributed RFID tags on objects and demonstrated that the feasibility of identifying daily activities solely from tracking object usage events. They achieved a precision of 88% and a recall of 73% classifying 14 activities

of daily living (ADLs [72]). In a separate study [117] they demonstrated the capability to classify 11 daily activities with 88% accuracy. However, like all other user-centric sensing systems, Philipose et al.'s system requires the user to wear an active sensing device continuously.

The recent advances in commercially available passive long-range RFID systems give rise to their popularity in distributed sensing applications. These systems operate at 900Mhz ultra-high frequency (UHF) ISM band. The long range and low-cost nature of such RFID technology make it scalable to instrument everyday objects. The RF channel parameters reported in the Gen 2 UHF RFID readers [33, 63] provides insights into the state of the tag and its surrounding environments. This system is previously studied for health and wellness sensing applications [162, 15, 68, 159]. Researchers have also studied the use of passive UHF RFID for activity recognition in trauma resuscitation scenarios [86, 114, 115]. However, the object usage classifiers described in their systems are based solely on reflected signal strength (RSSI). In addition to requiring multiple antennas, these previous work require excessive training for each object, making it infeasible to scale to a large number of objects.

Researchers have also explored modifying circuits of the commercial RFID system for enhancing its sensing capabilities. In particular, Smith et al. presented methods of adding mercury switches to tags in order to enable binary acceleration detection [121, 145]. They also leveraged this method to detect object usage and daily activities [144]. Prior work has also presented custom sensing platforms based on RFID communication, such as the UHF WISP[140, 146, 139] or the UMass Moo [183]. These sensing systems provide custom-built RFID tags with an embedded microcontroller to collect data from onboard sensors. Although expensive due to their special-purpose nature ($100 per unit), these tags are capable of many sensing tasks. For example, prior work describes their use for tracking tag movement within a large room[17]. Buettner et al were able to distinguish between 12 everyday household activities leveraging the tags onboard accelerometer.

In this thesis, I leverage various channel parameters reported by RFID readers to improve the sensitivity of object interaction detection while minimizing the antenna requirements. In Chapter 3, I present a system named IDSense [84] to enable scalable recognition of human object interaction and daily activities. Different from prior work, IDSense is capable of robustly classifying states of tagged objects using a single RFID reader and antenna. In addition, IDSense combines signal processing with machine learning to adapt to different environments without retraining. I also

demonstrate other applications that could be enabled via IDSense such as interactive storytelling with toys and understanding of customer interest in products in a retail setting.

2.4.2 RFID based Input Devices

Modern tools and techniques now make it quite easy to construct a graphical interface with only minimal training and effort. Advances in low-cost, micro-controller-based electronics (e.g., [7]) and accompanying tools such as hardware toolkits (e.g., [82, 42, 52]) have sought to bring that same ease of creation to physical interactive devices as well. However, because of physical constraints, such as the size of electronic circuits, required wiring, and the need to supply and/or replenish power to those circuits, creation of physical interfaces has remained difficult. Thus, construction of such devices still cannot really be considered easy, lightweight, or simple. As a result, we might normally start any design effort aimed at creating such an artifact with an exercise in paper prototyping [131]. We do this in part because the paper is *lightweight* – both in a physical and economic sense, but also with respect to the ease with which we can work with it and modify it on demand.

Enabling touch and gesture sensing on tags can make them an ideal alternative for use in control interfaces given their thin, flexible, passive nature. Early work on the use of RFID tags for input constructed power-free buttons from modified RFID tags [9]. Other sensing approaches that use RFID technology as an interface have required a large number of tags or reader antennas. For example, Asadzadeh and colleagues [8] implemented an 80cm by 80cm matrix with 64 tags and three antennas to detect hand gestures. Their system was able to classify 12 predefined hand gestures with 93% accuracy. Their system was able to trace tag position with a median error of 3.7cm. Marquardt and colleagues [96] explored the implementation of input buttons and sliders by modifying RFID circuits and adding electrical components around RFID chips. Together, these projects demonstrated the feasibility of RFID-based input sensing. However, constraints in system complexity or user burden make them undesirable for augmenting everyday objects in realistic scenarios.

Fine-grain motion detection can also be utilized as a mechanism for user input sensing. Fishkin *et al.* [36] were able to demonstrate the use of inexpensive, passive UHF RFID tags as a sensor for detecting motion. However, motions were detected at the cost of multiple tags per object and multiple readers. Motion-related gesture sensing is utilized for pairing objects instrumented with

RFID tags [79]. Even though binary motion detection is sufficient for object usage detection, it is not sufficient for sensing user input. The granularity of motion detection needs to be improved to support expressive gestures. More recently, Wang and colleagues [166] proposed a system for tracking tag movement that used 8 antennas in 4 pairs. Even though they were able to track motion traces with an error of a few centimeters, their system relied on heavy hardware installation, limiting its scalability in real-world scenarios. These previous research projects demonstrate the feasibility of using UHF RFID systems to sense user inputs. However, there are either only support binary motion detection or require extensive hardware, limiting its expressiveness or scalability for input sensing.

Techniques for printing RFID antenna on paper has been studied as a way for customizing RFID tag design [179, 133]. In Chapter 4, I present the PaperID system [83], which extends prior work by creating new tag types that can be quickly prototyped using conductive inks and an off-the-shelf RFID IC. Specifically, I introduce the 'half antenna tag' design, a passive RFID tag that can use the human body as part of the antenna structure for operation when touched. Additionally, a number of new input modalities are enabled that include 7 types of on-tag and free-air gesture interaction types (e.g., hand waving around the tag) as well as sensing the real-time motion velocity and trajectory of the tag relative to the reader. I also demonstrate how these input modalities can be customized to create passive, lightweight tangible input interfaces for IoT applications.

The PaperID project presented in this thesis was completed in 2016. More recently, a few interesting research projects are introduced to further improve the expressiveness of RFID based gesture sensing. In 2017, Zou [187] and colleagues presented a method for sensing 6 different mid-air hand gestures. In 2018, Jin and colleagues introduced a system for body frame tracking [70], which can be utilized for body gesture sensing. These new works further expanded the expressiveness of RFID gesture input sensing, yet still require multiple tags or antennas to achieve these functionalities. Future work in lowering the hardware setup will be necessary to improve the scalability of these systems.

2.4.3 RFID Localization

Given that RFID is capable of sensing the identity of tags, we can enable simultaneous identification and localization of tagged objects and users if tag localization can be achieved. Commercial solutions can achieve coarse-grained RFID localization using phased array antennas. For example, Impinj Xarray [66] can achieve a localization error of approximately 1.5m with 85% confidence. This level of accuracy is not sufficient for detecting object interactions.

Synthetic aperture radar (SAR) and angle-of-arrival (AoA) techniques are regularly used in wireless communication systems to locate active transmitters [176, 106, 58, 89, 26]. Many of these general approaches have been adapted to the UHF RFID space, where RF channel parameters such as received signal strength (RSSI) and RF phase. Using these features has led to localization accuracies on the order of 10-100 centimeters [108, 141, 57, 173, 163, 10] for well-controlled environments such as the anechoic chamber. When deployed in real-world scenarios, tag localization will be less accurate and insufficient for differentiating between neighboring users or objects.

To increase localization resolution of tags in unconstrained and multipath environments, several systems use synthetic aperture antennas on the RFID reader. However, these approaches face limitations when both the reader and the tag are moving. For example, Miesen et al. [101, 100], used an antenna on a linear actuator to create a synthetic aperture. However, this requires that the environment remains static while scanning occurs. Wang et al. [165] uses a spinning antenna to create a synthetic aperture but need densely spaced marker tags placed throughout the environment to disambiguate the mobile tag motion from reader antenna motion. Yang et al. [177] demonstrated the use of multiple RFID reader antennas that can locate various moving tags in harsh multipath environments. In their system, tags can only be located while traveling at a constant velocity along a known trajectory (i.e., on a conveyor belt).

A variety of mobile robotic systems have used RFID tags for navigation [45, 112] and to localize tagged objects [49, 28]. Although these robots have computer visions systems, they are typically used for object manipulation instead of real-time user tracking purposes. One notable exception is Grema et al. [152], who built an eight-element phased array into a mobile robotic platform to determine the AoA of people wearing RFID tags.

In general, prior work in RFID based localization has a high requirement for hardware installa-

tion (e.g., multiple antennae in the physical environment or multiple tags on each object), requires extensive calibration or fingerprinting for a specific physical environment, or imposes constraints on the possible state of tags (e.g. moving in a linear trace). Also, the accuracy that has been achieved in prior work (10s-100s cm) is not sufficient for differentiating between small objects that are in close proximity. Given these constraints, researchers have investigated combining RFID other sensing modalities to further improve its localization accuracy.

2.4.4 Sensor Fusion Approach towards Identification and Localization

Sensor fusion has been explored in a number of different sensor combinations for identification and localization purposes. Fusion between the motion sensor and computer vision has been examined by the CrossMotion project [172], where visible users can be correlated with their smartphone's acceleration data as they walk through a room. Such a system is feasible for identifying people since they carry their smartphones with them at all times, but it is not practical to instrument everyday objects with accelerometers.

Two hybrid computer vision and RFID systems have been reported with the goal of identifying and locating tagged people. Prior work by Cafaro et al. [18] demonstrates the ability to determine which of two people are standing on the left and right of an interactive display. The second example by Goller et al. [41] shows an occupancy counting scenario where the authors can simultaneously determine the direction of travel for two people. Both of these systems use RSSI fingerprinting, which requires extensive training and multiple reader antennas. As is the case with all fingerprinting based location schemes, changes in the RF environment will result in accuracy degradation and require the system to be retrained. Additionally, these systems are limited to tracking only two participants.

In Chapter 5, I present a hybrid system named IDMatch [85] that combines commercial UHF RFID with computer vision for accurate user localization. Different from prior hybrid approaches that have required complex antenna setups and RSSI fingerprinting procedures that are not scalable towards new environments, this work is based on trajectory correlation. Trajectory correlation, or the matching of motion traces across two different sensing domains, allows for fast alignment of human and tag motion without requiring retraining for new physical spaces. At the core of the

IDMatch system is a method for determining the fine grain change in distance between the reader and the tag using a reverse synthetic aperture radar (SAR) approach. These RFID SAR paths are then compared to the position paths of the people in the space as reported by a 3D depth camera. This core capability is augmented with a machine learning approach that classifies changes in low-level RFID communication channel parameters to the motion of a person in space.

Researchers have also explored the combination of computer vision and RFID for item tracking [40] and localization. [107, 138]. In those studies, the computer vision and RFID sensing systems are either collocated or installed at relatively fixed locations. In augmented reality applications, item interactions are best viewed from the user's perspective, whether through a smartphone or smart-glasses. In Chapter 6, I present that the two sensing systems, RFID and computer vision, can be decoupled for a mobile augmented reality object interaction scenario and still achieve high accuracy in item identification and localization. Similar to the techniques presented in the human-robot interaction work, I employ velocity correlation for fusing the two systems yet here the RFID and vision systems are decoupled where the RFID system is fixed onto the ambient environment while vision system is moving as the user conduct object interactions. To overcome this challenge, I implemented a dynamic coordinate transformation technique that can compare velocities measured from the two systems for correlating objects with identities stored in the RFID IC. In addition, I demonstrate that this approach works for short motion traces (on the order of tens of centimeters), which provide us with an unique opportunity for enhancing AR applications which are detailed in Chapter 6.

Chapter 3

UNDERSTANDING HUMAN OBJECT INTERACTIONS AND DAILY ACTIVITIES

3.1 Introduction

Effective means of identifying peoples activities in their indoor environments has the potential to enable a wide number of human computer interaction applications [59, 168, 169]. One key observation is that the objects we interact with provide rich contextual information about the state of our environment and the activities that we are doing. Whether its reading a book to a child, cooking a meal, or fixing a bicycle the objects that we use both define and reflect the activities we do in our daily lives. The challenge is to create an unobtrusive and general purpose approach to monitoring human object interaction via computer systems. A variety of sensing approaches have been proposed and shown that activity recognition is possible based on object interaction [114, 119, 154]. One common approach is to instrument objects with wireless sensor nodes using accelerometers (or other sensors) to infer object interactions. This approach can provide high fidelity streaming sensor data but due to their relatively high per unit cost, large size, and need for battery replacement these methods have found limited usage for object-based activity monitoring. In this Chapter I propose IDSense [84], a new human object interaction detection technique which uses commercially available passive Ultra High Frequency (UHF) RFID tags and readers to detect human object interactions in the form of motion and touch. Combined with the ID information inherently provided by the RFID tags, the approach enables interaction identification for a wide variety of daily objects. This is accomplished by observing changes in the physical layer signals of the communication channel between the RFID reader and the passive tags. The key insight is that the channel parameters reported by the RFID reader, such as Received Signal Strength Indicator (RSSI), RF Phase, and Doppler shift represent a unique signature of the RF environment of each individual tag. By observing changes in these parameters over time, inferences can be made about the state of the tag and thus the object the tag is attached to. IDSense is a scalable, real-time object interaction detections system that can robustly

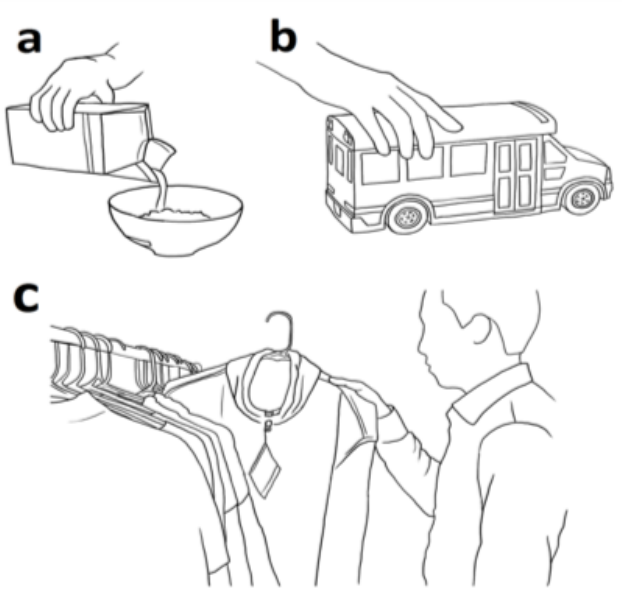


Figure 3.1: IDSense applications include: (a) activity inferencing of daily tasks (b) interactive story telling with low cost toys (c) identification of customer browsing habits

monitor large rooms and living spaces with a single RFID reader. Everyday objects can easily be retrofitted with RFID stickers or integrated with RFID tags by the manufacture. Users are able to naturally interact with 10s to 100s of tagged objects and the system detects events such as object motion and tag touch. By observing these interaction events over time, it is possible to enable a wide variety of applications such as inferencing of daily activities in the home (Figure 3.1a), interactive story telling using low cost tangible toys with computer based media (Figure 3.1b), enhanced retail experiences where interactions with tagged merchandise can be used to determine customer interests (Figure 3.1c). Generally speaking, IDSense makes sensing human interaction with everyday objects easy and unobtrusive, by minimally augmenting objects with low cost and long-lived RFID tags. I develop a new human object interaction detection technique based on commercially available, long range RFID technology. This system is capable of robustly classifying tagged objects using a single RFID reader and antenna in home and office environments.

3.2 System Overview and Implementation

The goal of this project is to use minimalistic hardware in the form of commercially available RFID tags to provide enough sensing capability to robustly detect basic human object interaction. Figure 3.2 shows an example of a toy ambulance augmented with a UHF RFID tag on its hood; the copper antenna is most visible in panel C. Good RF engineering practices should be observed when choosing the type of tag and its placement on the object. There are a wide variety of tag shapes and sizes to choose from, as well as tags specially designed for glass and metal object. Generally speaking it was not difficult to find good tag locations but some trial and error can be expected. Three different object states are investigated as well as two different touch events. The primary state is object still, meaning no interaction with object. The second state is object translation as shown in Figure 3.2a. Where an object translation is defined as movement of greater than 10cm within 2 seconds. The third state is rotation as depicted in Figure 3.2b, which consists of a 90 degree rotation around one of the objects axis. Swipe touch (shown in Figure 3.2c) consist of the user swiping their finger across the tag antenna within 2 seconds. Finally a cover touch is when a user touches more than half the tag antenna for a minimum of 1 second. It should be noted that small and/or extremely short duration movements were not specifically studied and only natural human interactions are explored.

This work focuses on common indoor environments such as the home and office with no special consideration given to building materials, room selection, or furniture placement. Figure 3.3 shows an image of the lab environment where initial testing and validation of the IDSense system was done. The RFID reader antenna is placed on top of a ceiling panel (highlighted in pink) pointed downwards. A coax cable leading back to the Impinj Speedway Revolution reader and host computer is visible. The read distance of UHF RFID tags in free space is +10 meters. In this example the reader coverage zone extends from the gray work bench on the left to the gray work bench on the right and includes the wooden table and floor in the middle of the frame. In this region per tag read rate (or sampling rate) is constantly between 15-40 reads per second depending on the size of the tag population. As a point of reference, it is possible to read tags on the far white shelves in the background but read rate is typically below 10 reads per second. As will be discussed later in this Chapter, proper RFID reader antenna placement is important to achieve good performance and

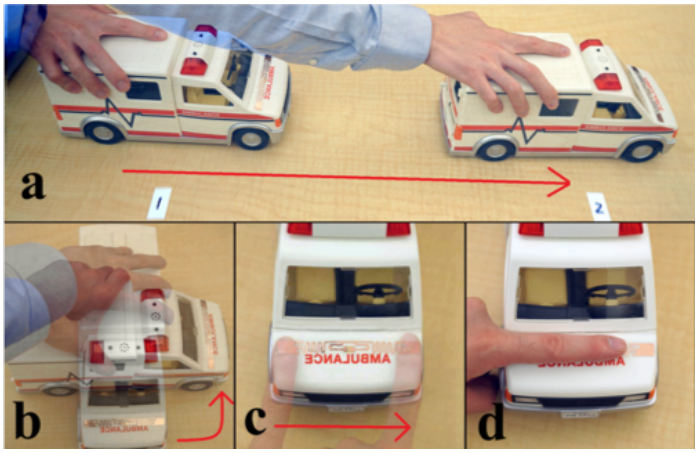


Figure 3.2: IDSense: types of human object interactions a) translation b) rotation c) swipe touch d) cover touch

reduce false detection caused by human activity, and a detailed discussion of room coverage can be found in [30].

3.2.1 Physical Layer Signals of UHF RFID Systems

The commercially available UHF RFID system used in this work is capable of reporting low-level channel parameters such as Received Signal Strength Indicator (RSSI), RF Phase, and Doppler shift as well as the unique identification number of each tag. The RFID reader interrogates tags within its range according to the ISO-18000-6C specification, based on the Slotted Aloha protocol, and has a maximum theoretical read rate of 1,200 tags/sec. Given proper reader settings, it has been observed that an individual tag can be read at 90 reads/second and a population of 10 tags can be read at 330 reads/sec with individuals reading at 30 reads/sec. A large population will begin to saturate the system. For instance, we observed that 60 tags had a total read rate of 616 reads/sec and thus the individual rate dropped to around 10 reads/sec. To overcome this limitation it is possible to programmatically mask sub-populations of tags to increase read rate. This technique was not needed for this work and as a rule of thumb the typical per tag read rate was between 15-40 reads per second for all experiments.

In RFID systems, RSSI is a measurement of the signal power received at the reader and is

Figure 3.3: Lab environment used for initial IDSense testing. The RFID reader antenna is placed on top of a ceiling panel (highlighted in pink) pointing downwards.

predominantly affected by large changes in the distance between the tag and the reader. RF phase is a measure of the phase angle between the RF carrier transmitted by the reader and the return signal from the tag. Phase is dominated by small changes in distance and/or in carrier frequency and repeats every wavelength. Finally Doppler is the frequency shift between the transmitted and reflected signals caused by quickly moving objects. Each time a tag is read, the RFID reader measures these physical layer channel parameters and reports them along with the tag ID and the transmit frequency to the real-time host application. To retrieve the low-level data streams, a reader communication software was implemented in C# using Octane SDK provided by Impinj.

A plot showing 60 seconds of raw RSSI and Phase data for a single tag is depicted in Figure 3.4, panel A. During the first 20 seconds the tag is still, next the tag is moved for 20 second, and for the remaining 20 seconds it is still. This sequence of events can be inferred from the RSSI data but the phase data does not show a discernable trend. This is due to FCC regulations which require RFID readers in the 915MHz ISM band to pseudo-randomly change their transmit frequency in order to minimize interference with other devices. The result is that the RFID reader must frequency hop across 50 channels from 902MHz to 928MHz (in the USA) at an interval of approximately 0.2 seconds. This causes significant discontinuities in the RF phase reported by the reader as a function

of time (see Figure 3.4a), which makes detecting tag movement particularly difficult. However, the RFID reader also reports which channel (aka frequency) was used when a tag is read. Thus, re-mapping the window of the RF phase data from time into transmitted frequency (as shown in Figure 3.4c) reveals well-defined structures that can be used to build classification features. One of the key insights is that these low-level channel parameters represent a snap shot of the RF environment that is unique to each tag. Each tags RF environment is comprised of, the far-fields signal path from the reader to the tag (including all multipath elements), as well as the objects within the near-field region of the tag, which has an effective radius around the tag of a half wavelength (16cm). Thus, any changes in distance and/or tag orientation will result in altering the signal paths and will be reported as change in RSSI and/or RF phase. By watching the change in these parameters over time, the state of an individually tagged object can be inferred. Furthermore, changes in the near-field region of the tag (such as hand touch) will alter its resonant frequency and/or the impedance match between the RFID IC and the antenna. Both of these effects will be reported as changes in RSSI and RF phase as reported by the reader.

3.2.2 Feature Selection and Mchine Learning

Through experimentation, it was determined that 2 seconds is approximately the upper bound needed for the participants to complete translation, rotation, swipe touch, and cover touch interaction in a natural fashion. Thus a 2-second sliding window (which is advanced each second) was employed to segment the RSSI and RF phase data stream to generate features as inputs for the object interaction classifier. A longer window could be used to identify longer object interaction but it would also increase the latency of the real time system. Figure 3.5 shows examples of the raw RSSI and RF Phase signals for the same object undergoing four different types of interactions. Figure 5a shows a still tag (i.e., no human interaction) with the RSSI vs time plot on the top and the RF phase vs. transmit frequency plot on the bottom. The RSSI vs time plot is relatively stable for a two second time window and the RF phase vs. frequency plot shows phase decreasing at a constant slope. Translation (Figure 3.2b) and rotation (Figure 3.2c) has a major influence on phase variation while swipe touch (Figure 3.2d) has a major influence on RSSI variation. To minimize the influence of RF signal multipath effect, the features are based on differentials rather than absolute values of

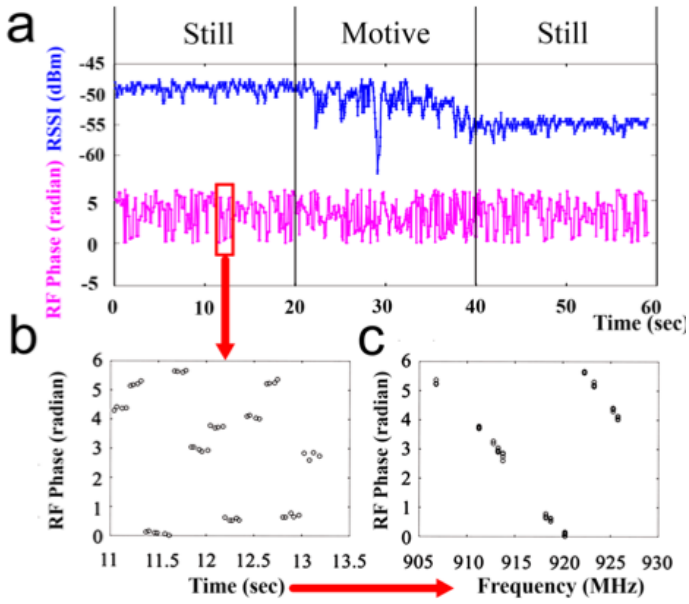


Figure 3.4: IDSense: raw RSSI and Phase data for a single RFID tag undergoing an interaction event (panel a). Due to pseudo-random frequency hopping a two second windows of phase data (panel b) must be sorted by channel (i.e. frequency) to reveal the tag expected phase behavior (panel c).

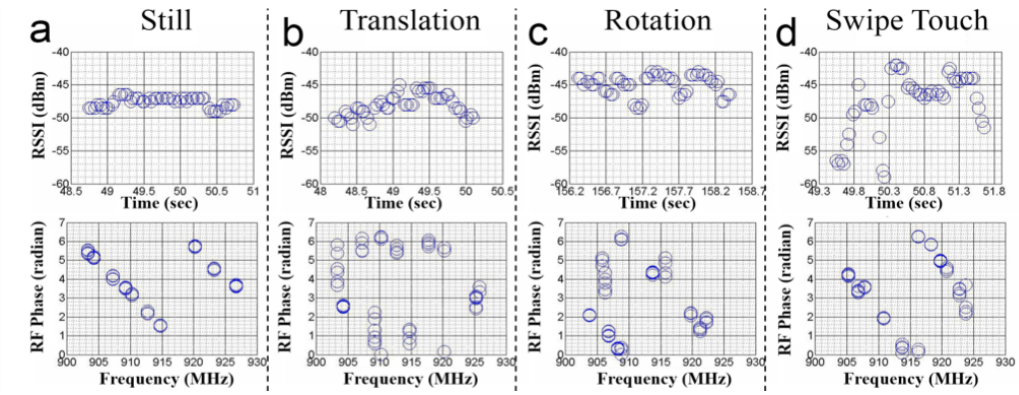


Figure 3.5: IDSense: tag RSSI vs time and phase vs frequency for Still, Translation, Rotation, and Swipe Touch over a 2 second window

RSSI and RF Phase. Eight features have been implemented from RSSI, RF Phase, and read rate.

RSSI Features

Generally speaking, changes in RSSI are predominantly caused by changes in the distance between the reader and the tag as well as the orientation of the tag antenna. However, it is well known that multipath effects can cause unpredictable variations in signal strength between a transmitter and receiver. In real world settings multipath increases the spatial variation in RSSI and thus provides a greater likelihood of detecting motion events. To identify these changes the following features have been selected.

- Standard Deviation of RSSI signal
- Mean of RSSI Standard Deviation within each frequency
- Mean of RSSI difference

RF Phase Features

RF phase is sensitive to smaller changes in distance between the tag and reader and is particularly useful for detecting translation. Additionally in Figure 3.4, the frequency hopping effect demon-

strates RF Phase dependency on channel frequency, which results in the phase related features being divided into two subgroups: the Constant Frequency Phase Rate (CFPR) and the Variable Frequency Phase Rate (VFPR). Since the RFID reader performs many tag reads on a single frequency before hopping to the next channel, changes in the phase are a good indicator of an interaction event. Here the Constant Frequency Phase Rate is defined as

$$CFPR = Phase[i+1] - Phase[i] \tag{3.1}$$

Where $Phase[i+1]$ and $Phase[i]$ are neighboring RF phase measurements at the same frequency (for a given time window). I employ the following three features to represent variations of Constant Frequency Phase Rate caused by human tag interaction.

- Median of the CFPR
- Sum of the absolute CFPR
- Standard Deviation of the CFPR

When the RFID reader does frequency hop from one channel to another, the change in frequency adds an additional dimension of information to infer human tag interaction. Equation (2) shows that the distance between the reader and tag is proportional to the partial derivative of the phase with respect to the derivative of frequency, as described in [108]

$$d = -\frac{c}{4\pi}\frac{\partial\theta}{\partial f} \tag{3.2}$$

Therefore the Variable Frequency Phase Rate is defined in equation (3) as the incremental change in phase divided by the incremental change in frequency

$$VFPR = \frac{Phase[i+1] - Phase[1]}{Frequency[i+1] - Frequency[i]} \tag{3.3}$$

Finally, since VFPR is proportional to the distance of the tag to the reader, the standard deviation is used to determine tag motion.

- Standard Deviation of VFPR

Read Rate per Tag

A cover touch event on a tag will dramatically weaken the received signal strength, which results in decreased read rate. Read rate of an uncovered tag ranges from 15 to 40 per second, while a covered tag (half covered or fully covered) usually has a read rate less than 10 reads per second.

- Read Rate: Number of packets received from each RFID tag per second.

Doppler Features

Doppler shift is used in a number of radio sensing scenarios to infer the relative motion of two radio systems. several features were developed using the Doppler information reported by the reader; unfortunately due to the relatively low speeds of human motion and the large amount of noise in the signal, these features did not prove to be expressive enough when compared to the other eight features. It should be noted that RFID based Doppler shift features would be useful in fast moving scenarios such as outdoor sporting activities and automotive settings.

3.2.3 Motion and Touch Event Classifier

In a pilot study conducted in the lab setting depicted in Figure 3.3, I determined that the system is able to detect horizontal translation greater than 20 centimeters in distance, vertical translation greater than 10 centimeters, and rotation of more than 45 degrees; all in a 2 second window. One limitation of the system is that when objects are moved very slowly, there may not be a significant RF signal change (in the two second sliding window) to be classified as an interaction event. At the other extreme, it could be possible to start and complete an object moment so quickly that the data would only appear as a brief impulse, thus making classification difficult. This edge condition was difficult to produce and was not observed in practice. Note that we later take this motion sensing capability one step further to track one dimensional traces of moving tags. Please refer to Chapter 4 Section 4.2 for more details.

In a pilot study, 600 instances of interaction events were recorded with one participant interacting with one tagged toy doing 5 types of interactions: still, translation, rotation, swipe touch and cover touch. For translation, the toy is moved by 20 centimeters and in the rotation class, the toy is rotated by 90 degrees. Interactions are conducted on a table top with the RFID reader antenna mounted on

the ceiling facing downwards at a distance of approximately 3 meters. A real-time classifier using Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel [21] was implemented in Matlab, which received streaming RFID read events from a Java application over TCP/IP.

A 5-class classifier is trained based on the 600 annotated instances and tuned parameters in the RBF kernel by 10 fold cross validation. This approach was able to achieve an 86.0% accuracy (table 3.1). However, the major classification confusion occurred between the translation class and the rotation class, which could only be distinguished with a 74.7% accuracy. It is believed that using multiple tags on a single object could improve these results in the future. Since the goal of this work is to robustly identify human object interactions, rotation and translation classes were combined into one motion class. The final classifier detects human tag interaction including still, motion, swipe touch and cover touch. The 10-fold cross validation results in a 4-class classifier, which shows an improved accuracy of 95.7%. It should be noted that the classifier is invariant to which object was used to train the classifier. Thus annotated data from all objects are being used to train a uniformed model. In later sections I demonstrate that this classifier can be used by different participants, without the need for retraining. This make sense, since the system is detecting that an object is being interacted with, not a gesture or action that is unique to an individual.

The real-time classifier reports results once per second for each tag, which is accomplished by sliding the two-second window over the data stream in one-second intervals. The following sections investigate several applications using the IDSense system and are implemented using the real time classifier.

Classified as	**Still**	**Trans**	**Rotate**	**Swipe**	**Cover**
Still	97.5%	1.7%	0%	0.8%	0%
Trans	0.8%	68.3%	28.3%	2.5%	0%
Rotate	0.8%	20.0%	74.2%	4.2%	0.8%
Swipe	0%	3.3%	3.3%	93.3%	0%
Cover	0%	0.8%	0.8%	1.7%	96.7%

Table 3.1: IDSense: 10 fold cross validation result for 5-class classifier

3.3 IDSense Application and Evaluation

IDSense sacrifice the rich sensing data of an objects state, for a low cost method of instrumenting that object with an passive RFID tag. The previous section demonstrated that the system is fundamentally capable of identifying basic human object interaction in a normal office/lab environment. To gain a better understanding of the capabilities of the system it is useful to focus on a few application spaces.

The following sections explore three application scenarios and provide a deeper analysis of system performance. It should also be noted that human subject approval was obtained for all studies in this work and the RFID equipment used is commercially available and meets the FCC regulations for health and safety as well as radio interference.

3.3.1 Interactive Storytelling with Physical Toys

An interactive story telling application is shown in Figure 3.6 where a stuffed toy lion is enhanced with a low cost RFID collar that communicates with a higher cost game console (RFID reader) connected to a computer or TV. When a child plays with the real toy, interaction events are recorded by the IDSense system which triggers actions by a virtual character on a computer screen. For instance a swipe touch near the collar is interpreted as petting the lion, while a cover touch triggers the character to take a nap. Likewise shaking the lion causes the digital character to dance. Any of these actions can advance the plot line of the story and be used to trigger visual and audio feedback. Since RFID tags inherently provides unique identification information, multiple toys can be used simultaneously to create complex and dynamic stories. Additionally, each toy can be personalized based on previous story lines or users preferences using a database or the writeable memory in the RFID tag. Ultimately IDSense offers unobtrusive way to bridge interactive digital media with real world toys and objects.

Study Design Overview

In this study I evaluate the performance of the system at classifying events necessary to support the interactive storytelling scenario. These events include the toy being: still, in motion, swipe touched, and cover touched. In this case the motion class includes both toy translation (Figure 3.2a) and

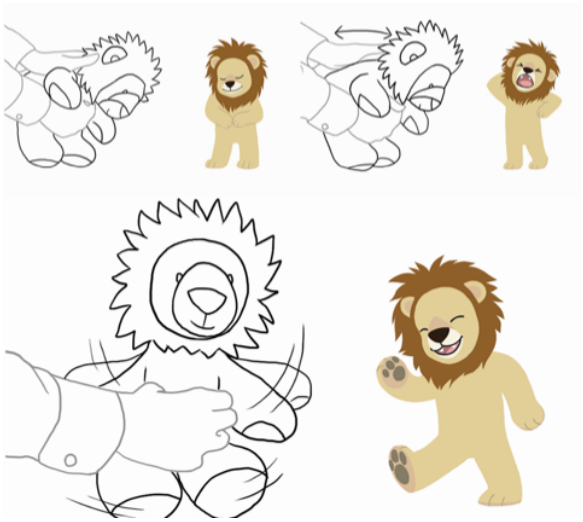


Figure 3.6: IDSense: an example of interactive storytelling where interaction with a physical toy lion, such as petting, holding, and shaking triggers digital character actions and plot events.

rotation (Figure 3.2b).

Five tagged toys were placed on a table measuring 140cm x 70cm as shown in Figure 3.7, which is in the same location as shown in Figure 3.3. The table was divided into 6 sections marked with number 1-6. For move events, the toys were moved from one section to another, distance between neighboring sections was approximately 20 to 35 centimeters, and non-neighboring sections approximately 40 to 80 centimeters.

Study Procedures

11 participants were recruited including 7 males and 4 females, with a mean age 25.7 years. Each participant finished the study independently, spanning a period of over two weeks. During the study, each participant was asked to follow visual instructions on a monitor to perform 10 instances of each of the following interactions on 2 randomly selected toys. Instructions were given once every 5 seconds. The instruction script is used as ground truth and users mistakes are manually annotated. I also monitor interaction records on the 3 unselected toys to test false alarms triggered by interaction with nearby objects.

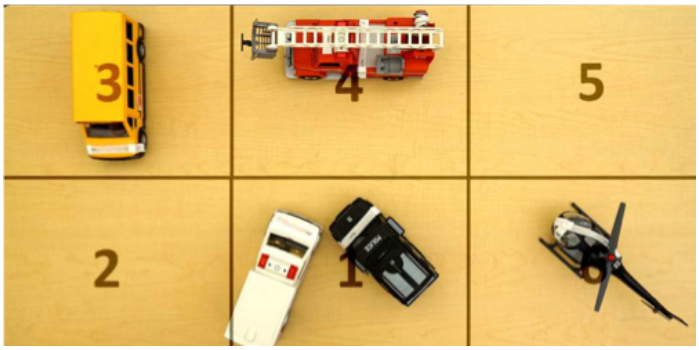


Figure 3.7: IDSense: study setting for toy interaction

1. Translation: Translate 2 toys simultaneously between 6 sections for 10 instances (Figure 3.2a, Figure 3.7). 5 translations to neighboring sections and 5 translations to non-neighboring sections. Note that 2 toys were following different paths during translations.
2. Rotation: Rotate 2 toys simultaneously by approximately 90 degrees for 10 instances (Figure 3.2b)
3. Swipe touch: Perform 10 swipe touches on 2 toys simultaneously. (Figure 3.2c)
4. Cover touch: Perform 10 cover touches on 2 toys simultaneously. (Figure 3.2d)

Training & Testing

A 4-class interaction classifiers was trained based on annotated data from one participant and tested the classifier for all other 10 participants. Classification results are reported in real time (and recorded for post processing), but were not made visible to the participants during testing. The script with annotated mistakes was then compared to data collected by the RFID reader.

Evaluation Results

The system achieved an average of 93.7% (SD=2.1%) classification accuracy for 5 toys on 1600 instances across 4 classes collected from 10 participants, and a 2.8% (2400 instances) false alarm rate on the 3 still reference toys. Table 3.2 shows detailed classification results. These results show

that the system is capable of classifying multiple tagged objects even when simultaneous interaction events occur. During natural interactions with toys, participants would have to reach over still toys in order to pick up adjacent objects when prompted. This could cause a change in the RF signature of the still toy since the arm of the participant would partially blocked some of the RF signals from the reader, potentially causing a false positive. However, the training session included this type of interference and the results show that it did not cause a significant false positive rate.

Classified as	**Still**	**Motion**	**Swipe**	**Cover**
Still	94.5%	2.5%	3.0%	0%
Motion	2.3%	92.3%	5.5%	0%
Swipe	1.0%	5.5%	93.5%	0%
Cover	0%	3.0%	2.5%	94.5%

Table 3.2: IDSense: classification results for user toy interaction events. Five toys where interacted with in total and at any one time two toys where being interacted with

I also evaluated the system in the same setting with only one tagged toy to get an upper bounds on performance. This accuracy is only slightly higher than then the 5 tag scenario with an accuracy of 95.3% (SD=4.0%).

3.4 An Activity Recognition Study in a Home Environment

To validate the effectiveness for the activity recognition approach, we conducted a naturalistic study across 4 different spaces in a living environment: a living room (Figure 3.8a), a kitchen (Figure 3.8b), a dining room (Figure 3.8c), and a bathroom (Figure 3.8d). We instrumented 110 commonly used objects with RFID stickers. RFID reader with a single antenna is placed in each of these environments, the positions of the antennas are highlighted with the red boxes in Figure 3.8. 10 participants were recruited, who are undergraduate and graduate students (7 males, 3 females), aged from 21 to 28 to participate in the study. The study happened in 7 different days during a period of 1 month. On each day, there are either 1 participant individually or 2 participants working together to finish all or a subset of the 24 activities outlined in Table 3.3.

Preparing meals	1. Blend smoothie, 2. Make cereal, 3. Make a sandwich, 4. Cook rice, 5. Garlic shrimp, 6. Fried egg with tomato, 7. Roasted chicken and potato, 8. Sweet and sour ribs, 9. Spinach almond salad; 10. Pan seared salmon; 11. Caesar salad; 12. Potato salad; 13. Pork chops; 14. Make tea.
Eating	15. Eat dinner
Mobility	16. Climb stairs
Medication	17. Take vitamin pills
Housework	18. Mop floor
Entertainment	19. Watch TV; 20. Read a book
Personal hygiene	21. Wash hands; 22. Brush teeth; 23. Wash face; 24. Dry hair

Table 3.3: IDSense: activities conducted in the user evaluation study

Figure 3.8: In home study environments and example activities (a) kitchen: making smoothie; (b) bathroom: comb hair; (c) dining room: take medicine; (d) living room: read book

3.4.1 Data Collection and Processing

26 hours of RFID data in total was collected as participants went through 24 different activities in these environments. These 24 different activities belong to a few categories of ADLs and instrumental ADLs, referring to people's daily self-care activities proposed by Sidney Katz [72]. These activities cover categories including preparing meals, eating, taking medication, housework, reading, personal hygiene, mobility, and entertainment. Table 3.3 provides more details on these activities. Examples activities are shown in Figure 3.8. A ceiling-mounted camera is deployed to collect ground truth information about object usage and activities. We created a video analysis tool to manually annotated object usage and activities happened in the study. The data was annotated using 1-second sliding windows. 2 annotators were employed to label these videos independently and the sections in which their labeling are consistent was utilized as ground truth (97%) and the inconsistent sections are ignored (3%).

3.4.2 Sensing Object Usage

The way we utilize daily objects varies with the nature of the object. To improve performance in object usage sensing, daily objects are grouped into 2 categories and apply different tagging and usage detection strategies. The 2 categories include objects that are mobile while being used and objects that are static while being used. For example, floor mop (Figure 3.9a) are under motion when performing cleaning tasks, while sitting bench (Figure 3.9b) are static when seated.

Tagging strategy

Objects that are mobile under usage are instrumented with a single tag (Figure 3.9a). Most of the objects used in the evaluation study come from this category. For objects that are static under usage (Figure 3.9b), a tag is attached every 30 cm on the object; a single tag is applied to objects that are smaller than 30 cm.

Tag states and object usage correlation

Object usage was detected leveraging its correlation with object states. In Figure 3.10 I visualize the time series usage data and the object states of a floor mop (Figure 3.10a) and a sitting bench (Figure 3.10b) for 100 seconds. These data are collected in real-world living environments. For each object, the data is segmented into 2 sections. In the first 50 seconds, the object is not in use while in the next 50 seconds it is being used. For mobile objects, the motion is a direct indication of object usage. However, other states are also detected including blocked, still and a few interfered examples. This is a typical situation for mobile object since the pauses in between motions will generate still states while interference and blocking from the human body will generate interfered and blocked states. The distinctive signature for usage is the increased variations in object states when compared to sections where the object is still. I apply a moving buffer with length x seconds to capture this variation and determine usage when the SD of object states (0 to 3) in the buffer breaks threshold p. The buffer size and the threshold will be optimized using training data. For tags that are attached to static objects, such as the ones on the bench, usage can be determined by the absence of tags when users are seated.

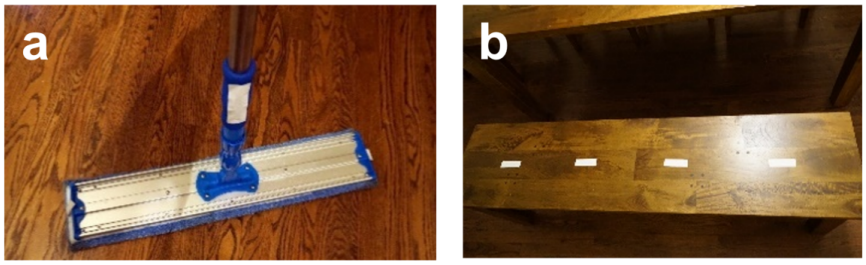


Figure 3.9: IDSense: Example objects for each of the two categories: a. mop, a mobile object, b. bench, a static object

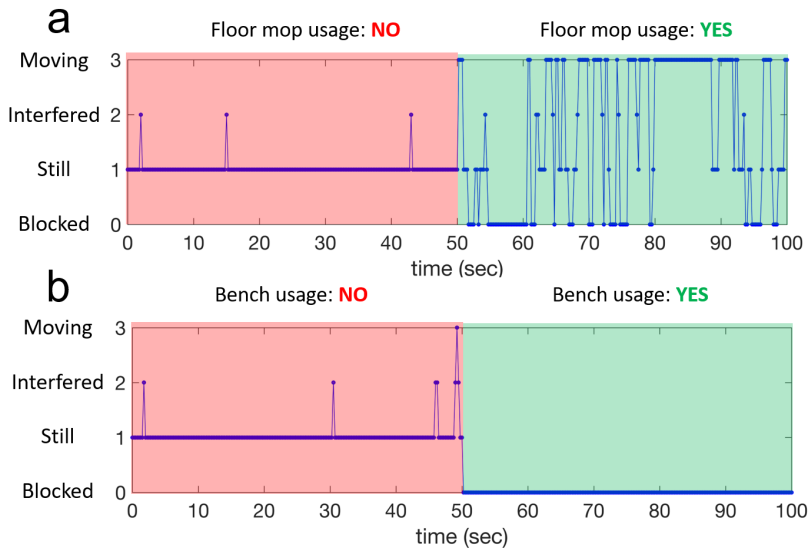


Figure 3.10: IDSense: correlation between tag states and object usage: a. mobile objects, b. static objects

Evaluation

To evaluate the object usage sensing approach, The RF channel parameters of each tag are processed into object states using the object state classifier trained in the controlled object state study. There are in total 7 days of data. Annotated object usage data were collected during one day for optimizing parameters in the object usage classifier by maximizing object detection F1 score. We then evaluate this object usage model (tag absence for static objects and $x = 21s, p = 0.94$ for motive objects) on other 6 days of annotated object usage data. the system achieved object usage detection with a precision of 92.4% and a recall of 90.9% for all objects involved in the study. It is worth noting that the false positive is low (0.03%). During the study, each object will generate on average about 1 false positive every hour.

3.4.3 Activity Recognition Approach

Previous literature has demonstrated how to recognize daily activities from object usage [117, 17]. In this work, I employ one commonly used methods: hidden markov models (HMM). Note that the contribution here is not meant to be of machine learning, but it serves as an important piece to demonstrate the feasibility of the RFID based object usage data towards activity recognition.

HMM Emission & Transition Probabilities

The HMM classifier considers the activity sequence to be the hidden states and the object used to be the emission states. There are three variables that need to be optimized from the training data, the initial probability of all activities, the transition probabilities between different activities and the emission probability from the activates to the objects. These concepts are explained using the following example. Consider a list of two activities, each with a set of utilized objects

- Blend smoothies: Orange, Apple, Cutting Board, Ice, Blender, Milk, Water Cup
- Make a sandwich: Cutting Board, Plate, Bread, Cheese, Ham, Plate

Here, the activities are hidden states, and their corresponding objects are the observations. The emission probabilities for activity ”blending smoothies” refers to the conditional probability of each

object given the activity. (E.g. the probability of using ”orange” given activity ”blending smoothies”). The transition probability control how the activity is chosen given the state of the previous activity. For example, what is the chance that a user is making a sandwich given that he was previously blending smoothies? The initial probability is the prior likelihood distribution of the two activities. These parameters are learned using training data in the evaluation section. The system infer the most likely sequence of activities by maximizing the joint probability of the activity sequence and the observations of objects. The most likely sequence of activates for any given sequence of objects can be effectively determined using the Viterbi algorithm [38]. For more details on the HMM-based activity recognition, please refer to the evaluation section.

3.4.4 Activity Recognition evaluation

In this section, I leverage the object usage data to classify 24 different activities outlined in table 3.3 . Given that there is a dedicated antenna installed in each room, we can refer to the corresponding antenna number to narrow down the search space for activity recognition. In particular, activity #1 to #13 is conducted in the kitchen, activity #14 to #17 is conducted in the dining room, activity #18 to #20 is conducted in the living room and finally activity #21 to #24 is conducted in the bathroom. Next, time series data of object usage are segmented into ”episodes”. For example, if the object is continuously used, for 30 seconds, that will be considered a single episode of object usage. Each episode will be considered one emission state from an activity (a hidden state) in the HMM model which are ordered by their starting time. The activity recognition results will be compared with manually labelled ground truth for evaluation. For example, given a continuous usage of a ”floor mop” from 0 to 30 seconds, the HMM inferred the most likely activity for this 30-second episode is ”cleaning the floor”. However, ground truth labeled 20 to 30 seconds as ”watching TV”. So that period will be considered as false positives for ”cleaning floor” and false negatives for ”watching TV”. To evaluate the HMM for activity recognition, I apply leave-one-out cross validation. The HMM is trained on data from 6 days and then evaluated on the data collected on a remaining day. In Figure 3.11, I compare the activity recognition accuracy using manually annotated object usage with usage generated by RFID tags. Given that different activities have different time duration. I normalize activity recognition results by time and evaluate the average F1 score accuracy across all

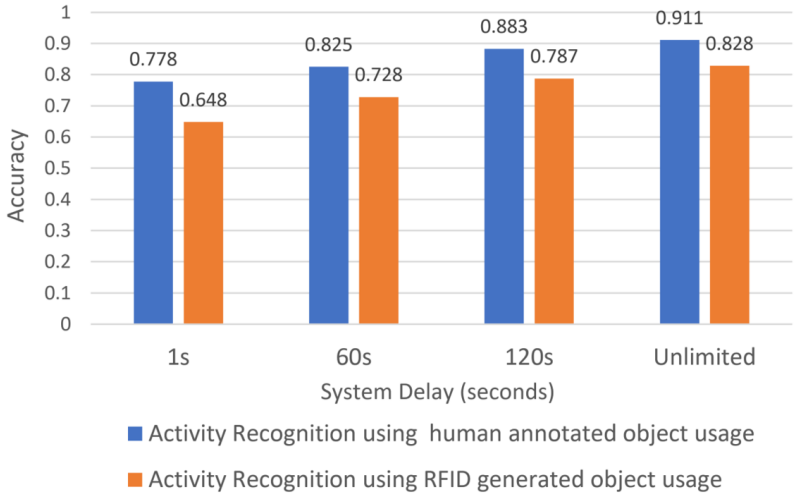


Figure 3.11: IDSense: average activity recognition F1 score for 24 different daily activities

24 daily activities. A random guess in this case would yield an F1 score of 4.2%, and the system was able to achieve an average F1 score of 82.8% recognizing 24 different daily activities.

Figure 3.4 shows the classification accuracy for each of the 24 activities described in Figure 3.3. In general, the accuracy for recognizing sub-activities in the food preparation category (#1 to #14) is low when compared to all other activities since they take place in the same environment and share a number objects, making it a challenge to differentiate between these activities. Activities that take place in other environments have a comparatively distinctive set of objects, making it easier to accurately recognize and classify these activities. In addition, given activities across different rooms does not involve any shared objects, the system did not generate any false positives and negatives among activities conducted in different locations.

3.5 A Preliminary Study in a Retail Environment

To better understand the effectiveness and limitations of the approach, I conducted an evaluation study in a retail gift store. Items are placed on multiple layers of shelves in the store (Figure 3.12). In the prior studies, the maximum number of tags visible to a single reader is around 100. Given the opportunity of the high density of products in the store, I evaluated the system performances when a single reader is exposed to approximately 100, 200 and 300 items and see how well the algorithm

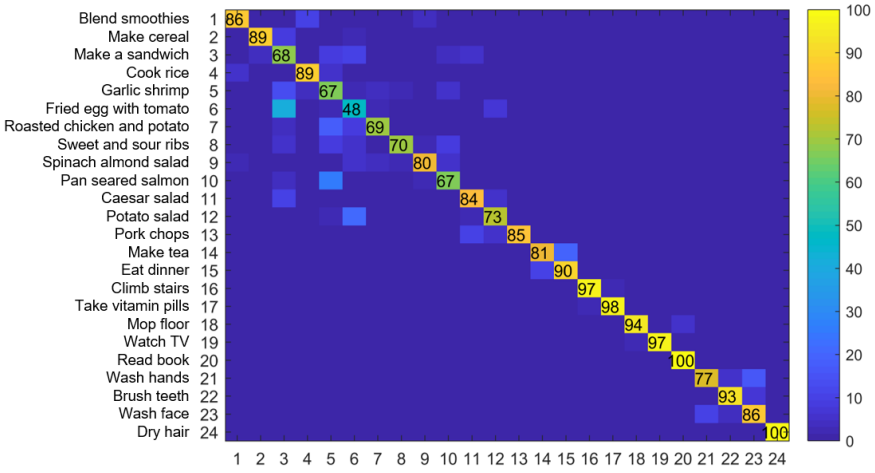


Table 3.4: IDSense: activity classification confusion matrix by class

could capture user pick up events in these scenarios.

In a staged study, I evaluated three different sections of the store including a clothing section, a gift box section as well as a section with hand-held bags. In three separate session of studies, I controlled the number of tags visible to each reader at approximately 100, 200, and 300. The study covered 63 shelves from these sections and we conducted 20 pick up events on each shelf. The false positive rate of pick-up events was evaluated during an hour-long study when the store is empty, and no interaction events occurred. The results for detecting these pickup events are as follow. Under the 100 tag scenario, the system achieved an average recall rate of 81% with a false positive rate of 0.005%. Under the 200 tag scenario, the method achieved an average recall rate of 65% with a false positive rate of 0.003%. Under the 300 tag scenario, the system achieved an average recall rate of 38%, with a false positive of 0.002%. We can see here that the recall rate of sensing pick up events is impacted by the number of items visible to the reader.

The decreased recall rate is primarily affected by the insufficient read rate for each tag. The Gen-2 UHF reader has a maximum read rate of 1000+ per second. However, in reality, maximum read rate was observed at between 700-800 reads per second. The increase in overall tag volume will dramatically impact the read rate for each tag, creating gaps in reads and making it more challenging to capture these pick-up events. Even though features across different channels were utilized to

Figure 3.12: IDSense: a section with multiple shelves in the gift store for evaluation purposes

mitigate the effect caused by channel frequency hopping, having a read rate of lower than 5 per second means each RF channel will only capture a subset of the tag visible to the reader, making it challenging to understand the motion state of tags. Prior research has demonstrated methods to improve the overall read rate through modification of the communication protocol to minimize collisions between tags. If these methods could be incorporated into the reader firmware upgrade or the design of the next generation of RFID, it would help the case of interaction sensing, especially under the scenario where items are densely populated.

I also investigated the effect of signal coverage across different layers of the shelf. Take the clothing section as an example, In Figure 3.13 items placed at four different locations (A, B, C, and D) were evaluated. There are in total about 100 tags exposed to the RFID reader. The picture wasn't able to capture the antenna but it is placed on top of location B as illustrated in the figure with a 30dBm power setting, making it very easy for detecting items placed at that location. Location B has a pick-up recall rate of 94%. However, for location A which is towards the edge of the antenna beam angle and location C, D which are covered by many layers of clothing, it is still challenging for the reader to query tags with the same speed as location B. As a result, the recall rate is reduced to 85%, 69% and 78% accordingly. The RFID industry is improving the sensitivity of reader by approximately 1dBm per year, which will help with interaction sensing challenges related to signal

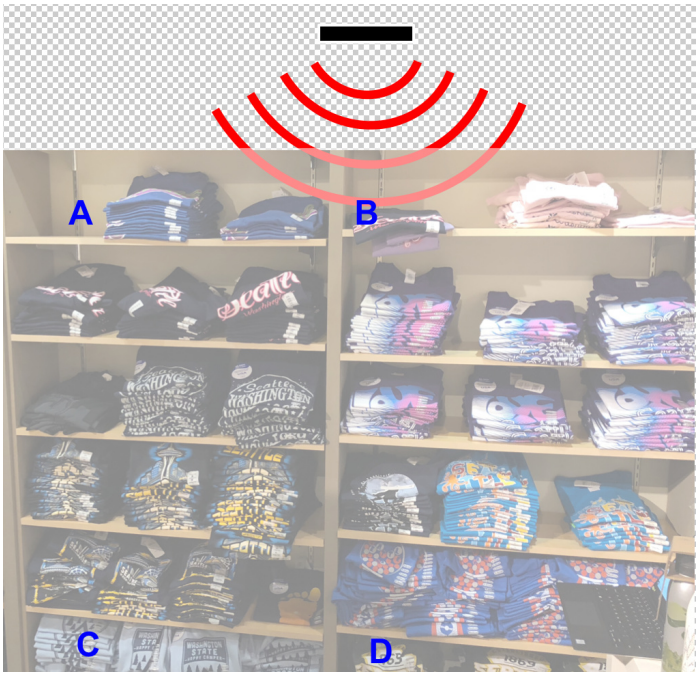


Figure 3.13: IDSense: a clothing section with variable sensing locations A, B, C and D.

coverage.

Several principles will help with the case of increasing signal coverage for interaction sensing proposes. One should try to reduce materials with a high dielectric constant in the line-of-sight path as well as in proximity to the tags. In addition, it will also be beneficial to reduce the overlapping between tags. For example, in the case of figure 3.13, if the antenna is placed right on top of the shelf, there will likely be many layers of overlapped items. If the antenna can be moved outwards and then tilted towards the shelf surface, that would improve signal coverage at the bottom of the shelf. Finally, reducing tag density will also be helpful. When tags are in the near field range (approximately 15cm for UHF tags) from each other, it will be challenging for the reader to pick up these tags at long ranges.

It is also worth pointing out that the sensing accuracy for pick-up events is also impacted by the reader's management of power between antennas. When there is only one antenna attached to the reader, the antenna is powered 100% of the time; however, the reader can support up to 4 different antennas. When all antennas are plugged in, the reader can only power one antenna at any given

time. And each antenna is working under 25% duty cycle. In the experimentation, I noticed that it could take up to 0.7 seconds before the reader come back to the same antenna, which means any interactions happened in between these gaps will not be tracked by the reader.

One interesting observation is that the trade-off between recall (sensitivity) and false positive can be tuned by thresholding the confidence output of the machine learning classifier. Consider a 2 class classifier with 'still' and 'pick-up' as outputs. The default confidence threshold for classification is 0.5, which means that the classifier will output a result as 'pick-up' when its confidence level is more than 0.5. Lowering this confidence threshold will increase the system's sensitivity and improve the recall rate on recognizing 'pick-up' events, but will also generate more false positives. For location C in Figure 3.13, the default SVM classifier has a recall rate of 69% with a false positive rate of 0.005%. This level of false positive will generate about four false pick-ups during a daytime. The false positive needs to be kept low so that it does not overwhelm the real pick-up events. Now consider a case where we bring in other sensing mechanisms, such as present sensing system using IR sensors. The tolerance for false positive will be much higher since user presence is a prerequisite for item interactions. The RFID system only need to pay attention to the period of time when the IR sensing system is triggered by user presence. For example, the IR system determined that there is only 10 percent of the time during a day when a user is standing in front of the location C. In this case, a false positive rate of 0.5% on recognizing the 'pick-up' events will generate approximately the same amount of false positives during a day time. Given this improved tolerance, the system can achieve a much improved recall for detecting these 'pick-up' events. Consider another scenario where more information can be retrieved from other parallel sensing system. For example, if we can introduce cameras with computer vision techniques to look at when these pick up events happened. In this case, the RFID system only needs to recognize and disambiguate similar objects. As a result, most items can be recognized by querying the velocity of all tags in proximity and pick the one with the highest speed. This will help further eliminate the possibility of false positives and improve the overall accuracy for detecting item interaction events.

Please see Chapter 7 for more discussions on improving reader performance, sensor fusion, and scalability in real-world applications.

3.6 Conclusion

This Chapter presents a robust method for enabling human object interaction by using minimalistic instrumentation in the form of passive UHF RFID tags. I have shown that it is possible to infer tag motion and touch events by measuring changes in the RF communication channel between the tag and reader with 93% to 97% accuracy. I also evaluated the system performance in a naturalistic study and achieved an F1 score of 82.8% recognizing 24 different daily activities. IDSense achieved similar or better results when compared with previous work where multiple antennas and extensive training data are required. I demonstrated an unobtrusive and scalable way for understanding activities in living environments. The minimal instrumentation and low maintenance nature of commercially available RFID systems lower the barrier for its adoption to real-world sensing applications.

One of the limitations of activity recognition is the requirement of excessive training data. In this work, this problem mitigated by instrumenting RFID antennas at a room level thus reducing the potential classes of activities in the HMM classifier. As a result, our system achieved promising recognition results given the relatively small set of activity data we collected in the study. In addition, the training data and testing data are generated by different participants, demonstrate the feasibility of applying the activity recognition approach towards new users without the requirement of retraining. For evaluation purposes, we manually labeled video recordings to collect ground truth. In real world applications, this will not be necessary since object usage events will be automatically determined by the object usage classifier. The only annotation required would be the annotation of daily activities for training supervised activity recognition mode. An alternative to video recording would be self-reporting, which can be supported by smart phone applications.

One thing that could be improved in the study is the tagging strategy. A single tag is applied to each of the mobile objects being utilized in the study. At times when the tag is being occluded by the hand of the users, it is challenging to determine the state of the object. This situation can be improved by having an additional tag at a different location of the object, which will reduce the probability of occlusion. In addition, it can also help us detect fine-grain actions such as holding on to the object which can be achieved by comparing the state of the 2 tags attached to the same object. However, having more tags on objects will increase the user burden of putting tags on objects. This

issue can be improved by embedding tags in objects during manufacturing.

In the future, I am interested in further explore reducing the training required for activity recognition. Instead of using a list of predefined activates, I want to explore the possibilities to automatically cluster activities into different categories leveraging the object state and human presence data provided by our system. In addition, I am interested in studying how to detect and differentiate multiple activities conducted by multiple users simultaneously. I consider these as important next steps towards the goal of creating an intelligent system that provides ubiquitous and fine-granular activity recognition as a service to support smart environment, personal wellness and assisted living applications.

Chapter 4

CREATING PASSIVE BATTERY-FREE WIRELESS INTERFACES ON PAPER

4.1 Introduction

In this chapter, I describe a technique named PaperID [83] that allow inexpensive, ultra-thin, battery-free Radio Frequency Identication (RFID) tags to be turned into simple paper input devices. Signal processing techniques was utilized to determine how a tag is being manipulated by the user via an RFID reader and show how tags may be enhanced with a simple set of conductive traces that can be printed on paper, stencil-traced, or even hand-drawn. These traces modify the behavior of contiguous tags to serve as input devices. Our techniques provide the capability to use off-the-shelf RFID tags to sense touch, cover, overlap of tags by conductive or dielectric (insulating) materials, and tag movement trajectories. Paper prototypes can be made functional in seconds. Due to the rapid deployability and low-cost of the tags used, PaperID can create a new class of interactive paper devices that are drawn on demand for simple tasks. These capabilities allow new interactive possibilities for pop-up books and other paper-craft objects.

4.2 Technical Overview

The process for creating an RFID-enhanced paper interface employs a familiar set of supplies and techniques consisting of paper, RFID tags in the form of stickers, and pens with conductive ink, along with extras such as glue, scissors, tape, and markers. Here, I provide a brief overview of the necessary steps and techniques as an outline for the in-depth discussion in the following sections where I focus on the fabrication, signal processing, and machine learning techniques related to prototyping interactive interfaces using off-the-shelf tags as well as customized tags.

As shown in Figure 4.1, fabrication using tags with customized antennas typically begins with placement of an RFID tag on the paper or by drawing the antenna itself and adding a loop IC sticker to form a fully functional RFID tag. The ability to hand draw, ink-jet print, or use commercial RFID tags provides a great deal of flexibility in visual design, RF performance, and the types of

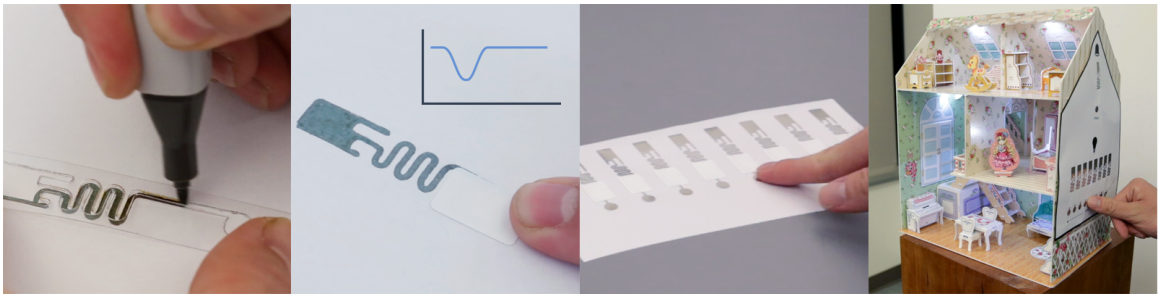


Figure 4.1: Example PaperID prototyping: a user creates an antenna with a stencil, affixes a loop IC, builds a slider, and uses the slider to control light brightness in a dollhouse.

interactions. I discuss the fabrication of these antennas in the Custom Tag Fabrication section.

In order to harness the wireless sensing capabilities of commonplace RFID tags (Figure 4.2 panel d), PaperID monitor changes in the low-level channel parameters of the reader / tag communication to infer human / tag interaction events. In the Signal Detection section, I discuss in detail the signal processing and machine learning techniques applied to low-channel parameters of this commonplace tag to enable a wide variety of on-tag and free-air interaction sensing capabilities.

Once the mechanics of building and sensing interactions are established, a number of primitives can be built in the form of paper knobs, sliders, pop-ups, etc. These building blocks are then combined with on-tag and free-air RFID tag gestures to create a variety of interaction types that can be used to build and prototype new interfaces and applications.

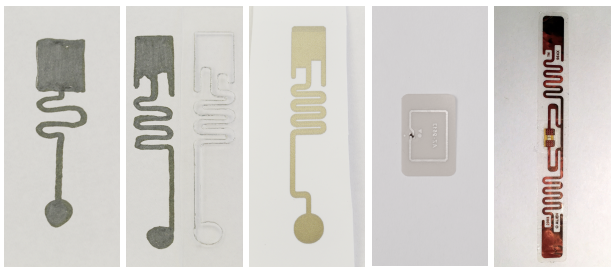
4.2.1 Background

In this work, a commercially available EPC Gen 2 UHF RFID system [32] that operates with a carrier signal at a frequency from 902 MHz to 928 MHz is employed and is capable of powering and reading hundreds of RFID tags simultaneously within a range of 6-10 meters. RFID readers vary in shape, size and performance with prices from $200 to $1,500. Generally, any full-featured reader that reports RSSI and RF phase can be used. We utilized commodity UHF Squiggle RFID tags and UHF loop tags that are inexpensive (10 cents), disposable, and completely passive in conjunction with user-customized tag antennas created with conductive ink.

4.3 Custom Tag Fabrication

UHF RFID tags are readily available in a wide variety of shapes and sizes that typically take the form of two-dimensional stickers or labels. While commercial tags offer the greatest possible performance in terms of read range, they are not suitable for dense deployment due to near-field interference. In this chapter, I implemented overcome this limitation by presenting a half-antenna design together with a sticker RFID IC as depicted in Figure 4.1(a,b,c). The design acts as an ungrounded monopole antenna, which performs poorly at harvesting RF energy thus making the tag invisible to the reader. When the antenna pad is touched, the human body (which is reasonably conductive at these frequencies) acts as a ground plane, which improves received signal power allowing the tag to harvest enough power for operation. The design of these 'silent' touch buttons will dramatically reduce interference between near-field tags. For rapid touch interface prototyping, I outline a few antenna construction methods that can be used to fabricate this type of customized RFID tag.

One standard approach to designing custom antennas is to use an inkjet printer with conductive ink [132, 178], which was used to print the half-antenna in figure 4.2(e). This provides the greatest amount of control and repeatability. One challenge to this approach is bonding the RFID IC (which is approximately 0.5mm x 0.5mm) to the printed antenna, which cannot be hand-soldered. This can be overcome by re-purposing an ultra small form-factor, near-field UHF RFID tag [155] into what is referred as a loop IC, as shown in Figure 4.2(d). Instead of having to mechanically and electrically bond an IC to the antenna, I use the loop IC to inductively couple to the spine of the printed antenna, thereby forming a fully functional custom RFID tag (i.e., antenna plus IC).



(b) Drawn (b) Stencil (e) Printed (e) Loop IC (e) Sticker

Figure 4.2: PaperID tag antenna types

An alternative approach is to use conductive ink pens to draw the antenna freehand (Figure 4.2(a)). While it is quite simple to draw a straight line design, other more compact antenna shapes can also be drawn, as shown in Figure 4.2. While hand drawn antennas do not have the 10-meter read range of their commercially available counterparts, with a little trial and error, it is quite easy to make hand drawn RFID antennas with a functional range of 5-6 meters. In order to help users become comfortable with drawing UHF antennas, a plastic antenna stencil can be used as a guide (Figure 4.2(b)). This also insures consistent performance from one tag to the next while still allowing people to prototype their interface quickly. In practice, I noticed that the conductive ink should have a resistance less than 10 ohms per inch in order to ensure an antenna with reliable performance.

4.4 Signal Detection

In addition to fabricating conductive ink antennas for loop ICs to enable responsive hand touching detection, PaperID also utilize other commodity tags to support interactions, including on-paper or in-air hand gesture detection as well as motion tracking of the paper interface.

In order to integrate rich hand interaction types that include touch, wave, swipe, cover and tag movement into a single tag (Figure 4.5(a)-(e)), PaperID employ the channel parameters reported by the RFID reader, such as Received Signal Strength Indicator (RSSI), RF Phase, and Read Rate, which represent a unique signature of the RF environment of each individual tag. Each tag's RF environment is comprised of the far-field signal path from the reader to the tag (including all multi-path elements), as well as the near-field region of the tag, which has an effective radius around the tag of half of a wavelength (~16cm). Thus, any changes in either the far-field or near-field regions of the tag, such as hand touch, wave, cover gestures, or tag movements, will result in altering the signal paths and be reported as changes in RF channel parameters. By putting tags onto the paper interface and observing and extracting the patterns of each tag's channel parameters, PaperID can establish a quantitative understanding of people's interactions with the interface. In the following subsections, I explain how PaperID make use of low-level channel parameters including RSSI, RF Phase and read rate to enable tag motion tracking, as well as on-tag and free air gesture detection.

4.4.1 Understanding Channel Parameters

RSSI

When an RFID tag receives power from a reader, part of the received power is reflected back to the RFID reader. This reflection is called backscatter and RSSI is a power measurement of the backscattered signal from each tag received by the reader. RSSI is predominantly determined by the distance to the tag as well as the power level of the reader, as shown in equation 4.1.

$$RSSI = 10log(P_r) = 10log(\frac{G_t\lambda^2\sigma}{(4\pi)^3 d^4}) + 10log(P_t) \tag{4.1}$$

where P_r = backscatter signal power, P_t = reader transmit power, G_t = reader antenna gain, λ = carrier wavelength, σ = tag radar cross section, and d = distance between the reader and tag.

Phase

The difference in phase between the signal transmitted by the reader and the backscattered signal from the tag as seen by the reader provides additional insight into the state of the tag. RF phase is dominated by the antenna-tag distance d as well as signal carrier frequency $f = 1/\lambda$, and repeats every wavelength. The constant shown in Equation 4.2 is introduced by the transmit circuits, receiver circuits, and the tag's reflection characteristic.

$$\theta = 2\pi\frac{2d}{\lambda} mod(2\pi) + constant \tag{4.2}$$

Read Rate

The RFID read rate is defined as the number of packets received from each tag per second. Read rate can be influenced by dramatic changes in signal strength as a result of blocking or capacitive coupling of the human body. To get a better understanding of these RF parameters, I attached a tag to a sheet of paper and show 15 seconds of the raw RSSI and phase signal streams of the tag in Figure 4.3. There was no interaction during the first 5 seconds, and for the next 5 seconds, the tag is undergoing constant hand waving in the near-field region. For the last 5 seconds, the tag is moving in a circular trace. The still state with no interaction can be distinguished from other

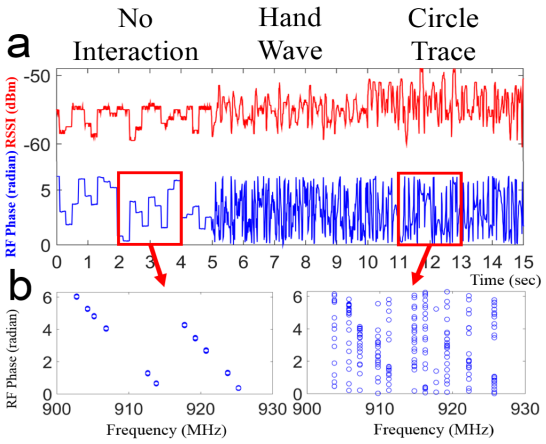


Figure 4.3: PaperID RF channel parameter example: (a) raw RSSI and Phase data for a single RFID tag remaining still or experiencing near-field hand waving and circular movement. (b) due to pseudo-random frequency hopping, a 2-second window of RF phase is sorted by channel to reveal characteristics of these events.

states by observing the RSSI and phase raw signal variations shown in Figure 4.3(a), noting larger variations in RSSI as well as Phase signal.

We can notice the discontinuity in RSSI and phase samples in the first 5 seconds of Figure 4.3(a). When there is no interaction involving the tag, this is a result of the RFID reader constantly changing its transmit signal frequency. FCC regulations require RFID readers in the 915 MHz ISM band to pseudo-randomly change their transmit frequency in order to minimize interference with other devices. To satisfy this requirement, RFID readers frequency hop across 50 channels from 902 MHz to 928 MHz (in the USA) at a time interval of approximately 0.2 seconds.

To better reveal the underlying characteristics of phase hidden by frequency hopping, I take a 1-second slice of phase from the still state and one from interaction state and re-plot against channel frequency in Figure 4.3(b). For the first 5 seconds, phase is linearly correlated with the carrier signal and wrapped into segments within $[0, 2\pi]$. Hand waving around the tag as well as circular movement of the tag resulted in dramatic phase variations within each channel, as depicted in the right panel of Figure 4.3(b).

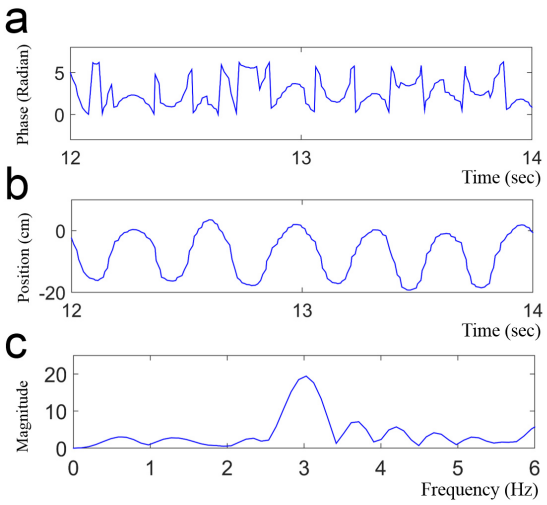


Figure 4.4: PaperID Motion trace reconstruction: (a) Two segments of raw phase data; (b) application of the trace recovering technique to reveal the moving trace relative to the RFID reader; (c) applying a Fast Fourier transform to reveal the frequency of the movement.

4.4.2 Trace Recovering

Here, I demonstrate a technique for tracking the movement traces of tags using the phase signal. This technique enables the ”continuous tracking” of the tag’s velocity, motion magnitude, and relative direction of motion towards and away from the reader. Figure 4.4.1 is a 2-second slice of seconds 12 to 14 shown in Figure 4.3. According to the phase definition in Equation 4.2, after unwrapping the phase signals, the distance change d_1 and d_2 can be calculated with the following equation:

$$d_1 - d_2 = \frac{1}{2}\frac{\lambda}{2\pi}(\theta_1 - \theta_2). \tag{4.3}$$

I apply this equation to the phase signal in panel (a) and then apply signal smoothing between timestamps where frequency hopping occurs. circular movement of the paper is recovered as represented in Figure 4.4.1(b). Note that according to Equation 4.2, distance d is the relative distance between the tag and the reader. In this case, the movement represented in the recovered trace is a 1-dimensional component of the circular trace, which is close to a sine wave. More details of this technique is discussed in Chapter 5 section 3

A Fast Fourier transform is applied to this trace signal to get the frequency component of the trace in panel (c). This trace-tracking technique is later applied to paper-based toy applications. Consider the "conductors wand" application where the frequency at which the wand is waved side-to-side is mapped to the tempo of the music while the magnitude of motion is mapped to volume. This technique provides a rich user interface, in contrast to the binary motion detection in IDSense [84], which would only be able to turn the music on or off.

4.4.3 Multi-Gesture Classifier

Here, I discuss the technical details of a multi-gesture classifier that is later utilized to enable controlling applications. PaperID support interactions including finger touch, cover, swipe and in-air directional hand waving on a single tag.

In order to enable a variety of gesture input methods on paper interfaces, I apply machine learning to these lower RF channel parameters to describe the unique RF environment of each tag on paper interfaces. For each segment, I calculate the following features.

Read Rate

- Read Rate: This feature is defined as the number of packets received from each RFID tag per second. This feature is effective for characterizing interactions that dramatically change the ambient RF environment, such as a cover gesture, or detuning events happening on the tag, such as a touch gesture.

Phase Features

- Tag position change: The relative position change within each segment characterized by the trace recovering technique. This feature is effective for describing motions related to a paper interface.
- Standard deviation of phase within each channel: This feature is effective for describing periodic hand gestures in the near-field region of the tag, such as swiping or touching the tag surface.

- Standard error of the linear regression of phase versus channel frequencies: As demonstrated in the left panel of Figure 4.3(b), when no interaction is happening on the paper interface, the phase is linearly correlated with the channel. Here, I unwrap the phase samples and calculate the error of the linear fit, which is useful for separating non-interaction states from other interactions on the paper interface.

RSSI Features

- Average Standard Deviation of RSSI per channel: As demonstrated in Figure 4.3, RSSI variation will increase when there are interaction events happening with the paper interface. The magnitude of the variation is dependent on the interaction type, so this feature is included here to describe the variation in signal strength related to different interaction types.
- Sum of RSSI difference per channel: This feature describes the directional changes of the RSSI magnitude, which is useful for monitoring immediate change in reflected signal as a result of near-field interference, such as hand touching.
- RSSI Average Value: The RSSI value can vary according to the relative distance, antenna orientation, and power level of the reader, so uncontrolled RSSI values cannot provide effective information about the interaction states of the paper interface to which it is attached; however, the initial RSSI value is calibrated to approximately -25dBm. In this case, the RSSI Average Value (within each segment) can indicate the interaction types that will change the RSSI, including on-tag and free-air interactions.

Machine Learning

Here, I introduce the machine learning pipeline for the purpose of distinguishing multiple simultaneous gestures on RFID tags. More details on how I have applied this pipeline to gesture classification applications will be discussed in later sections. First, RF channel parameters segmented by moving windows. Through experimentation, a window size of 1 second that advances every half second with 50% overlap was employed. This achieves a good balance between classification accuracy and delay in real-time. A longer window may help to further boost classification accuracy, but it would also increase the latency of the real-time system.

Phase, RSSI and read rate features are utilized to implement a Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel. Training data is collected offline. Training data for each of the interaction classes at multiple locations was recorded and labelled for training to improve the robustness of the classifier. Then, I optimize parameters of the RBF kernel by maximizing 10-fold cross-validation results. This classifier makes real-time predictions based on the most recent 1-second RF parameters and refreshes prediction results every 0.5 seconds.

4.5 Interaction Methods and Primitives

Thus far, I have presented several methods for quickly building custom RFID antennas using construction techniques consistent with prototyping on paper. By applying the signal processing techniques described above, PaperID is able to turn the custom tags and commercial tags (that normally only report their ID number) into battery-free, ultra-thin, wireless user input devices.

In this section, I describe several user interaction primitives that can easily be made with RFID tags. The first set of primitives consists of a single RFID tag that can be implemented in several ways. Next, I combine multiple RFID tags together to make complex interactive structures, such as rotary dials and sliders. Finally, free air gestures and actions are presented wherein the user is either interacting with a static paper object or is manipulating the paper object dynamically.

4.5.1 Single Tag Primitives

This is the simplest interface as it only requires the use of a single RFID tag. I describe gestures including cover, two types of finger touch, swipe touch, and hand waving (represented in Figure 4.5(a)-(d)). By applying the SVM classifier onto the RF features discussed earlier, all of these primitives can be integrated onto a single tag for real-time interaction detection and classification.

Waving

This technique is for detection of the free air hand waving gesture proximate to a paper interface. It can be applied to a single tag or multiple tags on one paper interface.

With a single tag, PaperID can detect the binary wave / no wave as represented in Figure 4.5(a). This is achieved by applying the SVM classifier to RF features. With the support of multiple tags

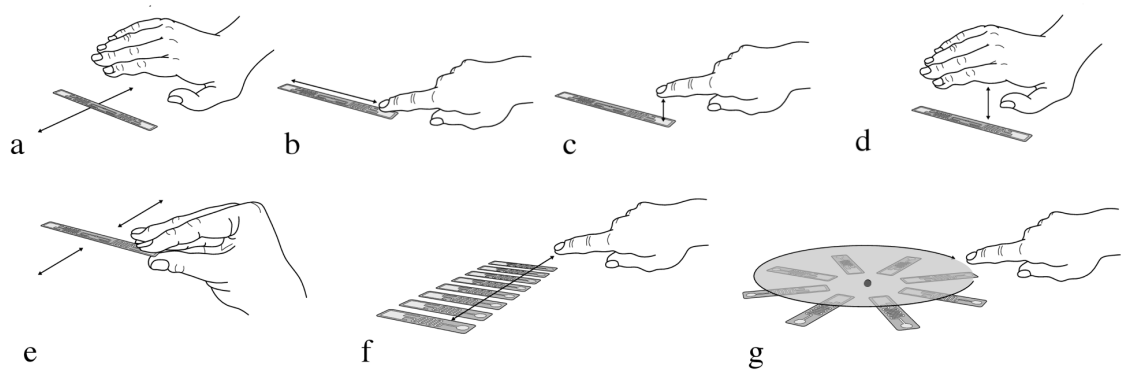


Figure 4.5: PaperID supported User interactions: (a) Wave (b) Swipe (c) Finger Touch (d) Cover Touch (e) Free air tag motion (f) Slider (g) Knob

on the paper interface, PaperID can also detect the direction of waving by monitoring the sequence of variation in RF features of the two corresponding tags.

Button Touch

Button touch interactions can be accomplished in two different ways. First, a commonplace RFID tag will be affected when a finger touches either end of its antenna. This type of touch will cause a change in the RSSI and RF phase as reported by the reader, but will not block the signal such that the tag can no longer be read as is the case with cover touch. Using SVM, it is easy to reliably classify touch events when the tag is still. However, if this type of button is used to prototype a remote control for a TV, it would be challenging to detect button pushes while the remote control is in motion.

To overcome this limitation, a second type of touch button based on a “half antenna” has been developed, as described in the Antenna Fabrication section. In this configuration, the tag operates as a button or key press. Multiple tags can be set up on a single sheet of paper and multiple buttons can be touched simultaneously. Because the actual operation of the tag is binary, it is very robust to the motion of the prototyped device the tag is on. However, since the button is *normally off / momentarily on*, the RFID reader will not know that the device is present until the button is pushed. If desired, the designer can add a second normal tag to the prototype to indicate if the button is

within view of the reader.

Cover Touch

One of the most basic RFID interaction methods is covering the RFID tag such that the signal from the reader is completely blocked or such that the tag's antenna becomes detuned to the point where it cannot receive enough power for operation (Figure 4.5(d)). This can be done by covering the tag with a person's hand or body or with a conductor like copper tape. The most effective way to detect these cover events is to measure the read rate of the tag and set a threshold for activation. One drawback is that when the tag leaves the interrogation zone of the RFID reader, its read rate will also appear to drop (i.e., go to zero) and will be registered as user input. To overcome this edge condition, a time-dependent interaction can be employed, such as covering the tag for a window of 1-2 seconds and then uncovering it.

Swipe Touch

PaperID also supports swipe touch interaction, which is the interaction of swiping a fingertip across the tag surface (Figure 4.5(b)). This gesture brings dramatic variation to both RSSI and phase signals, which can be characterized using the RF features.

4.5.2 Multi-Tag Primitives

The 'silent' nature of the customized touch button tag type makes them ideal for dense deployment. Below, I introduce more complicated primitives using multiple customized touch button tags.

Slider

For the slider interface, the user slides his or her finger across a row of customized button tags as depicted in figure 4.5(f)). Multiple tags with half antennas are placed on one static layer of paper; the finger is placed near the edge of the tab. As the finger is moved across the static half antennas, it can couple with each one to provide enough power to backscatter signals to the reader. Thus, the reader can detect the position of the finger as a discrete touch state sensor. For example, each static

tag can be paired with a different light bulb/LED, sound effect, or onscreen event that is activated when the moving hand and static half antennas align.

Rotary

This interface relies upon a group of static tags with half antennas and a single moving half antenna that triggers a unique response when moved to pair with each static tag. In the sample implementation, the static tags are placed on the front of a single sheet of paper in a spoke-like design. On top of this, a single circle is attached that has a single half antenna placed on one radius. The circle can spin freely, rotating the half antenna on the circle relative to the static tags (Figure 4.5(g)). When the rotating antenna contacts a static tag, it provides enough power to backscatter signals to the reader for detection and triggers a set response.

Note that in addition to using the hand as a ground plane to enable touch interaction, one can also overlay a second antenna to complete the circuits as a dipole antenna, which can also trigger the same effect as hand touching. In this case, the state of the slider and rotary will stay in place even if the hand moves away.

4.5.3 Free Air Motion Primitives

PaperID also create tag-based interfaces that track motion signals over single and multiple tags. Below, I describe a free air system with a single tag moving in space. This primitive can be applied to a paper interface of a stand-alone RFID tag where PaperID provide functionality for fine-grain trace tracking of the interface relative to the RFID reader as well as frequency tracking for periodic movements. This primitive is supported by the trace tracking technique introduced in the Signal Processing section.

4.5.4 LED User Feedback

Most of the RFID tags described thus far are geared to user input. However, it is possible to construct an LED-enhanced RFID tag that harvests enough power for the RFID reader to flash an LED. Although not as complex and full-featured as prior work [137], this tag does not need a “loop IC” to communicate with the reader. Instead, the RFID reader can programmatically modulate its output

power from high to low to turn the LED on or off. The LED tag can also be made into a half tag such that the user completes the antenna and lights the LED when touched.

4.6 Applications

In this section, I present several applications that use combinations of the primitives described above to build new interactive paper interfaces that come to life when read by an RFID reader.

4.6.1 Polling Device

One of the most widely used test-taking and grading methods involves the Scantron test form. In this classic paper interface, users fill in bubbles with a number two pencil to denote answers to multiple choice test questions. By using the RFID half tag concept, PaperID re-imagine test taking as a real-time, interactive experience. In Figure 4.6.1, I demonstrate two methods to use RFID tags to create polling/multiple choice devices. First, users can indicate a choice among multiple tags by either covering a complete RFID tag (touch interaction) or by completing a half antenna with a finger (button interaction). Second, users can complete one of multiple half antennas by coloring in the end of the antenna with a conductive pen. In both cases, data can be collected and assessed in near real time. This method allows nearly instantaneous processing of users' opinions or knowledge without requiring that each user has access to an individual wired or otherwise powered system, resulting in lower cost.

4.6.2 Magic Wand/Conducting Baton

Besides augmenting a static 2D piece of paper with touch controls, the "free air" gestures described above can be used to transform a piece of paper into a tangible gesture interface. In this example, a commercially available RFID tag sticker is placed on a single sheet of paper that is then rolled up into a cylinder to form a magic wand, as shown in Figure 4.6.1. As a user waves the wand back and forth through the air, the RFID reader streams the RF channel parameters to the signal processing pipeline (on a host PC) to extract wand motion in terms of frequency content and velocity. These two metrics are then used to control the tempo and the volume of music played through the host PC. The effect is that the user can control the tempo of the music by the rate at which they swing the wand

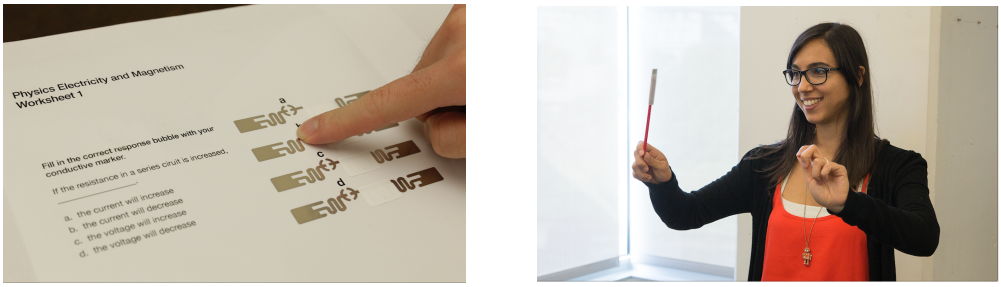


Figure 4.6: PaperID example applications left: In this polling application, a student can pick a response on a worksheet and receive live feedback. Right: The wand frequency is mapped to music tempo and the velocity is mapped to gain.

back and forth. To increase the volume of the music, the user increases the size of the gesture (i.e., the velocity it takes to complete one repetitive swinging motion). Thus, large, repetitive gestures make the music play louder and small, repetitive gestures make the music play softer.

4.6.3 Pinwheel

A corollary to the wand application is the pinwheel made of construction paper shown in Figure 4.7 left. For this application, A single RFID tag sticker is attached to a pinwheel. A user can blow on the pinwheel and use the velocity of the tag to control the animation flow of blowing particles with varying number and force, as shown in the supplementary video.

4.6.4 Pop-Up Books

Pop-up books offer a wide array of mechanical motions where pieces of paper are sliding across each other, bending, folding, and lifting. Using the RFID interaction primitives described above, these motions can inexpensively be instrumented to stream interaction events back to a computing device via the RFID reader.

For this application (Figure 4.7 right), I create an interactive, pop-up book page with a barn. First, An LED light circuit is created which turns on when the page is opened and the barn pops

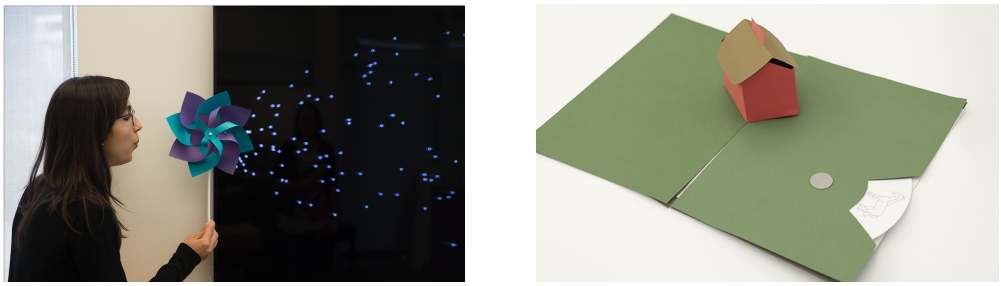


Figure 4.7: PaperID example applications (1) The speed of the spinning tag on the pinwheel is mapped to onscreen graphics; (2) This pop-up book contains a knob interactor which selects audio tracks and a shielded tag in the pop-up barn.

out. Next, a single tag is placed on an exterior wall of the barn that is blocked by a piece of copper tape when the page is closed. Opening the page stops the copper from blocking the RF signals so that the tag can be read, which in turn triggers the host PC to play the song 'Old McDonald Had a Farm'. Second, I present a rotary dial that allows the user to trigger various sounds. By aligning the customized button tag on the dial with the half antenna on the paper underneath, the RFID reader can determine the state of the rotary and prompt the sound that corresponds with that particular state. In the pop-up book example, the rotatory dial is covered with images of barnyard animals. Thus, when the child rotates the wheel to an image of a sheep, a 'baa' sound is played.

4.6.5 Dollhouse

In the dollhouse example shown in Figure 4.8, both the button and the slider are used. When a user presses the button, the house emits a doorbell sound. For the slider, there a set of stationary half tags connected to LEDs in each room in the house. As a finger is dragged across the stationary tags, it completes each tag antenna to adjust the brightness of the lights.

4.6.6 Interactive Light Control

cover, swipe, and wave interactors are combined into a single RFID tag to create a control system for a desk lamp with a Phillips Hue light bulb that can be controlled remotely by Bluetooth® and

Figure 4.8: PaperID example applications: paper interfaces can be used to create new sensors to control the lighting and doorbell of this foam doll house.

offers a full-color spectrum at varying levels of brightness (Figure 4.9). I use the cover action to turn the light on and off. The touch interaction changes the color of the light bulb across the light spectrum, and the swipe and wave gestures control the brightness.

4.6.7 RFID Powered LED

Although RFID is primarily designed for communication, it also transmits RF power that is harvested by the RFID tag for operation. This mechanism of wireless power transfer can be hijacked and used to wirelessly power LEDs. In the example application shown in Figure 4.10, an image of a cartoon skull is augmented with two LED enabled RFID tags in the eyes, and one hand-drawn stub antenna for a button tag in the mouth. When the user touches the button tag. the RFID reader is commanded to increase its output power level and cause enough power to be harvested by the two LEDs for them to turn on, creating the effect of shining eyes in the paper skull.

4.7 Evaluation

In this section, I evaluate the interaction methods and primitives used in the applications. One factor influencing performance is the reflected signal strength. To control this factor, the antenna power is

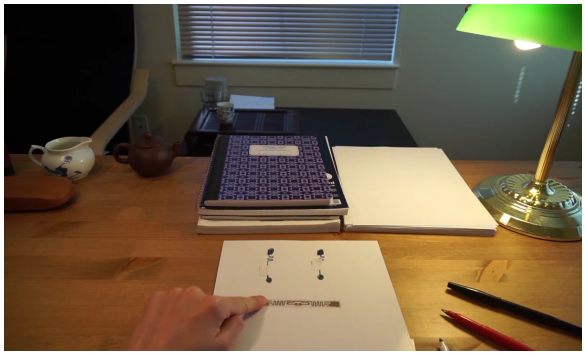


Figure 4.9: PaperID example application: light controller using a single RFID sticker

automatically calibrated so that RSSI is at approximately -25dBm.

4.7.1 Gesture Sensing Evaluation

Interaction primitives including wave (Figure 4.5(a)), swipe (Figure 4.5(b)), touch (Figure 4.5(c)) and cover (Figure 4.5(d)) are integrated into a tag gesture-sensing application using features and classification methods described in the Signal Detection Section. The training data was recorded at 3 locations (2m, 4m, and 6m away from the reader). The classifier has 5 classes, including the 4 gestures and the still state when no interaction is performed. 30 instances from each class are recorded, 10 at each of the 3 locations. Training is done offline by optimizing parameters in the RBF kernel maximizing the 10-fold cross validation results of gesture classification. 5 participants were recruited (4 male, 1 female) with a mean age of 24.4 years. Each participant was asked to follow visual instructions on a monitor to perform 20 instances of each of the 5 gesture classes. Each participant repeated the task 3 times at each of the 3 locations (2m, 4m, and 6m away from the antenna). Visual gesture instructions were given once every 4 seconds, and pauses between gestures were utilized to generate still instances. The instruction script was used as ground truth and user mistakes were manually annotated. The classifier achieved an average of 94.1% accuracy classifying gestures into 5 categories with a 2.4% STD across participants and 1.4% STD across locations. This study result demonstrates the high performance of the system across all locations

Figure 4.10: PaperID example applications #2: (1) The intensity and color of light can be altered using wave and touch gestures; (2): LEDs are wirelessly powered using the same antenna for input sensing, A single touch tag is used as the input to increase or decrease the reader's power output.

with low variations, showcasing the promising potential for deploying the technique for real-world user interfaces.

4.7.2 Customized Touch Button Evaluation

The effectiveness of the customized touch button design is evaluated in this study for touch vs no-touch states using a read rate threshold of 5 reads/per second. 10 hand-drawn touch buttons are arrayed in a row on a sheet of paper. Note that there are 3 ways to fabricate the button antenna: hand-drawn, stencil-traced, and inject printing. The most challenging hand-drawn ones are evaluated to establish a performance baseline. 5 participants were recruited (3 male, 2 female) with a mean age of 26.0 years. Each participant was asked to follow visual instructions on a monitor to perform 10 touches on each of the 10 tags, one tag at a time. The study was repeated 3 times at 3 different locations (2m, 4m, and 6m away from the reader). Instructions were given once every 2 seconds. The instruction script was used as ground truth and user mistakes were manually annotated. Pauses between touches were utilized to generate no-touch instances. This approach achieved an average of 96.2% accuracy on detecting touch events, with a 1.1% STD across participants and a 0.4% STD

across locations. Because the Knob and Slider primitives were designed based on customized touch button elements, their performance will be similar to what's shown in this evaluation.

It is worth noting that the number of squiggles in the antenna tag design will have an impact on the feasible operating rage of the tags. When there are too many squiggles, the tags will automatically turn on when being close (e.g. 2m away) form the reader, making it challenging to detect finger touches. On the other hand, if there are too little squiggles, it will be hard to sense the touch event when the tag is far away from the reader antenna. During the experimentation, having a squiggle number of 5 as demonstrated in the figure 4.1 will have the best balance between long and short ranges. Other factors such as conductivity of the ink, as well as the size of the top rectangle area of the antenna design will also affect the tag performances at different locations. The investigation of these factors are beyond the scope of this work, but would be interesting to explore through RF stimulation and experimentation methods.

4.7.3 Velocity Metric Evaluation

The magic wand described in the Applications section is utilized to evaluate the Trace Recovering technique. A tagged wand was moved between 2 points with a separation of 20 centimeters back and forth at 1, 2, 4, 6 Hz per second from 2, 4, and 6 meters away from the antenna, each for 20 seconds. Velocity is calculated by accumulated displacement divided by time difference and frequency is estimated by the most significant frequency component after the Fast Fourier Transform is performed. These two metrics are calculated in real time using window length of 2 seconds and updated every 1 second. Study results demonstrated that velocity sensing had an average standard error of 6.4% and frequency sensing had an average standard error of 8.2%.

4.8 Conclusion

In this chapter, I demonstrated techniques to create simple, lightweight, recyclable paper interfaces that rely on RFID technology. Using a single RFID reader antenna and inexpensive passive tags, PaperID can create applications with numerous controllable features, including lights, sounds, software control, and animations. Subtle changes in the low-level radio frequency signals that typically retrieve tag identification information are leveraged to determine how each tag is manipulated by

the user. Additionally, I introduce multiple techniques for creating RFID tag antennas in order to customize tags for different contexts. This methodology expands the functionality of RFID tags to quickly and easily create user interfaces on paper, a ubiquitous, lightweight medium.

Chapter 5

SENSOR FUSION FOR NATURAL HUMAN ROBOTIC INTERACTIONS

5.1 Introduction

In this chapter, I present ID-Match [85], a hybrid computer vision and RFID system that recovers the relative motion paths of an RFID tags worn by people and correlate that to physical motion paths of individuals as measured with a 3D depth camera. This allow autonomous robots and computer systems to quickly recognize and interact with individuals in a group setting has the potential to enable a wide range of personalized experiences. Results show that the real-time system is capable of simultaneously recognizing and correctly assigning IDs to individuals within 4 seconds with 96.6% accuracy and groups of ve people in 7 seconds with 95% accuracy. In order to test the effectiveness of this approach in realistic scenarios, groups of ve participants play an interactive quiz game with an autonomous robot, resulting in an ID assignment accuracy of 93.3%

5.2 System Overview

ID-Match is a real-time, hybrid computer vision and UHF RFID system that can simultaneously track, and individually identify multiple people wearing RFID tags in multipath rich, real world environments.

Figure 5.1 shows a typical implementation of the ID-Match system consisting of an Impinj Speedway Revolution UHF RFID reader [65] with a single antenna [64], along with a Kinect v2 depth camera [98]. Users wear low cost (7-15 cent each) UHF RFID tags in the form of clip-on name badges or lanyard. As people walk within view of the system the RFID tags are continuously read and the Kinect tracks the position of the unknown skeletons in 3D space. Since passive UHF RFID tags have a read range of up to 10 meters and a single reader can cover 50-150 square meters, the challenge is to determine which ID belongs to which skeleton.

ID-Match accomplishes this using two independent correlation pipelines, which are combined to provide a fast and accurate determination of the precise location and ID of individuals. The

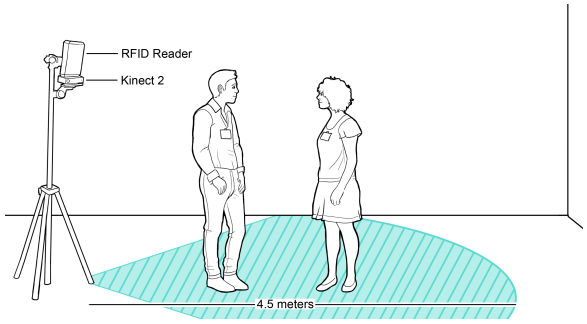


Figure 5.1: An illustration of a typical implementation of the ID-Match system consisting of a Kinect depth camera and UHF RFID reader.

procedure is outlined below and is covered in greater detail in subsequent sections.

Reverse Synthetic Aperture Pipeline:

1. Use a reverse synthetic aperture technique to determine the relative radial path of each of the RFID tags to the RFID reader's antenna
2. Use the Kinect to track each person in 3D and determine his or her equivalent radial path in the coordinate frame of the RFID reader antenna
3. Compare and rank each of the tags synthetic aperture paths to the visual motion paths and place in a voting buffer

RF Motion SVM Pipeline:

1. Record the RSSI, phase, and channel number for each tag and extract seven RF motion features
2. Track the location of each skeleton using the Kinect and extract three motion features
3. Use SVM machine learning to classify similar RF motion features to physical motion features and place results in a voting buffer

Final Step:

1. Combined the two independent voting buffers and determine the final tag ID to skeleton correspondence

5.3 Reverse Synthetic Aperture

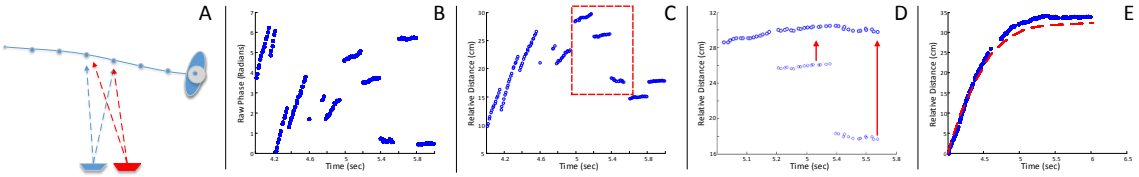


Figure 5.2: ID-Match: overview of the signal processing algorithm used to associate RFID SAR data to Kinect motion data

Synthetic Aperture Radar (SAR) techniques have been widely used to increase the imaging resolution of radar systems [26] and the localization accuracy of both radio transponders [89, 176] as well as UHF RFID tags [177, 101]. This is accomplished by moving the objective antenna (i.e. reader antenna) along a known trajectory at a constant velocity, to synthesize a large array of antennas. Localization accuracy is increased since the objective antenna can take multiple samples of a stationary target from different angles.

In contrast, the goal of the ID-Match system is to locate and identify people wearing RFID tags as they naturally walk and play in their environment. This breaks traditional SAR approaches since both the objective antenna and the target object would be simultaneously moving at unknown velocities. This would normally result in blurry radar images and large localization errors.

In this work the SAR paradigm is reversed as shown in Figure 5.2, panel A. Here the RFID reader is placed in a fixed location and the motion of the RFID tag (when worn by a person moving) creates a synthetic aperture. Since the exact trajectory of the people wearing tags is not known (i.e. they are 'free roaming') it is not possible to directly compute the exact location of the tags. Panel A shows a motion path of a person wearing and RFID tag while being continuously measured by the ID-Match system (top view). Panel B shows a two-second snapshot of the raw RFID phase data as a function of time. In panel C, the phase data has been converted to the radial distance from the antenna. In Panel D the phase discontinuities cause by frequency hopping are "stitched" back

together. Finally, panel E shows the RFID SAR trajectory (blue circles) compared to the radial motion of the person as measured by the Kinect (dashed red line)

One of the key attributes of UHF RFID systems is that the tags do not actively generate and transmit radio waves. Instead the tags back-scatter (i.e. reflect) the RFID reader's carrier wave back to the reader in order to send data. This unique feature of the UHF RFID physical layer means that in a single read event, the RFID reader can precisely measure the phase angle between its transmitted signal and the reflected signal received from the tag. Figure 5.2, Panel B shows a two-second snapshot of the raw phase measurement as a function of time, for a person wearing an RFID tag walking towards the ID-Match system. The discontinuities in the reported phase are partially due to 2π radian wrapping as can be seen at 4.2 seconds. An additional source of error is due to an unknown phase offset introduced when the reader pseudo-randomly frequency hops from one carrier to another as required by government regulations [34]. This manifests in panel B as a grouping in the phase data. Where each group consists of a series of data points read consecutively at the same frequency.

Using equation 5.1, it is straightforward to calculate the relative radial distance from the tag to the reader, where ϕ is the phase angle reported by the RFID reader, f is the frequency of the RF carrier, and c is the speed of light.

$$Distance(relative) = \frac{\phi c}{4\pi f}; 0 < \phi < 2\pi \tag{5.1}$$

Figure 5.2, Panel C shows the phase data converted to distance. However, since the phase angle between the transmitted and reflected signal will rotate 2π radians for every λ wavelength, it is not possible to calculate the absolute distance, but instead rather the relative distance. Given the inertia of a person walking, and the high sampling rate of the reader, it is reasonable to assume that large discontinuities in distance are not due to human behavior, but rather an artifact of the RFID reader frequency hopping. Thus, given the channel number reported by the RFID reader it is possible to "stitch" the disconnected groups back together. This is shown in Figure 5.2 panel D, where the slope of the trailing points of one group are aligned with the slope of the leading points of the next group.

Working in conjunction with the RFID reader the ID-Match system uses a Kinect 3D depth camera to track the location of people with in its frame of view. In this work people wear RFID

tags on their torso, as either name tags or lanyards. By computing the distance of the person's spine (as measured by the Kinect) to the RFID reader antenna, it is possible to correlate the physical path of the person, to the RFID tag's SAR path. It should be noted that due to phase wrapping, there is an equivalent distance wrapping, and the absolute distance of the tag to the reader cannot be determined. Thus for a given time window both the Kinect distance data and RFID SAR distance data are aligned at the beginning. The final result is shown in Figure 5.2, Panel E, which shows good agreement between the Kinect (red dashed line) and RFID reader (blue circles).

When multiple people are walking in front of the ID-Match system it will result in multiple RFID SAR paths and multiple Kinect motion paths. By taking the standard error between the sets of paths it is possible to determine which path belongs to which person. Details on the ranking algorithm are presented in the ID Association section.

5.4 Probabilistic Correlation

In addition to the synthetic aperture approach described above it is possible to use low-level channel parameters along with machine learning techniques to determine the correlation between bodies in the view of the Kinect and tags within the view of the RFID reader. The key idea is that each time the RFID reader interrogates a tag, it reports channel parameters such as Received Signal Strength Indicator (RSSI), RF Phase, and Doppler shift, along with the channel number, representing a unique signature of the RF environment of that individual tag. By observing the changes in these channel parameters over time, it is possible to correlate the motion signature of a tag, with the motion of the person wearing it.

In Chapter 3 I discussed using RFID channel parameters including RSSI, RF Phase, and tag Read Rate to classify a tag as either being still, moving, cover by a hand, or swipe touch. Building upon this approach new features have been specifically designed for the task of correlating tag motion to the motion of people as measured by a depth camera.

5.4.1 RFID Phase Features

Note that all features are calculated using RF channel parameters within each segments. The first two RF phase features are based on the tags velocity using equation 5.2, which is calculated from

consecutive tag reads on the same channel [108].

$$v_r = \frac{\lambda(\theta_1 - \theta_2)}{4\pi(t_1 - t_2)} \tag{5.2}$$

1. Radial Velocity: the average of v_r.
2. Absolute Radial velocity: the average of the absolute v_r.
3. Standard error of the simple linear regression of unwrapped phase versus channel frequencies. $stderror(linearfit(channel, unwrap(phase)))$
4. Average of phase change divided by frequency change when carrier signal frequency hops. $\sum_{i=1}^{k-1}((phase(i+1) - phase(i))/(frequency(i+1) - frequency(i)))/(k-1)$

5.4.2 RFID RSSI Features

Change in distance will create variations in the RSSI signal which is employed in the following three features.

1. Average RSSI standard deviation within each channel
2. Average RSSI difference between consecutive samples within each channel
3. Absolute value of the RSSI

5.4.3 RFID Read Rate

The read rate of a given tag is primarily correlated to the amount of RF power it can receive. This typically means that RFID tags with a low read rate are on the edge of the read zone which is well outside the field of view of the Kinect. Thus, this feature is mainly employed to exclude tags that are unlikely to be a viable candidate for ID to body matching.

1. Read rate: number of packets received from each RFID tag per second

5.4.4 Kinect Motion Features

Both RSSI and phase are sensitive to movement in the radial direction to the reader. Thus, the position of the spines on the skeletons reported by the Kinect are transformed from the Kinect's Cartesian coordinate system to a polar coordinates centered at the reader. Once the skeleton tracking is started, the following three features are utilized.

1. Radial component of the skeleton's velocity relative to the RFID reader antenna

2. Azimuth (i.e. non-radial) component of the skeleton's velocity relative to the RFID reader antenna

3. Distance between the skeleton's spine and the RFID reader antenna

5.4.5 Data Segmentation & Machine Learning

A Support Vector Machine (SVM) classifier with the Radial Basis Function (RBF) kernel was trained by recruiting three participants to wear RFID tags while walking freely in the view of the depth camera and the reader (ground truth was taken manually). The data stream was segmented with a window size of 2/3 second which was advanced every 1/3 second resulting in a 50% overlap. This achieved a good balance between matching accuracy and latency. The classifier was only trained to two prediction classes, "Match" and "Not-a-Match". The parameters of the SVM classifier where optimized by maximizing the 10 fold cross validation results.

Since the system is trained for large movements that are generic to a majority of people, and since the machine learning features are differential and/or relative to the reader antenna, training does not have to be redone for new environments or people. In fact, training was done on one set of participants at one location, while all testing was done at several other locations with many other participants over the course of several weeks. Finally, In the following section, the output of the SVM classifier is combined with the results from the SAR method to provide final decisions on body to ID matches.

5.5 Matching Algorithm

The true strength of the ID-Match system is that ID association is based on similarity ranking between a finite number of possibilities, rather than the raw ability of the RFID reader and/or the Kinect to precisely locate tags and people. This is accomplished by measuring small variations in tag and body motion of people as they walk in order to differentiate between them. While these motion difference can be hard to visualize, they are statistically distinct. In order to illustrate how the ID association algorithm works I present a 'toy example' of five people walking in a single file line, towards the RFID reader and Kinect mounted on a tripod. While this does not represent a real usage scenario (which is presented in detail in subsequent sections) for the sake of an example, it simplifies the motion the people and ensures that there are no blocking events where one person is visually occluded by another.

Figure 5.3 panel A, shows the distance traces for five participants (i.e. Persons 1-5) walking toward the system as measured with the Kinect. Panel B shows the relative distance of the RFID tags (Tags 1-5) as worn by the five people over the same period of time. Since the RFID reader can detect the tags at a range of 7-8 meters they are visible to the system before the Kinects depth camera can track them (which occurs at 4.5 meters).

The ID-Match system is implemented in Matlab and is capable of assigning IDs to people in real time. This is accomplished by applying a 2/3 second sliding window (which is advanced every 1/3 seconds) to the incoming data stream from the Kinect camera and RFID reader. During each time segment the system analyzes the buffered data and applies the SAR and SVM techniques previously described.

In the case of the SAR approach, the standard error between each of the five SAR tag traces is computed across each of the five participants position traces. An example can be seen in Figure 5.3 panel C which shows an expanded view of a 2/3 second time slice from panels A and B. Here it can be seen that 'Tag 1' (red circles) is the best fit for 'Person 1' (dashed red line), which will result in the lowest standard error and highest confidence. Thus, every 1/3 seconds the system outputs new matching events and places them in the SAR results buffer for each person seen by the Kinect. It should be noted that SAR data is only considered valid for people in motion.

During the same time window, the 8 RF features for the five RFID tags and 3 Kinect features

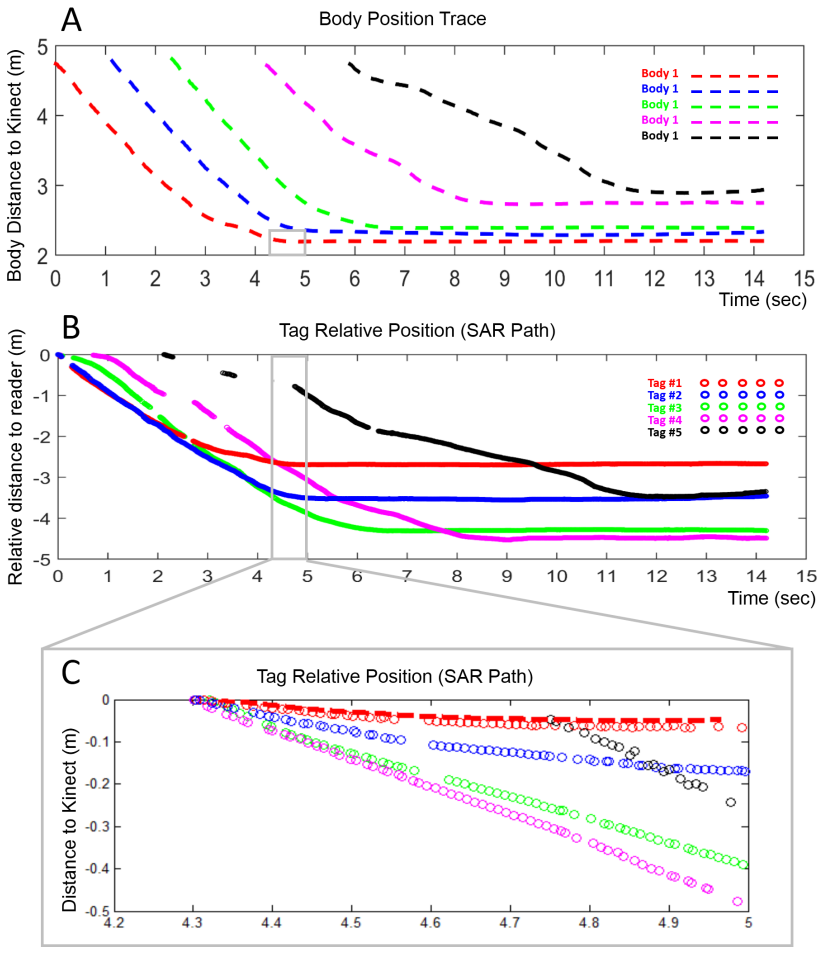


Figure 5.3: ID-Match: correlation body motion trajectories to tag motion trajectories

for the five bodies are passed to a trained SVM classifier which generates a probability estimation of each Tag and Person pair. For each individual body, if a given RFID tag has a probability with a margin of 30% or more when compared to all the other RFID tags, it is considered a matching event and the SVM results is placed in a result buffer.

Each of the five bodies is assigned their own final results buffer consisting of a FIFO of the last 10 prediction points from the SAR and SVM predictor. The arithmetic mode is taken as the final matching result, and the ID is assigned to the body. Since the prediction FIFO for each body seen by the system can be updated over time, misidentifications can be corrected for. Once sufficient confidence is obtained, the identity of the body (i.e. tag / body pair) is "locked in".

5.6 Performance and Evaluation

In this section ID-Match is evaluated under a number of controlled scenarios so that the underlining behavior of the system can be better understood and quantified. In particular, I evaluate the ability of the system to: recognize individuals in groups of people, distinguish between multiple tags undergoing the same motion, and reacquire people after the identity match has been lost.

5.6.1 Recognition of Individuals In Groups

When multiple people, walking as a group, approach the ID-Match system. This creates a number of dynamic events where people in the foreground can visually occlude (and possibly RF occlude) people in the background. Additionally, since the RFID SAR motion paths are a relative measure of distance and not a trace of position over time, there is the potential that the paths may not be unique, causing confusion in the matching algorithm.

In this study, two groups of five participants wore RFID tags on their upper torso. They were asked to start outside of the Kinect's field of view (approximately 6-7 meters away from the system), and walk as a group up to the ID-Match system. They were instructed to stop in one of five general locations such that upon arrival in front of the ID-Match system they would be standing *five a breast* and could visually see the system. Figure 5.4 panel a shows the final location of one set of participants of one trial as seen by the Kinect, and panel b shows the path they took as reported by the Kinect. For ground truth each person was assigned a "Body Number" (1-5) and the corresponding

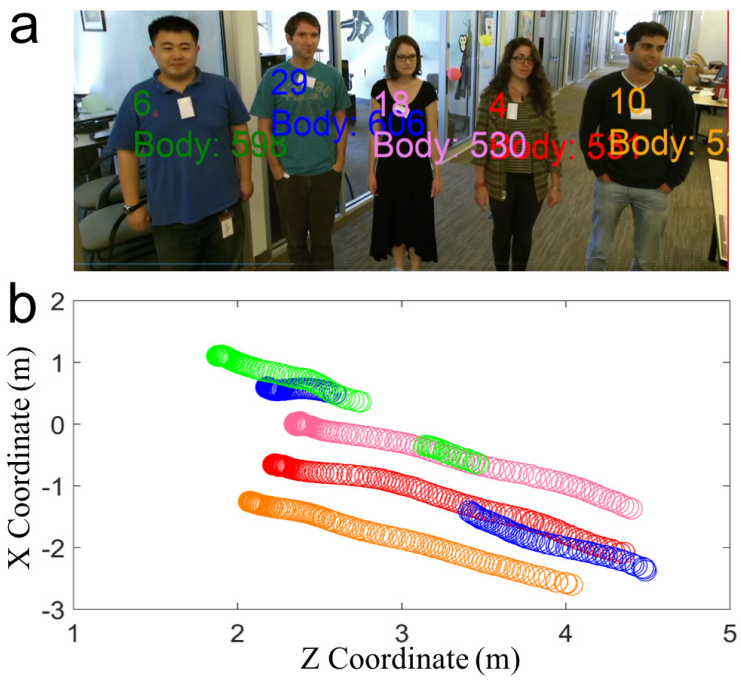


Figure 5.4: ID-Match: panel A shows the RGB image captured by the Kinect as five people walk up to the system. Overlaid on the image is the RFID tag ID that has been automatically associated with that person. Panel B shows a top view diagram of the paths of the participants as recorded by the Kinect.

RFID ID (1-5) was also recorded. To emulate a more real world scenario and to ensure the problem space was not unduly constrained, two dummy tags were also introduced into the view of the reader (IDs 6 & 7). Each group of participants was asked to repeat the experiment 6 times for a total of 60 potential matching events. The two experiments are folded together into the truth table shown in table 5.1.

These results show and overall ID association accuracy of 95%. There were two trials where the Kinect was not able to see one of the participants as indicated in the two columns to the right in table 5.1. Excluding this data point results in an accuracy of 98.3% when both the Kinect and RFID reader can view all people. In one trial an *ID* was not assigned to a *body* leaving body 5 “pending”, meaning the confidence threshold for a match has not been met. Observations of the trials suggest

		RFID Tags Visible to the RFID Reader							Error States	
	Assigned to ➔	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6 (dummy)	ID 7 (dummy)	Pending (not matched)	Not Seen by Kinect
People Wearing Tags	Body 1	11	0	0	0	0	0	0	0	1
	Body 2	0	11	0	0	0	0	0	0	1
	Body 3	0	0	12	0	0	0	0	0	0
	Body 4	0	0	0	12	0	0	0	0	0
	Body 5	0	0	0	0	10	0	0	1	0

*Ground Truth: Body 1 = ID 1, Body 2 = ID 2, Body 3 = ID 3, Body 4 = ID 4, Body 5 = ID 5

Table 5.1: A confusion matrix showing the results for assigning IDs to five participants as they approach the ID-Match system.

the errors are due to occlusions.

One of the important criteria of the proposed system is that it should be able to quickly identify and recognize individuals such that an autonomous robot can properly interact with them in a timely manor. However, for the ID-Match system calculating the exact acquisition time for groups of five people is a multi-dimensional problem consisting of (but not limited to); the walking rate of individuals, the total time all participants take to arrive at the destination zone, the fact the some people will enter the Kinect's depth field of view before others, the issue of people dynamically blocking the view of the Kinect, and the resulting need for target reacquisition and ID association.

Instead of quantitatively addressing each one of these variables individually the following plot qualitatively shows that, given all the above variables, ID-Match is indeed able to identify individuals in a timely manor. The red line in Figure 5.5 shows the accuracy of matching the correct *ID* to an individual as a function of time. Since there is variability in the time of arrival and walking speed, the data has been normalized to the point in time when each participant first walked into the view of the Kinect's depth camera. Results show that ID-Match is capable of correctly assigning IDs to individuals within 4 seconds with 96.6% accuracy.

The second blue line in the plot shows the accuracy of assigning the correct ID to all five participants in the group as a function of time. The reason this plot takes longer to converge is that the system must wait for all participants (regardless of how fast they walk) to enter the field of view. Results show that a group of five participants can be correctly matched within 7 seconds with 95% accuracy. It should be noted the line for the individual acquisition time has a higher accuracy than

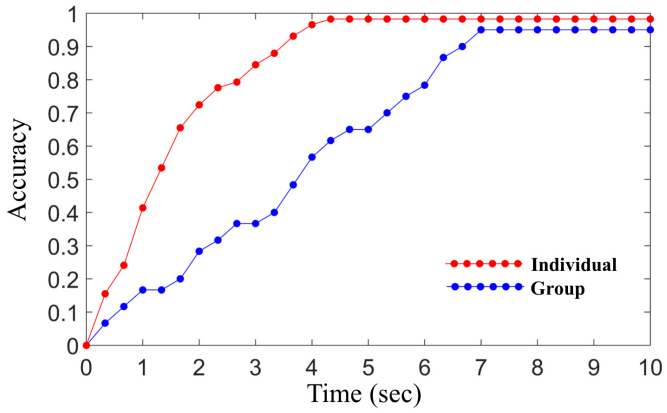


Figure 5.5: ID-Match prediction accuracy of the system overtime

the group case because the two people not seen by the Kinect are excluded.

A detailed examination of the SAR and SVM voting buffers for the above data shows the predictive power of each technique individually. For the data shown in Figure 5.5, at t=4sec for each individual, the SVM classifier has an accuracy of 70.1% and the SAR approach has an accuracy of 89.4%, while the combination of the two results in a total of 96.6% accuracy.

5.6.2 Disambiguating Similar User Motion

One important question is how unique does the motions of the users and the paths they take, have to be for the ID Association algorithm to work properly. One concern is that groups of people walking in *lock step* may have nearly identical RFID SAR paths. Since the measurements RFID SAR of relative distance and not absolute distance it may not be possible to distinguish them from each other.

Given the complexity of precisely and repeatedly controlling the motion of five people, an alternative and arguably more stringent test has been implemented. Figure 5.6, panel A shows an image of five RFID tags (in the form of name tags) attached to a wooden rod. They are spaced evenly approximately 40 cm apart to mimic five people standing shoulder to shoulder. For this study ten

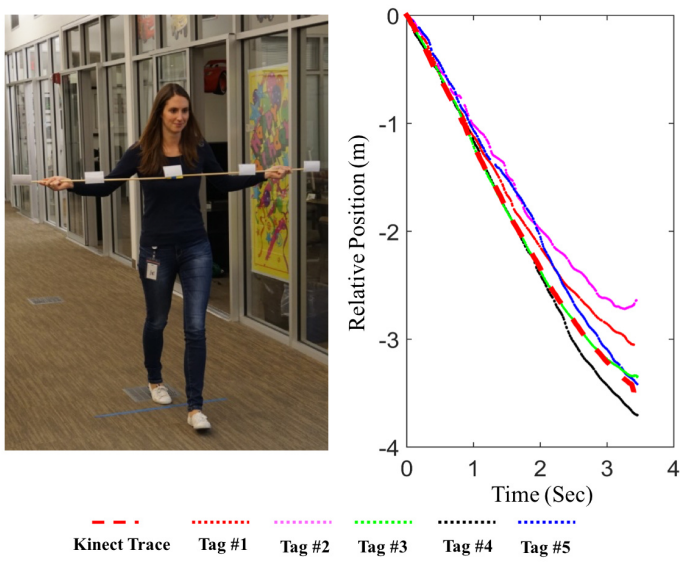


Figure 5.6: ID-Match experimentation: a person carrying a horizontal rod with five RFID tags mounted on it and their corresponding motion traces

participants were recruited and each person was asked to hold the wooden rod such that one of the tags was place firmly against their chest as if they were wearing the tag. Starting outside the field of view of the Kinect's depth camera the participants were ask to walk towards the ID-Match system. This procedure was repeated five times such that each participant held each tag on the wooden rod against his or her chest once, which resulted in a total of 50 runs. Again two dummy tags were placed in the environment to insure a more realistic experiment.

The results in Figure 5.7 show that ID-Match is capable of correctly matching the tag *ID* to the corresponding *body* position on the wooden rod with an accuracy of 96.7%. The probability of randomly guessing the correct answer for one instance is 1-out-of-7. Since each of the 60 trials was done by a single person (rather than a group) there were no occlusions, and the Kinects depth camera was able to track each person without errors. These results show that ID-Match is able to take advantage of the slight difference in radial direction of parallel movement to successfully assign the correct *ID* to the corresponding *body*.

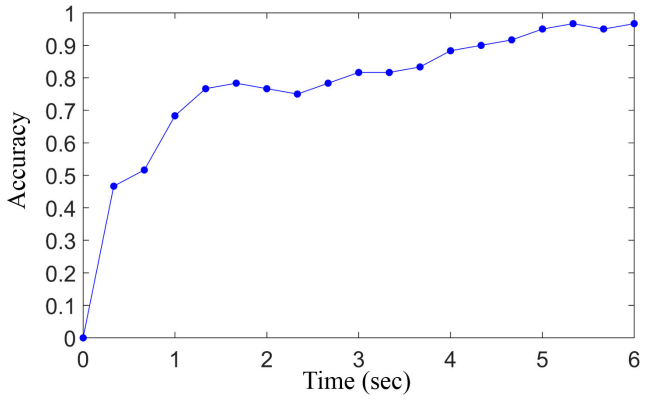


Figure 5.7: ID-Match: the accuracy of the system at matching the correct tag to the skeleton in the disambiguating similar motion evaluation. The blue line shows and average accuracy of 96.7% within 6 seconds, for the 60 trial.

5.6.3 Match Reacquisition

The Kinect depth camera has proven to be reasonably reliable at detecting the presence of a person, determining the position of their body and limbs (i.e. skeleton), and tracking the location of the person while in the field of view. However, all computer vision systems suffer from occlusions when the camera's view of the target person is blocked by an object or another person. In the case of the Kinect this means that when a person is occluded or goes out of view, their virtual skeleton (and corresponding skeleton ID) is discarded. Once the person reenters the field of view, they are given a new skeleton and skeleton ID. This means that each time the Kinect loses a person and they reenter the field of view the ID-Match system must reacquisition the person and perform ID association to determine who they are and where they are.

In order to investigate how effectively ID-Match re-acquisitions tagged people once they have been lost by the Kinect's depth camera, the following experiment has been devised. Here three participants have been asked to walk in a circle marked on the floor with a diameter of 3 meters. As they walk around the perimeter of the circle the person in the foreground as seen by the Kinect blocks the people in the background. Thus, each person is potentially occluded twice per revolution. The three participants are asked to walk in a circle for 5 minutes resulting in 86 blocking and

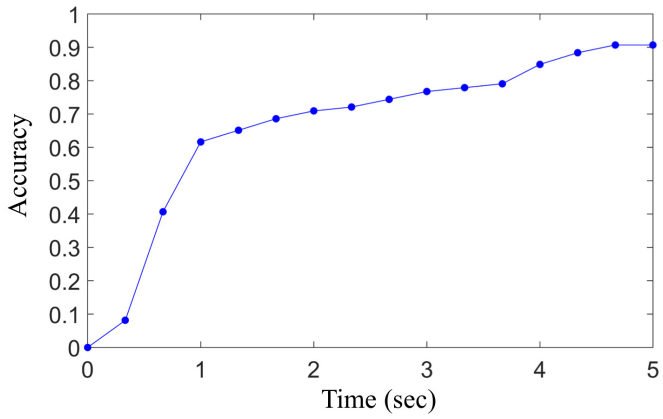


Figure 5.8: ID-Match accumulated accuracy overtime for re-acquisitioning a person after visual tracking is lost.

reacquisition events. Results for the percent accuracy for matching the correct ID to the correct person is shown in Figure 5.8. In this case the available time was only approximately five seconds (i.e. the time between a person becoming visible and then blocked again). This test shows that ID-Match performs quite well at reacquiring people once they have been lost by the Kinect, with a percent accuracy of 90.7% over the limited time interval of 5 seconds.

5.7 HRI and Occupancy Tracking Studies

In an environment where the system is supposed to capture which users are passing a gate, users will enter, continue moving, and leave the field of view of the system at different speeds and at different times, with no constraints on how long they would stay in the field of view.

In the following sections, I present two different scenarios to evaluate the accuracy of ID-Match with less control on the behavior of the users.

5.7.1 Human Robot Interaction Study

In this application, the ID-Match system has been integrated into *Furhat* [4], an interactive anthropomorphic robotic head with an rear-projected 3D face. Computer animation delivers dynamic facial movements such as gaze (eye movement), facial expressions (happy, sad, confused, surprised

etc.) and lip animation of basic phonemes which are synchronized to computer generated speech. Additionally *Furhat* is equipped with a pan-tilt neck which along with the animated eyes allows for convincing gaze simulation, and importantly for this work, allows the robot head to turn to address an individual person while looking at them in the eyes.

In this experiment *Furhat* has been programmed to autonomously host a multi-player quiz game [5]. In this interaction, users in groups freely approach the robot to compete in a gaming interaction scenario where the robot asks players different trivia questions and the users collect points when giving the correct answer. The interaction is constructed such that the robot needs to use the identity of the user and address the user by name (e.g. "Jack, here is the next question: what is..."). For the purposes of this game contestants were asked multiple choice questions and a commercially available speech recognition engine was used to capture users verbal answers, which were limited to "one", "two", "three", or "four".

The interaction was setup up in a large office room (4 x 6 square meters), with the robotic head placed on a pedestal in one of the corners, as shown in Figure 5.9. The RFID antenna and the depth camera was placed to maximize the field of view of the ID-Match systems and the coordinate frame of the Kinect was transformed to the robot's coordinate frame to account for the offset in location and pose between the two. Ten participants were recruited to interact with the robot in two groups of five users each. Seven of the users were male and three were female. All participants were graduate level students between 20 and 30 years old. Groups of five users carrying pre-registered RFID tags were instructed to enter the room and playing and quiz game with the robot. In order to allow the interaction to be natural, the participants were not told to stand in any predefined locations, order, or distance from the robot. However, the participants were asked to avoid standing directly in front of another person so as not to block the view of the other participants in the game. Each of the two groups of participants played the quiz game 6 times, giving a total number of 12 games, and 60 interactions (questions). Each of the 12 sessions lasted approximately 5 minutes.

Although the interaction was complex in terms of robot and task design, the ID-Match system task was simple: each time a new skeleton is tracked by the camera the system will attempt to match that skeleton to one of the RFID tags out of all the tags currently visible to the RFID reader. Figure 5.10 shows a plot of the average accumulated ID to human assignment accuracy of the system over time.

Figure 5.9: ID-Match application: a snapshot of the Human-Robot Interaction setup showing 5 participants in front of the robot playing a quiz game.

The "individual" (red line) shows the accuracy of the system at identifying each individual people approaching the robot in a group. It is important to mention here that at any moment in time, there was always two additional static tags in the environment that were not held by a person. The results show that the system is able to recognize each individual at 94.8% accuracy within 5 seconds starting from when the person enters the scene. This is consistent with the controlled studies presented earlier. This fast acquisition time gave the robot the ability to customize the interaction, and address the individual person by name, almost by the time they completed their approach and came to a stop by the robot. The"group" blue line in Figure 5.10 shows the accuracy of the system in identifying all individuals in the five-person group with time starting at the moment when the first person enters the scene. Here the system is capable of accurately identifying all individuals with 91.6% accuracy in 12 seconds. Although this group acquisition time is highly dependent on actions and speed of the participants, it is important to include this result as a benchmark for a group of people naturally approaches the robot for interactions.

5.7.2 Occupancy Monitoring Scenario

The second uncontrolled user study evaluated ID-Match's ability to passively monitor the flow of people in an office as they pass through a "virtual gates" and checkpoints. The systems task is to quickly recognize the identity of the passersby while they are visible to the camera and determine

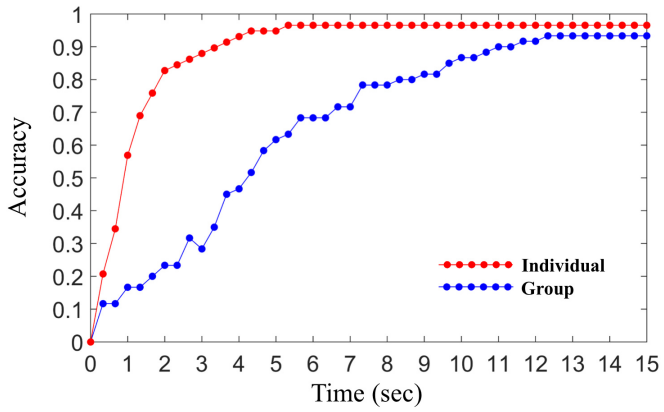


Figure 5.10: ID-Match evaluation: a plot showing the accuracy of the system over time

in which direction they are moving over a long period of time.

Physical Space

The study was conducted in an office setting as shown in Figure 5.11 consisting of a central hallway with offices one side and workstations in an open floor plan on the other side. The ID-Match system was configured to log data for a full working day (i.e. continuously for 7.5 hours). Out of 40 employees on site during that day, 16 were assigned RFID tags and registered in the system, the other employees had no other RFID tags. In the same area where the system was setup, two additional static RFID tags were placed and were visible to the reader at all times. Additionally, 3 of the 16 employees with RFID tags had workstations within the read range of the RFID antennas, being visible to the reader at almost all times, but not visible to the camera. This causes a greater level of uncertainty when attempting to match the IDs to people. The Kinect camera was set up pointing towards the hallway with a 30-degree side view as demonstrated in 5.11, panel A. In order to increase system reliability a second RFID antenna was used to insure tags were properly read as participants either walked towards and away from the Kinect.

As mentioned earlier, 40 participants took part in the experiment, 16 of them were carrying a registered RFID tag. The environment and the users were not controlled during the day. The participants were not informed about the purpose of the study in order to maintain their usual behavior.

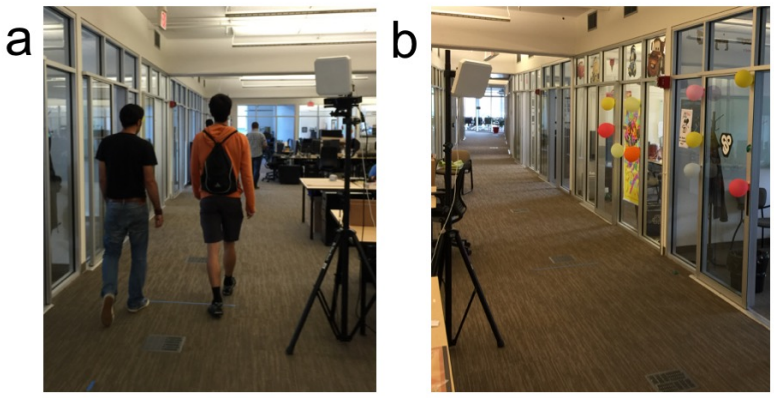


Figure 5.11: ID-Match physical setup of the occupancy study.

The only requirement placed on the participants was to wear the tags around their chest position, at the time they entered the office till the time they left for the day.

During the day-long experiment, 302 instances of people passed through the hallway where the experiment was set up. 129 of these were people who carried an RFID tag, and 173 did not carry a tag. The system's task was to recognize the identity of any visible skeleton, whether it belongs to one of the registered users, or whether that skeleton did not have a registered tag in the system. Out of the 129 instances of people wearing tags, the ID-Match system was able to correctly recognize 122 users (at an accuracy of 94.5%). For people not wearing tags, the system incorrectly assigned 5 identifications, yielding a 2.9% false positive rate, while the rest was correctly identified as not wearing tags.

By using the 3D skeleton tracking of the Kinect, the system was able to recognize the direction of movement of the correctly identified skeletons at a 100% accuracy. Although this study is set up in a highly uncontrolled environment, the main objective of the study is to show the ability to function equally for registered users and unregistered users, giving bigger context for the application space where the system can be deployed.

5.8 Conclusion

This work presents a novel hybrid computer vision and RFID system that is capable of seamlessly matching the identity of an individual (as stored on an RFID tag) to the 3D image of that person as captured by a depth camera. This real-time system is capable of determining the identity of individuals within 4 seconds at an accuracy of 96.6% and groups of five people in 7 seconds with 95% accuracy.

Users studies show that ID-Match is capable of robustly identifying people with enough speed and accuracy to enable a humanoid autonomous robot to naturally interact with up to five people simultaneously. The system has also been shown to be effective at passively monitoring both tag and un-tag participants without requiring active participation for identification and tracking.

In order to demonstrate that this approach is scalable to other usage scenarios several controlled lab experiments are presented that cover many of the edge conditions and worst case scenarios. Including participant re-acquisition after visual tracking is lost, and the systems ability to distinguish between nearly identical tag motion. Ultimately ID-Match is a novel sensor fusion technique providing a valuable method for enabling automated computing systems to quickly and accurately recognize multiple people simultaneously.

Chapter 6

SENSOR FUSION FOR ENHANCED AR OBJECT INTERACTION

6.1 Introduction

Augmented reality enables applications to seamlessly overlay virtual information onto the physical world, immersing a user's surrounding environment with interactive digital information. Recent advancements have brought dramatic changes to user experiences in many application spaces, including retail, gaming, education, and remote collaboration. In retail, for example, AR could enable systems that display customizable information about a product whenever a person interacts with it. That information could include online reviews or video introductions (Figure 6.1a). With the customer's permission, the store may be able to remind the customer of deals related to the product or push coupons directly to their phone. Such a system would allow shoppers to access rich information online while experiencing the details of products in the real world. In a residential setting, AR could also be used to provide real-time instructions as a user builds furniture or performs repairs. In an industry setting, AR could allow remote collaboration, getting expert knowledge to workers when and where it's needed.

None of these applications are possible without a precise understanding of the objects involved, when the objects are being manipulated by the user, and the task at hand. One way to automatically detect interactions with an object would be through an AR headsets outward-facing camera. State-of-the-art systems have begun to surpass the performance of humans at identifying objects [55, 54] ; however, these results are limited to classifying item categories (e.g., mugs, bowls, shirts) and each new category requires retraining of the classifier. Situations where objects have contextual differences but little or no visual difference, such as identical mugs owned by different people, pose serious challenges for systems solely based on computer vision. Scannable codes (bar or QR) provide another means of discriminating objects, but the code must be clearly visible to the user. Furthermore, the user must be able to locate the code to capture a clear photo of it, which is not trivial and adds to the users burden.

I propose IDCam, a hybrid system that combines two sensing modalities - computer vision and radio frequency identification (RFID) - to achieve accurate item identification for seamless information augmentation. I assume that each item is instrumented with an off-the-shelf RFID tag and that there are long-range RFID readers installed in the ambient environment (roughly 10 m apart) to query these tags in real time. In retail environments, such a deployment is plausible in the near future given the push by major retailers to deploy RFID systems throughout their stores [94]. As people interact with RFID-tagged objects, they disturb the RF signal that is reflected off the tag and back to the reader. Using insights from prior work [83], these disruptions manifest in the low-level channel parameters measured by an RFID reader. IDCam captures these changes to determine which items are being manipulated and how fast they are moving. Separately, the motion of the customer's hands is tracked via their AR headset using computer vision. When these traces are correlated to one another, users can be matched with interactions with objects.

IDCam is capable identifying objects and sensing user interactions even when multiple people in the same location interact with similar items. To support this claim, IDCam is evalauted through a simulated shopping scenario. Two groups of five participants were asked to interact with 20 clothing items on the same rack simultaneously, which is beyond the typical density of interactions at a normal retail store. IDCam managed to match user interactions with object identities with 82.0% accuracy within 2 seconds.

Our contributions are the following:

1. A technique that correlates velocity calculations from RFID and computer vision with different coordinates systems to match object interactions with users, and

2. An evaluation with two groups of 5 participants in a dense simulated shopping scenario to quantify the performance of our system.

6.2 System Implementation

In this section, I will outline the hardware requirements for IDCam. I will then describe the software components, including the processing of the data streams and how they are correlated together to match item interactions with the users.

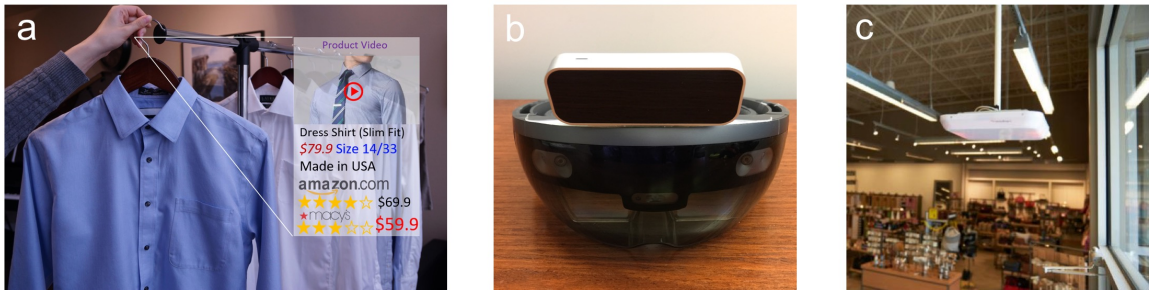


Figure 6.1: **(left)** IDCam enables information to be automatically displayed with a product when customer interacts with it wearing AR headset; **(middle)** a HoloLens headset with a Leap Motion controller mounted on top; **(right)** an RFID antenna installed on the ceiling of a retail store

6.2.1 Hardware

Figure 6.1 shows the different hardware components that comprise IDCam. To take advantage of IDCam, a user must wear an AR headset; this is necessary anyway in order to present real-time data about objects as an overlay. In the study, Microsoft's HoloLens [99] (figure 6.1b) is employed for augmenting visual information. Even though the HoloLens is equipped with an outward-facing stereo camera and IR structured light, a hand-tracking API has not yet been released. To get access to accurate hand-tracking data, I mounted a Leap Motion controller [81] on top of the HoloLens. RFID readers are mounted on the ceiling of the environment in question to provide good coverage of the space (figure 6.1c). An RFID tag is placed on each object to provide uniquely identifying information. RFID tags are inexpensive ($0.10), thin, and battery-free, making them as easy to use as the price tags that are already attached on products.

6.2.2 Approach

Consider a case where a user picks up a shirt instrumented with an RFID tag while wearing an AR headset in a store (Figure 6.2). Assuming the position of the RFID tag is fixed relative to the customer's hand as they take the shirt off of the rack, the velocity vectors of their hand ($\vec{V_h}$) and the RFID tag ($\vec{V_t}$) at any given moment should be approximately equal. As a result, the trajectory of the hand $\vec{Tr_h}$ and the tag $\vec{Tr_t}$ should be similar. Even when there are other customers interacting with

nearby items, it is unlikely that all of the customers will move items in the same way, especially with respect to the RFID antenna. Based on this observation, users can be matched with item interactions through a correlation of velocity vectors between the RFID and computer vision systems.

Our system was implemented as a HoloLens app and a MATLAB server running on a desktop PC. The Leap Motion and RFID antenna interface directly with this PC. The HoloLens app is implemented using the Unity 3D game engine. It streams information about the user's head position and orientation to the MATLAB server. The server is responsible for matching this information with data from the Leap Motion and RFID antenna, computing the RFID and hand velocities, and detecting whether an object is being held by the user.

The hand location (denoted as X, Y, Z in Figure 6.2) can be accessed using the Leap Motion controller's API. Computing the change in hand location over time leads to measurement of $\vec{V_h}$ in the Leap Motion's coordinate space. RFID readers can report the phase angle difference between the transmitted signal and the backscattered signal. Prior work has demonstrated that the motion trace of a tag relative to the antenna can be reconstructed by utilizing the phase information retrieved from multiple communication channels [83], thus yielding a measurement of $\vec{V_t}$. The velocity measurements from the computer vision and RFID systems are not directly comparable since they are measured relative to different reference points. The RFID velocity is computed as 1-dimensional radial movement relative to the ceiling-mounted antenna. The user's hand velocity is measured in 3D relative to the Leap Motion controller, whose position and orientation change as the user moves around in the retail store. As a result, the two coordinate systems must be unified before the velocities are compared.

6.2.3 Coordinate Transformation

The different coordinate frames of the HoloLens, Leap Motion, and RFID reader present a challenge in correlating motion from the RFID tags and hand. In this section, I describe our approach to integrate data from these different systems.

The reference frames relevant to computation is defined as below:

- **W**: The world reference frame, defined by the position of the HoloLens during a one-shot calibration procedure. The RFID reader is specified in this frame based on the geometry of

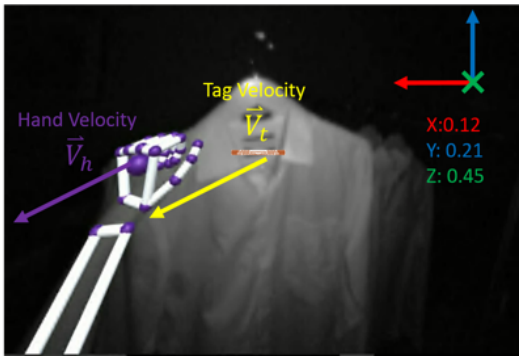


Figure 6.2: IDCam: a frame showing the output of the Leap Motion controller overlaid with velocity vectors for the hand ($\vec{V}_h$) and the RFID tag ($\vec{V}_t$)

the environment,

- **U**: The Unity world frame, defined by the placement of the app in the environment when launched,

- **C**: The HoloLens camera frame, defined within **U** by the tracking system of the HoloLens,

- **L**: The Leap Motion frame, defined by the physical placement of the Leap Motion on the HoloLens.

Qquaternions is employed for our calculation of the coordinate transformation to eliminate gimbal lock. To explain the transformation, notation from [95] is adopted. ${}^{A}_{B}\hat{q}$ represents the rotation of frame B relative to frame A, and ${}^{A}_{C}p$ represents the position of object C in frame A. The objective is to compute ${}^{W}_{H}p$, the position of the hand in world space.

On the HoloLens, the position and orientation of the HoloLens is computed in the Unity frame, ${}^{U}_{C}\hat{q}$ and ${}^{U}_{C}p$ respectively, and stream this to MATLAB for further analysis. To align the Unity (**U**) and world (**W**) reference frames, HoloLens is first positioned at the origin of the desired world coordinates, pointed toward the floor. The values are saved as ${}^{U}_{C_0}\hat{q}$ and ${}^{U}_{C_0}p$. HoloLens camera frame is defined during calibration, C_0, as the world frame **W**. This allows us to treat ${}^{U}_{W}\hat{q}$ as ${}^{U}_{C_0}\hat{q}$

New positions and orientations are transformed to compute the HoloLens position in world space.

$$ {}^{W}_{C}\hat{q} = {}^{U}_{C}\hat{q} \otimes {}^{U}_{W}\hat{q}^{\star} \qquad \text{Orientation of hand in world space} \tag{6.1} $$

$$ {}^{W}_{C}p = {}^{U}_{W}\hat{q} \otimes \left({}^{U}_{C}p - {}^{U}_{C_0}p \right) \otimes {}^{U}_{W}\hat{q}^{\star} \qquad \text{Position of hand in world space} \tag{6.2} $$

To integrate data from the Leap Motion hand tracker, we transform the raw coordinates of the hand in the **L** frame, ${}^{L}_{H}p$, to the HoloLens camera frame. This is done by transforming the right-handed coordinates to approximately match the left-handed Unity coordinates by Equation 6.3.

$$ {}^{U}_{H}p' = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \times {}^{L}_{H}p \tag{6.3} $$

The LeapMotion was physically tilted downwards relative to the HoloLens, so we accounted for that in the calculation by rotating estimated hand coordinates 20 degrees in the opposite direction to obtain a better estimate of hand position in the camera frame, ${}^{U}_{H}p$. Finally, the hand position is obtained in world coordinates by applying the HoloLens rotation and position quaternions:

$$ {}^{W}_{H}p = \left({}^{W}_{C}\hat{q}^{\star} \otimes {}^{C}_{H}p \otimes {}^{Q}_{C}\hat{q} \right) + {}^{W}_{C}p \tag{6.4} $$

6.2.4 Trace Correlation

Since there can be scenarios with thousands of RFID-tagged objects, it would be infeasible to compute the velocity correlation between all pairs of objects and users. IDCam first narrows down the potential items with which a user could be interacting by leveraging knowledge about the location of the multiple RFID antennas installed in the HoloLens' 3D mapping system. IDCam only considers the items that are seen by the closest antenna to the user.

The sensing area of each RFID antenna is approximately 200 sqft, so there may still be many tags that need to be checked. IDCam further reduces candidates by filtering out tags with insignificant motions. When a person takes a shirt off of a rack, for example, they are likely to jostle many of the other shirts on that rack by a small amount. I observed that these interactions result in a tag speed of no faster than 10 cm/s, so IDCam uses this threshold to ignore items with no significant motion.

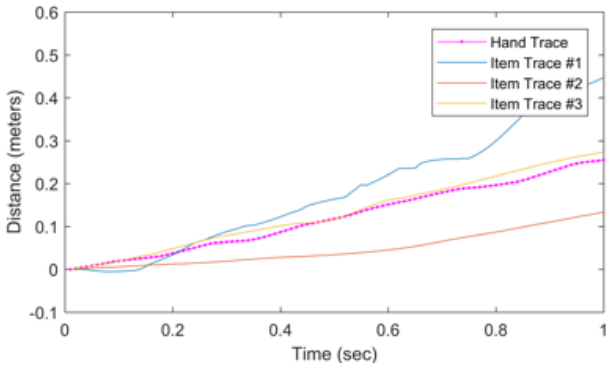


Figure 6.3: IDCam: a one second segment of the motion trace tracked of hand Tr_h when compared to 3 different RFID traces Tr_t1, Tr_t2 Tr_t3

Combining the location information of items and the speed threshold, IDCam narrows the list of potential items that the user could be handling by a significant margin.

Velocity Correlation

Figure 6.3 displays a 1-second segment of position data from a person's Leap Motion (pink) and the RFID system (blue, orange, yellow). The position in both domains was calculated by integrating the velocity measurements; given that the antenna can only infer the relative position change of items, all of the position traces start at the origin for easy comparison. As the user grabbed a tagged object, two other people moved nearby tagged objects with a velocity greater than 10 cm/s relative to the RFID reader. By overlaying all four trajectories, it is clear to see that the motion of Item #3 (yellow) is most correlated with the participant's hand (pink). IDCam computes trace similarity using the Pearson correlation coefficient (PCC) over 1-second windows with 80% overlap. The object that leads to the highest correlation coefficient is likely the one that the user is manipulating, so its information is displayed on the AR headset.

Figure 6.4 shows the distance of a user's hand relative to the antenna as they pick up 5 different items within roughly 30 seconds. While the user interacted with items, other users were also disturbing nearby tags. Each rising edge indicates when the user takes an item off the rack (and away from the RFID reader), while each falling edge indicates the opposite. Green sections represent when the

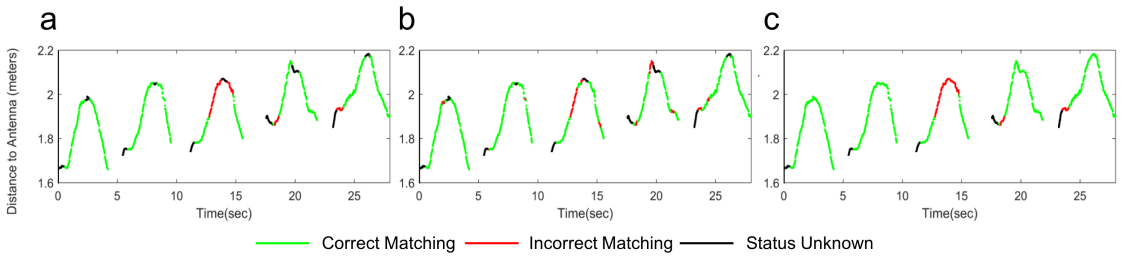


Figure 6.4: **(left)** An example scenario that shows the accuracy of using PCC alone for matching. In this case, the user handles five products sequentially. **(middle)** a 1 second voting buffer is applied to smooth the results. **(right)** nearest neighbor algorithm was applied to fill in motion gaps.

user was correctly matched with the item they were handling. Red sections indicate mismatches, which happen when interactions conducted by other users are falsely matched with the target user. Black sections indicate the that IDCam was uncertain because no significant motion was detected from the RFID tags or there was poor correlation between the 2 systems. The left part of Figure 6.4 shows an example of how well the described matching algorithm worked for that scenario. Using the PCC correlation, IDCam has a matching accuracy of 84.0%. Further assumptions can be made in order to remove brief, spurious detections. It is highly unlikely that a user will handle one item for 5 seconds, handle a second item for 0.5 seconds, and then go back to the original item for 5 seconds; it is more likely that the second item was mismatched with the customer. IDCam handles these cases by building a 1 second buffer of correlation results and then applying a majority vote across them for the best match. Using this method, matching accuracy in the example scenario is improved to 86.9% (Figure 6.4, middle). Nearest neighbor matching can also be used to reduce brief moments when no object motion is detected (e.g., the customer is holding the item, but not moving it). Using this method, matching accuracy of the example data is improved to 89.5% (Figure 6.4, right).

6.3 Evaluation

The accuracy of IDCam is a function of simultaneous item interactions in the area of a single RFID reader. Off-the-shelf RFID antennas can cover 200 sqft, but RFID systems that can localize objects

Figure 6.5: IDCam: each article of clothing had an RFID tag placed near the price tag.

(e.g., Impinj Xarray) can only do so with an accuracy of 5 ft, the rough size of a clothing rack. Increasing the number of interactions in that area leads to more candidates for correlation matching, making it a more challenging task. In real-world applications, item interactions can be sorted between RFID readers by individual shelves/racks, so users not in the proximity of an RFID antenna will not be considered for its matching calculation.

To evaluate the effectiveness of the matching algorithm, a simulated retail experience was set up. A single rack of clothing on hangers was placed in the middle of a laboratory space along with an RFID reader. 10 (6 males, 4 females) undergraduate and graduate students from a public university was recruited. The participants were separated into 2 groups, each with 5 participants, and asked to interact with multiple articles of clothing simultaneously (Figure 6.5), as if they were all interested in purchasing something in that area. Beyond the simultaneous interactions, clothing was also free to jostle on the rack, creating false positive item interactions. I believe this scenario makes the study representative and perhaps even more challenging than a real-world retail situation since so many people will likely not be in the same place at one.

6.3.1 Study Procedures

Twenty articles of clothing were hung on a 1.5 m-long rack. An RFID tag was placed in the center of each item's price tag (Figure 6.5). The RFID reader was placed 2 meters away from the shelf.

Figure 6.6: The simulated shopping experience used to evaluate IDCam involved five participants interacting with clothing on a rack. One of the five participants wore a HoloLens with a Leap Motion controller mounted on top

For each study session, all participants were asked to pick up an article of clothing by its hanger, read the product information on the price tag, and then put the hanger back on the rack. One of the participants was asked to wear a HoloLens with a Leap Motion controller while they picked up hangers. The participant wearing the HoloLens was asked to pick up each hanger in sequence for the ease of ground truth labeling, while all other participants were free to interact with items as they chose. The start and finish time for each interaction for the person wearing the HoloLens was annotated using a video recording. Each of the 5 participants took turns wearing the HoloLens. There were 893 item interactions in total by participants, 200 interactions of which were generated by users wearing the HoloLens. Each session took 5.1 minutes on average. Each item interaction took 13.7 seconds on average, which means that 4 items were typically in motion simultaneously.

6.3.2 Study Evaluation

The Leap Motion controller has difficulty tracking a person's hand with a cluttered background. Given that the contribution is not meant to be one of computer vision, but rather the fusion of RFID and visual traces, participants were asked to extend their hands in front of the Leap Motion controller as an initialization gesture. There were still 17 interactions among the 200 collected in the user

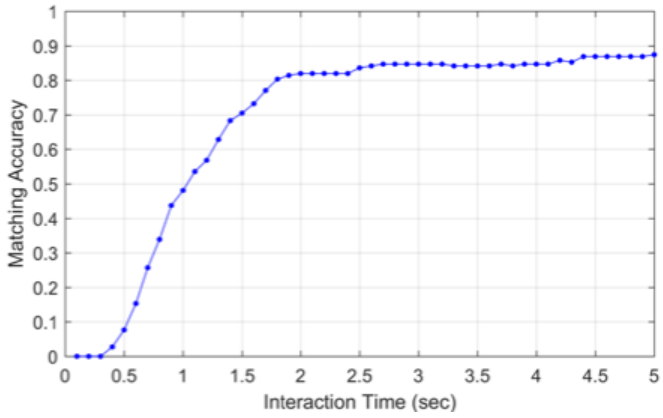


Figure 6.7: IDCam: a cumulative distribution function showing the accuracy of item identification as a function of the duration of the correlated traces.

study that were not properly tracked by the Leap Motion controller. For the 183 item interactions that were fully or partially successfully tracked, the hand motion traces were transformed into the RFID antenna's coordinate space and compared them to RFID tags.

Figure 6.7 is a cumulative distribution function showing IDCam's interaction matching accuracy as a function of the time duration of the interaction traces. After 2 seconds, IDCam achieved an accuracy of 82.0%. The accuracy improved to 88.0% after 5 seconds of user interactions. At the beginning of each interaction, it is hard for the system to make accurate decisions due to the lack of motion. As the user generate longer motion traces, the system makes better decisions using the sliding window and voting buffer approach mentioned in the previous section.

6.4 Discussion

The goal was to develop a method of precise object identification for augmented reality that would be easy to deploy and scale to large number of new objects. I addressed this goal by proposing IDCam, a system that fuses velocity information from computer vision and RFID to identify the precise object of interest for a user. IDCam is relevant to a number of augmented reality applications, In an evaluation study focused on a simulated shopping scenario, IDCam is tested with 5 participants interacting with items on the same rack simultaneously. In that study, IDCam correctly matched

users with the objects they were interacting 82.0% of the time within 2 seconds.

The current capabilities of hand tracking imposes a number of limitations on IDCam. There were 17 cases of item interactions where the Leap Motion controller failed to track the user's hand. This was primarily due to the hand being outside of its field of view. There were also additional 10 cases when the hand was only temporarily tracked. In general, I believe the evaluation study provides a lower bound for object identification in real-world scenarios and I believe hand tracking will be much more accurate in the near future given recent advancements in computer vision. This trend should continue to improve IDCam's accuracy and robustness. Also, AR systems like HoloLens already have hand-tracking capabilities that I hope will be available soon to replace the Leap Motion to help provide better hand tracking for the system.

Natural user motions can be challenging to track since they are unconstrained, so providing users with instructions can help improve IDCam's performance in edge cases. As an extreme example, IDCam could suggest that a user shakes an item to help with edge cases when the item is not being detected initially.

6.5 Conclusion

IDCam is a novel system that fuses long-range RFID with computer vision for linking the objects in physical world with digital content. IDCam matches object identity stored in the RFID tag to user interactions by correlating the motion traces tracked by the two systems. IDCam overcomes challenges of differentiating visually similar items while scaling well to a large number of new items without requiring retraining. IDCam system is evaluated in a simulated retail scenario with a high density of users. I believe IDCam has many other application potentials in fields such as remote instruction, collaboration and education and intend to explore and evaluate these application spaces in the future work.

Chapter 7

CONCLUSION

In this thesis, I presented a new class of sensing techniques leveraging low-level channel parameters of commercial RFID system to detect the state of everyday objects, users, and the interactions between the two parties. By applying signal processing and machine learning techniques and customizing RFID antenna design, I turned off-the-shelf RFID systems into sensors that measure object motion, detect user presence, and understand touch and mid-air gesture interactions. I also presented a sensor fusion system combining RFID with computer vision to precisely identify and localize individuals and objects in complex and dynamic physical environments.

These sensing capabilities have a variety of applications in understanding physical environments and connecting physical objects to digital media. In this thesis I proposed and evaluated applications such as (1) autonomous recognition of fine-grain daily activities, (2) customizable touch interfaces for everyday objects, (3) natural human-robot interactions, (4) enhanced AR object interactions as well as (5) shopper's interaction tracking in retail environments.

The major contributions of this thesis can be summarized into three different categories: (1) signal processing and machine learning algorithms that can be applied to both RF channel parameters and RF-vision sensor fusion to enable various sensing applications presented in this thesis. (2) Novel system design that combines RFID with other hardware components such as inkjet printed circuits, depth-sensing camera, robot hosts and AR head-mounted displays to facilitate interactive user experiences. (3) Designing and conducting user studies to benchmark the performance of the sensing techniques in lab controlled or real-world environments. A more detailed list of contributions is included in Chapter 1 Section 4. These contributions support and prove the thesis statement:

Applying signal processing and machine learning techniques to low-level channel parameters of commercially available RFID tags can enable novel sensing and interactive applications with everyday objects.

In the later sections of this chapter, I would like to discuss a few additional application domains that could be impacted from my work described in this thesis. Then I will discuss methodology, techniques, and lessons learned that are generalizable to other sensing platforms or other research domains. Finally, I will discuss directions that I would like to explore in my future research.

7.1 Potential Application Spaces

7.1.1 Enhancing Augmented/Virtual Reality Experiences

My research involves understanding peoples daily activities through precise identification and localization of objects and individuals. This can enable computer systems to digitize everyday objects and surroundings in the real world and combine them with virtual content generated by AR and VR technologies. For example, a VR device can use the hybrid system to identify and locate a chair, and then render it in the virtual world as an iron throne, allowing users to experience natural interactions with the virtual world through objects that exist in reality. This is related to the passive haptics research field which often relies on cameras to track the geometry of the objects. With our system, object identity and geometry can be digitally embedded to enable much faster tracking and natural interactions. By combining precise item identification with augmented reality (AR), one can also seamlessly merge online media with physical environments, bring immediate changes in high impact spaces such as health and wellness, remote collaboration, education, and retail.

7.1.2 Improving social interaction through fine-grained activity recognition

Social media connects billions of people with each other. My research involves connecting each user with hundreds of objects they interact with on a daily basis, creating interactive user experiences with the physical surroundings. I believe my research will help social media to achieve greater impact by bringing user interactions from online and mobile platforms to offline real-world scenarios, facilitating people-to-people interactions. As an example, imagine a scenario where your coffee mug, instrumented with a passive sensor, sends a notification to your friend when you pick it up and start moving to the coffee machine. Social media apps could encourage people who respond to that notification to get coffee together, thereby facilitating face-to-face interaction. Object that

users interact with define and reflect their daily activities. Leveraging the convergence of people and objects to facilitate user interactions will allow engaging social media experiences.

7.1.3 Smart Home Applications

My research focuses on creating interactive user experiences with the physical surroundings through detection of user interactions with daily objects. I believe there are many opportunities in this space leveraging low-cost, passive sensors and sensor fusion to detect fine-grained daily activities. Consider an environment where your water cup is aware of your hydration level, your pill bottle is aware of your medication intake and each physical space is aware of the current activity and can adjust lighting, temperature, and humidity accordingly. Understanding of users daily activities can enable many context-aware applications to improve the life quality of everyday users.

7.1.4 Business Intelligence Applications

When detection of object interactions is combined with business scenarios, techniques presented in this thesis can help business to understand customer behavior and to create novel user experiences. In the shopping scenario discussed earlier, interaction sensing can quantify consumer interest through their interactions with products. One can measure the conversion between user's visit at each store section towards item interactions and sales. These data will provide critical information towards business analytic platforms and can help retailers to understand customer journey and to conduct effective A/B testing on merchandising and pricing decisions. RFID-based sensing of object interactions can also be applied towards automatic (line-free) check-out and theft prevention, improving customer experiences while reducing loss for the business.

7.2 Improving RFID readers for Interaction Sensing Applications

So far, I have demonstrated a variety of applications that could be enabled by the UHF RFID. However, there are limitations in the current RFID systems that I would like to discuss here, with the goal of help informing the design of the next generation of RFID for interactive IoT applications.

7.2.1 Improving Reading Rate and Tag Sensitivity

The sensing accuracy is a function of read rate. When the reading rate of each tag drops below 5 reads per second, it is very challenging to guarantee accurate interaction detection. Low reading rate means bigger gaps between samples which makes it more likely to miss some of the user interaction events, creating false negatives. The industry standard today for UHF RFID reader is the EPC-Gen2, which in theory allow each reader to query tags at 1000+ reads per second. However, in reality, a maximum read rate between 700-800 reads per second were observed under ideal conditions. If there are more than 200 tags visible to the reader, the average reading rate of each tag will drop below 4 per second. In addition, insufficient sensitivity of the reader also resulted in decreased reading rate in tags approaching the boundaries of the reader's reading range (10m), or those within sensing range yet under challenging ambient RF environments, which makes it challenging to maintain a consistent sensing accuracy across large physical environments.

The current generation of RFID systems employs a slotted Aloha protocol for communication. Research in the field of RFID communication protocols has demonstrated promising results in improving sampling rate by mitigating collision problems between RFID tags. Integration of more advanced communication protocols will help improve the reading rate of the UHF RFID system. The pace of hardware research and engineering efforts improves the reader.by approximately 1 dBm every year. Having readers with better sensitivity and faster reading capabilities will dramatically improve the scalability of RFID systems for interactive sensing purposes under dynamic conditions and in complex environments.

7.2.2 Form Factor and Power Consumption

The smallest UHF reader module on the market today comes in size of about 1-inch square with 0.1-inch thickness and its power consumption is about 2 watts when tuned for 1-meter sensing range. The module size and power consumption today are still not ideal for integration into mobile platforms such as the smartphone or AR head-mounted displays. For the scope of this thesis, the readers are installed into the physical environments. However, with the current trend in reducing power consumption and form factor. Future RFID readers will have better opportunities to be integrated into mainstream mobile devices to facilitate broader application spaces.

7.2.3 Channel Hopping and Duty Cycle Control

Current APIs of commercial RFID readers do not allow control of physical parameters such as channel hopping frequency or power shifting frequency between multiple antenna ports. User control over these parameters may seem excessive to the RFID manufacture because the reader is designed for inventory tracking. Also, full control of these functionalities has the risk of validating FCC regulations. However, I would argue that certain level of control should be given to users for building applications beyond inventory tracking. Take the user-product interaction sensing in retail environments as an example - discontinuity of the phase signal caused by frequency hopping and antenna switching will reduce the effective samples for determining the item states which can be problematic especially when items are densely populated. There is plenty of room to improve this situation if partial control access to the reader firmware is granted without violating FCC regulations.

7.2.4 Specialty Tags

Commercial UHF tags work well for objects with low dielectric constant. However, it is challenging to find tags that work well for materials with high dielectric constants. The primary reason is the dielectric of these materials can still vary a lot from each other, and it is challenging to manufacture general purpose tags that works well for a wide range of objects without knowing their dielectric constant. Real-world scenarios can be even more challenging since certain objects consist of more than one material. One interesting idea that may be worth exploring is to manufacture tags with special antenna design whose size can be easily customized (trimmed) to adapt to materials with different dielectric constant. Another idea is to have specialized materials with adjustable thickness attached in between the tag and the object to help to improve the impedance matching for objects consisting of metal and liquid.

7.3 Lesson Learned from Building Sensing Systems

7.3.1 Generating Hypothesis from Physical Principles

Sensing is the discovery process of turning sensory data into the understanding of the physical world. To achieve this goal, one must first start with understanding the fundamental physical principles of the sensing platform. Take the RFID-based gesture sensing work in Chapter 4 as an example;

understanding of the fundamental physical properties of RSSI, RF phase, Doppler shift, as well as theory related to impedance matching and multipath reflection is required to hypothesize the correlation between user's gestures and the change in the RF channel parameters. In addition, having a comprehensive understanding of a variety of sensing modalities is crucial towards application-driven research, where one needs to find the most elegant and effective sensing approach among many potential candidates.

7.3.2 Importance of Interactive Data Visualization

Once a hypothesis is established between a physical process and its correlation with signals, the next step will be visualizing these signals in real-time to validate these hypotheses. Note that finding the best visualization method is usually an iterative process, where one can start with visualizing a moving window of the raw signals in time domain and frequency domains via FFT, then move on to customize the visualization according to the specific nature of the sensing system. This will likely be an interesting learning process to uncover the hidden correlation between the signal and the corresponding physical process that one is trying to sense. In the case where the initial hypothesis is disproven, good real-time visualization of signals may help discover new patterns beyond the original hypothesis.

7.3.3 Data Cleaning

In cases when the correlations that one tries to observe are hidden due to the noisy nature of the signal, signal processing can be helpful for noise reduction. Familiarity with noise removal filters in both time and frequency domains such as Wiener, Kalman or Butterworth filters are beneficial for data cleaning. Note that data cleaning and visualization are an interactive process where the user can continue to experiment with the filter selection until the signal demonstrates clear patterns that are either consistent with or contradicts the original hypothesis.

7.3.4 Feature Engineering

Machine learning can help turn raw signals into understanding of the physical environments. A typical output for machine learning is a multi-state classification or a regression correlation. The

deep neural network has simplified feature engineering by taking raw data as input, yet it requires lots of training data to perform accurately. Many classification tasks with smaller scale training data still require feature engineering using observations and domain knowledge. Feature engineering aims to find the variables as a function of the raw signal that has significant prediction power for machine learning models. Take RFID motion detection as an example, the phase signal is know to be correlated with the tag location wrapped in each wavelength. So when the tag is under motion, the phase signal collected from each communication channel will have a much larger variation compared to when the tag is still. Consequently, taking the standard deviation of phase for each channel will be a good feature to differentiate moving tags from still tags. When feature selection is completed, it is usually beneficial to visualize these features to confirm they can properly capture different physical events.

It is important to engineer features that match with physical intuitions. For example, hand waving gestures around the tags cause variation in the channel parameters. To detect this event, features looking at variations in Discrete data points such as differences of channel parameters in time were employed. Unlike absolute channel parameter values, variation features are not impacted by the relative physical location between the tag and the reader, allowing the system to detect hand waving events across multiple physical locations. In addition, variation features can better characterize the changes in signals when the hand gesture happens, which matches with the physical intuition. The sensing goal determines the feature engineering process. For example, if one wants to determine the absolute location of tags, absolute values of channel parameters will be much more useful than differential features.

7.3.5 Staged Experimentation for Machine Learning Exploration

To validate the effectiveness of the selected features for classification tasks. One can start with collecting a small dataset using staged experimentation. Methods such as video recording can be employed for annotation of the data. With labeled data, one can perform cross validation using popular machine learning models. Given the dataset collected in the staged experimentation is usually too small for deep learning approaches, one can start with popular methods such as Random Forest or SVM classifier. Implementation of Random Forest is straightforward, but SVM requires

more attention on kernel selection and parameter tuning. Parameters can be learned and optimized during the cross-validation process using the training data.

7.3.6 Scaling to Multiple Physical environments

If the small-scale staged experimentation turned out to be successful, the immediate next step would be trying to scale to a larger dataset. For sensing applications, it is essential to collect data from multiple users as well as multiple physical environments for training purposes so that the training data accounts for the variability present in the physical world. Then, one validates its accuracy by applying the machine model towards new incoming data. It is also important to examine the sensing accuracy via data collected from physical environments and users that were not included in the training data to evaluate its scalability.

Even with careful system design and machine learning considerations, dynamic and complex real-world environments can raise new scalability challenges, especially if the sensing application is based on a single sensing mechanism. It is often worth considering employing a few different sensing mechanisms that are complementary regarding their sensing advantages and disadvantages. If the sensing accuracy is a high priority, it is okay to introduce redundancy, in which case one event will be detected by several different sensing mechanisms. In fact, sensing redundancy is often necessary in real-world applications to guarantee system robustness. When deploying a system with multiple sensing modalities, it is also worth exploring the signal relationship between modalities to discover new sensing capabilities through a sensor fusion approach.

7.4 Sensing Methodologies

7.4.1 Generalizing to other platforms

The signal processing and machine learning techniques being applied towards the RFID system are based on a few signals such as RSSI, RF Phase, and data rate. The generalizability of the discussed techniques towards other platform is dependent on the availability of these signals. Even though all channel parameters explored in this work are not unique to the RFID system, their availability and granularity are different across various RF platforms. RSSI is the most common one that has been thoroughly investigated across WiFi, Bluetooth, and Zigbee systems for localization purposes.

I believe the application space for RSSI based sensing on these other RF devices can be extended towards hand gesture sensing leveraging techniques mentioned in this thesis. The support for phase signal varies across platforms. Commercially available Bluetooth or WiFi chips, in general, do not support API for phase difference measurement between devices. Zigbee allows open access towards phase difference measurement given its onboard phase measurement unit (PMU). Even though that phase measurement on Zigbee is based on differences between time-synchronized RF modules instead of the difference of transmitted and backscattered signal, its physical nature is very similar to the RF phase of the RFID system which describes device relative location within each wavelength. Prior research has demonstrated leveraging Zigbee phase difference measurement to do fine-grained distance estimation [142]. Given the availability of both RSSI and phase difference measurement on Zigbee devices, I believe this application space can be extended towards understand how people are manipulating physical objects instrumented with Zigbee, sensing the presence of users in the ambient environment as well as their daily activities. In addition, the velocity-based correlation methodology for fusing RFID with computer vision presented in this thesis could also be repurposed to enable sensor fusion between Zigbee and other sensing platforms.

7.4.2 Simplicity in Sensing Hardware

To achieve goals such as understanding object interactions or creating tangible touch interfaces, prior work has demonstrated many different approaches consisting of a variety of hardware and sensing modalities. Different from prior work which mostly took on the approach of building a new and highly customized system with dedicated sensors, in this thesis, I focus on taking simple and off-the-shelf devices that are already deployed. I then apply signal processing and machine learning techniques to enhance its sensing capabilities through a software approach. Such methods will have more design constraints when compared to building a new sensing systems from the ground up. However, it is beneficial towards scalability as well as immediate impact since software algorithms can be easily deployed. I believe this principle in sensing can be valuable for research and engineering beyond the scope of this thesis, especially in the real-world environment where one can expand the boundaries of sensing by adding more intelligence towards how signals of existing off-the-shelf devices are processed and trained to enable new applications.

7.4.3 Modeling through Physical Principles

Given the constraints of commercially available hardware, it is critical to understand the physical properties of the sensing system. Each sensing system has its unique physical principles due to the sensing modalities utilized and the firmware design specific to these sensing systems. A clear understanding of the physical principles and its correlation with potential application spaces is crucial for discovering new sensing opportunities that were not included in the original system design. During my research, I found it greatly beneficial to break down complex problems using physical principles and reason upwards to identify unconventional solutions. In addition to the scope of work presented in this thesis, I also applied this principle towards infrared and capacitive sensing platforms and discovered interesting results in gesture sensing and user localization.

7.5 Future Research Directions

7.5.1 Smart Environment with Artificial Intelligence

I am passionate about building sensing systems and creating compelling user experiences around them. Sensing tackles the problem space of detecting events happening in the physical world, and AI will bridge the gap between sensing and understanding. With the ever-increasing volume of devices in the IoT and major adoption of artificial intelligence across the industry, the line between what's physical and what's digital has started to blur. There are many great opportunities to create compelling user experiences leveraging computational resources and information in the digital world while at the same time allowing users to interact with the physical environments in a natural way. This direction of research extending AI applications through sensing the physical world is of great interest to me, and I would love to continue exploring this space in future research.

7.5.2 Mobile Sensing and Interactions

I am also interested in finding new ways to enrich the interactive experiences between humans and computers. I am particularly interested in sensing and interaction opportunities on mobile devices. Mobile phones today are capable of collecting a wide spectrum of sensory data such as capacitive, electromagnetic, infrared, NFC, acceleration, and audio signals. This sensory data indicates the state of the device and activities that the user is engaging in, which provides opportunities for contextual-

aware computing applications. Utilizing existing sensors or building new attachments to mobile devices could also enable new and interesting ways for users to interact with mobile devices and more satisfying user experiences.

7.5.3 Customizable Physical User Interfaces

Physical interfaces such as touch input devices are ubiquitous. Unlike digital interfaces which are easy to customize using computer graphics, physical interfaces usually come with fixed form-factors which makes it very challenging to customized once manufactured. One of my research interests is to explore easy and scalable ways to prototype physical interfaces that are customizable by users and can be easily fabricated into the physical objects. Following up on the touch interface work combining RFID chips with conductive ink-jet printing, I have been exploring ways to further enhance the sensing capabilities to support 2D or 3D continuous touch interactions using smart antenna design and signal processing. With the advancement in digital fabrication and the variety of materials that can be integrated, I believe there are many research opportunities for fabricating interactive objects, allowing the maker community to customize touch interactions in an easy and scalable way.

7.6 Summary

I hope this dissertation serves as an inspiration for anyone who is interested in leveraging RF channel parameters of wireless devices to build interactive IoT applications. I strongly believe that repurposing devices that has already been deployed in the physical world via signal processing and machine learning approaches could bring immediate changes to high impact application spaces such as smart cities, personal health, education, productivity etc. Sensing helps computer systems to make sense of the world, blurring the line between what is physical and what is digital. With emergence of the Internet of Things, there will be more and more sensing opportunities like the one discussed in this thesis waiting to be discovered by future researchers.

BIBLIOGRAPHY

[1] Heba Abdelnasser, Khaled A Harras, and Moustafa Youssef. Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 277–286. ACM, 2015.

[2] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-person motion tracking via rf body reflections. *Computer Science and Artificial Intelligence Laboratory Technical Report*, 2014.

[3] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 837–846. ACM, 2015.

[4] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems*, pages 114–130. Springer, 2012.

[5] Samer Al Moubayed and Jill Lehman. Regulating turn-taking in multi-child spoken interaction. In *Intelligent Virtual Agents*, pages 363–374. Springer, 2015.

[6] Toshifumi Arai, Dietmar Aust, and Scott E. Hudson. Paperlink: A technique for hyperlinking from real paper to electronic content. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 327–334, New York, NY, USA, 1997. ACM.

[7] Arduino. Arduino & Genuino Product Page. `https://www.arduino.cc/en/Main/Products`, 2018. [Online; accessed 3-May-2018].

[8] Parvin Asadzadeh, Lars Kulik, and Egemen Tanin. Gesture recognition using rfid technology. *Personal and Ubiquitous Computing*, 16(3):225–234, 2012.

[9] Daniel Avrahami and Scott E. Hudson. Forming interactivity: A tool for rapid prototyping of physical interactive products. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '02, pages 141–146, New York, NY, USA, 2002. ACM.

[10] Salah Azzouzi, Markus Cremer, Uwe Dettmar, Rainer Kronberger, and Thomas Knie. New measurement results for the localization of uhf rfid transponders using an angle of arrival (aoa) approach. In *RFID (IEEE RFID), 2011 IEEE International Conference on*, pages 91–97. IEEE, 2011.

[11] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer, 2004.

[12] Paul Beardsley, Jeroen Van Baar, Ramesh Raskar, and Clifton Forlines. Interaction using a handheld projector. *IEEE Computer Graphics and Applications*, 25(1):39–43, 2005.

[13] Luiz Belussi and Nina Hirata. Fast qr code detection in arbitrarily acquired images. In *Graphics, Patterns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on*, pages 281–288. IEEE, 2011.

[14] Rahul Bhattacharyya, Christian Floerkemeier, and Sanjay Sarma. Low-cost, ubiquitous rfid-tag-antenna-based sensing. *Proceedings of the IEEE*, 98(9):1593–1600, 2010.

[15] S Boncinelli, P Citti, E Del Re, G Campatelli, L Pierucci, and L Bocchi. Real time detection and tracking of gauzes by rfid uwb technique. In *RFID, 2010 IEEE International Conference on*, pages 97–101. IEEE, 2010.

[16] Maurizio A Bonuccelli, Francesca Lonetti, and Francesca Martelli. Tree slotted aloha: a new protocol for tag identification in rfid networks. In *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, pages 603–608. IEEE Computer Society, 2006.

[17] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp '09, pages 51–60, New York, NY, USA, 2009. ACM.

[18] Francesco Cafaro, Alessandro Panella, Leilah Lyons, Jessica Roberts, and Josh Radinsky. I see you there!: Developing identity-preserving embodied interaction for museum exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1911–1920, New York, NY, USA, 2013. ACM.

[19] Xiang Cao, Clifton Forlines, and Ravin Balakrishnan. Multi-user interaction using handheld projectors. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 43–52. ACM, 2007.

[20] Daniel Castro, Steven Hickson, Vinay Bettadapura, Edison Thomaz, Gregory Abowd, Henrik Christensen, and Irfan Essa. Predicting daily activities from egocentric images using deep learning. In *proceedings of the 2015 ACM International symposium on Wearable Computers*, pages 75–82. ACM, 2015.

[21] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[22] Jianfeng Chen, Jianmin Zhang, Alvin Harvey Kam, and Louis Shue. An automatic acoustic bathroom monitoring system. In *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, pages 1750–1753. IEEE, 2005.

[23] Saisakul Chernbumroong, Shuang Cang, Anthony Atkins, and Hongnian Yu. Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications*, 40(5):1662–1674, 2013.

[24] Gabe Cohn, Erich Stuntebeck, Jagdish Pandey, Brian Otis, Gregory D Abowd, and Shwetak N Patel. Snupi: sensor nodes utilizing powerline infrastructure. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 159–168. ACM, 2010.

[25] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Transactions on pattern analysis and machine intelligence*, 25(5):564–577, 2003.

[26] John C Curlander and Robert N McDonough. *Synthetic aperture radar*. John Wiley & Sons, 1991.

[27] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[28] Travis Deyle, Hai Nguyen, Matt Reynolds, and Charles C Kemp. Rf vision: Rfid receive signal strength indicator (rssi) images for sensor fusion and mobile manipulation. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 5553–5560. IEEE, 2009.

[29] Travis Deyle, Matthew S Reynolds, and Charles C Kemp. Finding and navigating to household objects with uhf rfid tags by optimizing rf signal strength. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 2579–2586. IEEE, 2014.

[30] Antonis G Dimitriou, Aggelos Bletsas, Anastasis C Polycarpou, and JohnN Sahalos. On efficient uhf rfid coverage inside a room. In *Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on*, pages 1–5. IEEE, 2010.

[31] Changxing Ding and Dacheng Tao. Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[32] EPC. EPC™ Radio-frequency identity protocols generation-2 UHF RFID specification for RFID air interface, Version 2.0. `http://www.gs1.org/sites/default/files/docs/epc/uhfc1g2_2_0_0_standard_20131101.pdf`, 2015. [Online; accessed 3-May-2018].

[33] GS EPCglobal. Epc radio-frequency identity protocols generation-2 uhf rfid; specification for rfid air interface protocol for communications at 860 mhz–960 mhz. *EPCglobal Inc., November*, 2013.

[34] FCC. Title 47: Telecommunication, Part 15 Radio Frequency Devices. `https://www.gpo.gov/fdsys/pkg/CFR-2010-title47-vol1/pdf/CFR-2010-title47-vol1-part15.pdf`, 2010. [Online; accessed 3-May-2018].

[35] FCC. EPC Radio-Frequency Identity Protocols Generation-2 UHF RFID. `http://www.gs1.org/sites/default/files/docs/epc/Gen2_Protocol_Standard.pdf`, 2015. [Online; accessed 3-May-2018].

[36] Kenneth P Fishkin, Bing Jiang, Matthai Philipose, and Sumit Roy. I sense a disturbance in the force: Unobtrusive detection of interactions with rfid-tagged objects. In *International Conference on Ubiquitous Computing*, pages 268–282. Springer, 2004.

[37] Kenneth P Fishkin, Matthai Philipose, and Adam Rea. Hands-on rfid: Wireless wearables for detecting use of objects. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, pages 38–41. IEEE, 2005.

[38] G David Forney. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, 1973.

[39] Gartner. Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016. `https://www.gartner.com/newsroom/id/3598917/`, 2017. [Online; accessed 3-May-2018].

[40] Thierry Germa, Frédéric Lerasle, Noureddine Ouadah, and Viviane Cadenat. Vision and rfid data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding*, 114(6):641–651, 2010.

[41] Michael Goller, Christoph Feichtenhofer, and Axel Pinz. Fusing rfid and computer vision for probabilistic tag localization. In *RFID (IEEE RFID), 2014 IEEE International Conference on*, pages 89–96. IEEE, 2014.

[42] Saul Greenberg and Chester Fitchett. Phidgets: Easy development of physical interfaces through physical widgets. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 209–218, New York, NY, USA, 2001. ACM.

[43] Patrick J Grother, George W Quinn, and P Jonathon Phillips. Report on the evaluation of 2d still-image face recognition algorithms. *NIST interagency report*, 7709:106, 2010.

[44] SMARTRAC Technology Group. Smartrac uhf inlays and tags. https://www.smartrac-group.com/files/content/Products_Services/PDF/SMARTRAC_UHF_Inlays_and_Tags_Overview.pdf, 2018. [Online; accessed 3-May-2018].

[45] Wail Gueaieb and Md Suruz Miah. An intelligent mobile robot navigation technique using rfid technology. *IEEE Transactions on Instrumentation and Measurement*, 57(9):1908–1917, 2008.

[46] François Guimbretière. Paper augmented digital documents. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, pages 51–60, New York, NY, USA, 2003. ACM.

[47] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139–148. ACM, 2010.

[48] Gaëtanne Haché, Edward D Lemaire, and Natalie Baddour. Wearable mobility monitoring using a multimedia smartphone platform. *IEEE transactions on instrumentation and measurement*, 60(9):3153–3161, 2011.

[49] Dirk Hahnel, Wolfram Burgard, Dieter Fox, Ken Fishkin, and Matthai Philipose. Mapping and localization with rfid technology. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 1015–1020. IEEE, 2004.

[50] Derek Hao Hu, Sinno Jialin Pan, Vincent Wenchen Zheng, Nathan Nan Liu, and Qiang Yang. Real world activity recognition with multiple goals. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 30–39. ACM, 2008.

[51] Aki Harma, Martin F McKinney, and Janto Skowronek. Automatic surveillance of the acoustic activity in our living environment. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 4–pp. IEEE, 2005.

[52] Björn Hartmann, Scott R. Klemmer, Michael Bernstein, Leith Abdulla, Brandon Burr, Avi Robinson-Mosher, and Jennifer Gee. Reflective physical prototyping through integrated design, test, and analysis. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 299–308, New York, NY, USA, 2006. ACM.

[53] Md Kamrul Hasan, Husne Ara Rubaiyeat, Yong-Koo Lee, and Sungyoung Lee. A reconfigurable hmm for activity recognition. In *Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on*, volume 1, pages 843–846. IEEE, 2008.

[54] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[55] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[56] Jeremy M. Heiner, Scott E. Hudson, and Kenichiro Tanaka. Linking and messaging from real paper in the paper pda. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, UIST '99, pages 179–186, New York, NY, USA, 1999. ACM.

[57] Cory Hekimian-Williams, Brandon Grant, Xiuwen Liu, Zhenghao Zhang, and Piyush Kumar. Accurate localization of rfid tags using phase difference. In *RFID (IEEE RFID), 2010 IEEE International Conference on*, pages 89–96. IEEE, 2010.

[58] Jeffrey Hightower, Roy Want, and Gaetano Borriello. Spoton: An indoor 3d location sensing technology based on rf signal strength. *UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA*, 1, 2000.

[59] Steve Hodges, Alan Thorne, Hugo Mallinson, and Christian Floerkemeier. Assessing and optimizing the range of uhf rfid to enable real-world pervasive computing applications. In *International Conference on Pervasive Computing*, pages 280–297. Springer, 2007.

[60] David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. Paper windows: Interaction techniques for digital paper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 591–599, New York, NY, USA, 2005. ACM.

[61] Sherry Hsi and Holly Fait. Rfid enhances visitors' museum experience at the exploratorium. *Commun. ACM*, 48(9):60–65, September 2005.

[62] Scott E. Hudson and Jennifer Mankoff. Rapid construction of functioning physical interfaces from cardboard, thumbtacks, tin foil and masking tape. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 289–298, New York, NY, USA, 2006. ACM.

[63] Impinj. Application Note - Low Level User Data Support Impinj Support Portal. `https://support.impinj.com/hc/en-us/articles/202755318-Application-Note-Low-Level-User-Data-Support/`, 2015. [Online; accessed 3-May-2018].

[64] Impinj. Impinj Circular Polarity RFID Panel Antenna. `https://support.impinj.com/hc/en-us/article_attachments/200774708/IPJ-A1000-A1001-USA_Antenna_Spec.pdf`, 2015. [Online; accessed 3-May-2018].

[65] Impinj. Impinj Speedway Reader Product Brief Datasheet. https://support.impinj.com/hc/en-us/article_attachments/202942578/Impinj_SpeedwayReaders_ProductBrief_080715.pdf, 2015. [Online; accessed 3-May-2018].

[66] Impinj. Xarray. https://www.impinj.com/platform/connectivity/xarray/, 2018. [Online; accessed 3-May-2018].

[67] Stephen S Intille, Kent Larson, Emmanuel Munguia Tapia, Jennifer S Beaudin, Pallavi Kaushik, Jason Nawyn, and Randy Rockinson. Using a live-in laboratory for ubiquitous computing research. In *International Conference on Pervasive Computing*, pages 349–365. Springer, 2006.

[68] Asangi Jayatilaka and Damith C Ranasinghe. Towards unobtrusive real-time fluid intake monitoring using passive uhf rfid. In *RFID (IEEE RFID), 2016 IEEE International Conference on*, pages 1–4. IEEE, 2016.

[69] Bing Jiang, Kenneth P Fishkin, Sumit Roy, and Matthai Philipose. Unobtrusive long-range detection of passive rfid tag motion. *IEEE transactions on instrumentation and measurement*, 55(1):187–196, 2006.

[70] Haojian Jin, Zhijian Yang, Swarun Kumar, and Jason I Hong. Towards wearable everyday body-frame tracking using passive rfids. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):145, 2018.

[71] Walter Johnson, Herbert Jellinek, Leigh Klotz, Jr., Ramana Rao, and Stuart K. Card. Bridging the paper and electronic worlds: The paper user interface. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 507–512, New York, NY, USA, 1993. ACM.

[72] Sidney Kaiz. Assessing Self-maintenance: Activities of Daily Living, Mobility, and Instrumental Activities of Daily Living. *Journal of the American Geriatrics Society*, 31(12):721–727, dec 1983.

[73] Tai-Wei Kan, Chin-Hung Teng, and Wen-Shou Chou. Applying qr code in augmented reality applications. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry*, pages 253–257. ACM, 2009.

[74] Marije Kanis, Sean Alizadeh, Jesse Groen, Milad Khalili, Saskia Robben, Sander Bakkes, and Ben Kröse. Ambient monitoring from an elderly-centred design perspective: What, who and how. In *International Joint Conference on Ambient Intelligence*, pages 330–334. Springer, 2011.

[75] Hiroko Kato and Keng T Tan. Pervasive 2d barcodes for camera phone applications. *IEEE Pervasive Computing*, 6(4), 2007.

[76] Yoshihiro Kawahara, Steve Hodges, Benjamin S Cook, Cheng Zhang, and Gregory D Abowd. Instant inkjet circuits: lab-based inkjet printing to support rapid prototyping of ubicomp devices. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 363–372. ACM, 2013.

[77] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. Bringing gesture recognition to all devices. In *NSDI*, volume 14, pages 303–316, 2014.

[78] Cory D Kidd, Robert Orr, Gregory D Abowd, Christopher G Atkeson, Irfan A Essa, Blair MacIntyre, Elizabeth Mynatt, Thad E Starner, and Wendy Newstetter. The aware home: A living laboratory for ubiquitous computing research. In *International Workshop on Cooperative Buildings*, pages 191–198. Springer, 1999.

[79] Lito Kriara, Matthew Alsup, Giorgio Corbellini, Matthew Trotter, Joshua D Griffin, and Stefan Mangold. Rfid shakables: Pairing radio-frequency identification tags with the help of gesture recognition. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 327–332. ACM, 2013.

[80] Gierad Laput, Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. Em-sense: Touch recognition of uninstrumented, electrical and electromechanical objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 157–166. ACM, 2015.

[81] Inc Leap Motion. Leap motion. `https://www.leapmotion.com/`, 2018. [Online; accessed 3-May-2018].

[82] Johnny C. Lee, Daniel Avrahami, Scott E. Hudson, Jodi Forlizzi, Paul H. Dietz, and Darren Leigh. The calder toolkit: Wired and wireless components for rapidly prototyping interactive devices. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, pages 167–175, New York, NY, USA, 2004. ACM.

[83] Hanchuan Li, Eric Brockmeyer, Elizabeth J Carter, Josh Fromm, Scott E Hudson, Shwetak N Patel, and Alanson Sample. Paperid: A technique for drawing functional battery-free wireless interfaces on paper. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5885–5896. ACM, 2016.

[84] Hanchuan Li, Can Ye, and Alanson P Sample. Idsense: A human object interaction detection system based on passive uhf rfid. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2555–2564. ACM, 2015.

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

[86] Xinyu Li, Dongyang Yao, Xuechao Pan, Jonathan Johannaman, JaeWon Yang, Rachel Webman, Aleksandra Sarcevic, Ivan Marsic, and Randall S Burd. Activity recognition for medical teamwork based on passive rfid. In *RFID (IEEE RFID), 2016 IEEE International Conference on*, pages 1–9. IEEE, 2016.

[87] Sikun Lin, Hao Fei Cheng, Weikai Li, Zhanpeng Huang, Pan Hui, and Christoph Peylo. Ubii: Physical world interaction through augmented reality. *IEEE Transactions on Mobile Computing*, 16(3):872–885, 2017.

[88] Weiyao Lin, Ming-Ting Sun, Radha Poovandran, and Zhengyou Zhang. Human activity recognition for video surveillance. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pages 2737–2740. IEEE, 2008.

[89] Hui Liu, H. Darabi, P. Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067 –1080, nov. 2007.

[90] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille. A long-term evaluation of sensing modalities for activity recognition. In *International conference on Ubiquitous computing*, pages 483–500. Springer, 2007.

[91] Minghuang Ma, Haoqi Fan, and Kris M Kitani. Going deeper into first-person activity recognition. *arXiv preprint arXiv:1605.03688*, 2016.

[92] Wendy E Mackay, Daniele S Pagani, L Faber, B Inwood, P Launiainen, L Brenta, and V Pouzol. Ariel: augmenting paper engineering drawings. In *Conference companion on Human factors in computing systems*, pages 421–422. ACM, 1995.

[93] Wendy E. Mackay, Guillaume Pothier, Catherine Letondal, Kaare Bøegh, and Hans Erik Sørensen. The missing link: Augmenting biology laboratory notebooks. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*, UIST '02, pages 41–50, New York, NY, USA, 2002. ACM.

[94] Macy. Macy's inventory will be 100 percent rfid-tagged by 2017 supply chain dive. `http://www.supplychaindive.com/news/Macys-RFID-inventory-tracking/428937/`, 2017. [Online; accessed 3-May-2018].

[95] Sebastian Madgwick. An efficient orientation filter for inertial and inertial/magnetic sensor arrays. *Report x-io and University of Bristol (UK)*, 25, 2010.

[96] Nicolai Marquardt, Alex S Taylor, Nicolas Villar, and Saul Greenberg. Rethinking rfid: awareness and control for interaction with rfid systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2307–2316. ACM, 2010.

[97] Ross Messing, Chris Pal, and Henry Kautz. Activity recognition using the velocity histories of tracked keypoints. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 104–111. IEEE, 2009.

[98] Microsoft. Kinect for Windows. `https://dev.windows.com/en-us/kinect/`, 2015. [Online; accessed 3-May-2018].

[99] Microsoft. Hololens. `https://www.microsoft.com/en-us/hololens`, 2018. [Online; accessed 3-May-2018].

[100] R. Miesen, F. Kirsch, and M. Vossiek. Holographic localization of passive uhf rfid transponders. In *RFID (IEEE RFID), 2011 IEEE International Conference on*, pages 32 –37, april 2011.

[101] R. Miesen, F. Kirsch, and M. Vossiek. Uhf rfid localization based on synthetic apertures. *Automation Science and Engineering, IEEE Transactions on*, 10(3):807–815, July 2013.

[102] Pranav Mistry and Pattie Maes. Sixthsense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, page 11. ACM, 2009.

[103] MIT. House_n. `http://web.mit.edu/cron/group/house_n/`, 2010. [Online; accessed 3-May-2018].

[104] David Molyneaux and Hans Gellersen. Projected interfaces: enabling serendipitous interaction with smart tangible objects. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 385–392. ACM, 2009.

[105] Gordon E Moore. Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp. 114 ff. *IEEE solid-state circuits society newsletter*, 20(3):33–35, 2006.

[106] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004.

[107] Theresa Nick, Sebastian Cordes, Jurgen Gotze, and Werner John. Camera-assisted localization of passive rfid labels. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–8. IEEE, 2012.

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

[109] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang. Human activity detection and recognition for video surveillance. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, volume 1, pages 719–722. IEEE, 2004.

[110] Omar Oreifej and Zicheng Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In *Computer vision and pattern recognition (CVPR), 2013 IEEE conference on*, pages 716–723. IEEE, 2013.

[111] Alexandros Pantelopoulos and Nikolaos G Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):1–12, 2010.

[112] Sunhong Park and Shuji Hashimoto. Autonomous mobile robot navigation using passive rfid in indoor environment. *IEEE Transactions on Industrial Electronics*, 56(7):2366–2373, 2009.

[113] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.

[114] Siddika Parlak and Ivan Marsic. Detecting object motion using passive rfid: A trauma resuscitation case study. *Instrumentation and Measurement, IEEE Transactions on*, 62(9):2430–2437, 2013.

[115] Siddika Parlak, Ivan Marsic, and Randall S Burd. Activity recognition for emergency care using rfid. In *Proceedings of the 6th International Conference on Body Area Networks*, pages 40–46. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[116] Shwetak N Patel. *Infrastructure mediated sensing*. Georgia Institute of Technology, 2008.

[117] Donald J Patterson, Dieter Fox, Henry Kautz, and Matthai Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, pages 44–51. IEEE, 2005.

[118] Harry E Pence. Smartphones, smart objects, and augmented reality. *The Reference Librarian*, 52(1-2):136–145, 2010.

[119] Matthai Philipose. Large-scale human activity recognition using ultra-dense sensing. *The Bridge, National Academy of Engineering*, 35(4), 2005.

[120] Matthai Philipose, Kenneth P Fishkin, Mike Perkowitz, Donald J Patterson, Dieter Fox, Henry Kautz, and Dirk Hahnel. Inferring activities from interactions with objects. *IEEE pervasive computing*, 3(4):50–57, 2004.

[121] Matthai Philipose, Joshua R Smith, Bing Jiang, Alexander Mamishev, Sumit Roy, and Kishore Sundara-Rajan. Battery-free wireless identification and sensing. *IEEE Pervasive computing*, 4(1):37–45, 2005.

[122] Ronald Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.

[123] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Gesture recognition using wireless signals. *GetMobile: Mobile Computing and Communications*, 18(4):15–18, 2015.

[124] Daniele Puccinelli and Martin Haenggi. Wireless sensor networks: applications and challenges of ubiquitous sensing. *Circuits and Systems Magazine, IEEE*, 5(3):19–31, 2005.

[125] Jie Qi and Leah Buechley. Sketching in circuits: designing and building electronics on paper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1713–1722. ACM, 2014.

[126] Rajeev Ranjan, Vishal M Patel, and Rama Chellappa. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[127] Zulqarnain Rashid, Enric Peig, and Rafael Pous. Bringing online shopping experience to offline retail through augmented reality and rfid. In *Internet of Things (IOT), 2015 5th International Conference on the*, pages 45–51. IEEE, 2015.

[128] Ruth Ravichandran, Elliot Saba, Ke-Yu Chen, Mayank Goel, Sidhant Gupta, and Shwetak N Patel. Wibreathe: Estimating respiration rate using wireless signals in natural settings in the home. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 131–139. IEEE, 2015.

[129] Shaoqing Ren, Kaiming He, Ross Girshick, Xiangyu Zhang, and Jian Sun. Object detection networks on convolutional feature maps. *IEEE transactions on pattern analysis and machine intelligence*, 39(7):1476–1481, 2017.

[130] Xiaofeng Ren and Matthai Philipose. Egocentric recognition of handled objects: Benchmark and analysis. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2009.

[131] Marc Rettig. Prototyping for tiny fingers. *Communications of the ACM*, 37(4):21–27, 1994.

[132] A. Rida, Li Yang, R. Vyas, and M.M. Tentzeris. Conductive inkjet-printed antennas on flexible low-cost paper-based substrates for rfid and wsn applications. *Antennas and Propagation Magazine, IEEE*, 51(3):13–23, June 2009.

[133] Amin Rida, Li Yang, Rushi Vyas, and Manos M Tentzeris. Conductive inkjet-printed antennas on flexible low-cost paper-based substrates for rfid and wsn applications. *IEEE Antennas and Propagation Magazine*, 51(3), 2009.

[134] Lawrence G Roberts. Aloha packet system with and without slots and capture. *ACM SIGCOMM Computer Communication Review*, 5(2):28–42, 1975.

[135] Mahsan Rofouei, Andrew Wilson, A.J. Brush, and Stewart Tansley. Your phone or mine?: Fusing body, touch and device sensing for multi-user device-display interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1915–1918, New York, NY, USA, 2012. ACM.

[136] Max Roser and Hannah Ritchie. "technological progress" published online at ourworldindata.org. `https://ourworldindata.org/technological-progress`, 2018. [Online; accessed 3-May-2018].

[137] Alanson Sample, Craig Macomber, Liang-Ting Jiang, and Joshua R Smith. Optical localization of passive uhf rfid tags with integrated leds. In *RFID (RFID), 2012 IEEE International Conference on*, pages 116–123, April 2012.

[138] Alanson P Sample, Craig Macomber, Liang-Ting Jiang, and Joshua R Smith. Optical localization of passive uhf rfid tags with integrated leds. In *RFID (IEEE RFID), 2012 IEEE International Conference on*, pages 116–123. IEEE, 2012.

[139] Alanson P Sample, Daniel J Yeager, Pauline S Powledge, Alexander V Mamishev, and Joshua R Smith. Design of an rfid-based battery-free programmable sensing platform. *IEEE transactions on instrumentation and measurement*, 57(11):2608–2615, 2008.

[140] Alanson P Sample, Daniel J Yeager, Pauline S Powledge, and Joshua R Smith. Design of a passively-powered, programmable sensing platform for uhf rfid systems. In *RFID (IEEE RFID), 2007. IEEE International Conference on*, pages 149–156. IEEE, 2007.

[141] Simo Sarkka, Ville V Viikari, Miika Huusko, and Kaarle Jaakkola. Phase-based uhf rfid tracking with nonlinear kalman filtering and smoothing. *IEEE Sensors Journal*, 12(5):904–910, 2012.

[142] Stefan Schwarzer, Martin Vossiek, Markus Pichler, and Andreas Stelzer. Precise distance measurement with ieee 802.15. 4 (zigbee) devices. In *Radio and Wireless Symposium, 2008 IEEE*, pages 779–782. IEEE, 2008.

[143] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[144] Joshua R Smith, Kenneth P Fishkin, Bing Jiang, Alexander Mamishev, Matthai Philipose, Adam D Rea, Sumit Roy, and Kishore Sundara-Rajan. Rfid-based techniques for human-activity detection. *Communications of the ACM*, 48(9):39–44, 2005.

[145] Joshua R Smith, Bing Jiang, Sumit Roy, Matthai Philipose, Kishore Sundara-Rajan, and Alexander Mamishev. Id modulation: Embedding sensor data in an rfid timeseries. In *International Workshop on Information Hiding*, pages 234–246. Springer, 2005.

[146] Joshua R Smith, Alanson P Sample, Pauline S Powledge, Sumit Roy, and Alexander Mamishev. A wirelessly-powered platform for sensing and computation. In *International Conference on Ubiquitous Computing*, pages 495–506. Springer, 2006.

[147] Morgan Stanley. Morgan stanley releases the mobile internet report. *https://www.morganstanley.com/press-releases/morgan-stanley-releases-the-mobile-internet-report/*, 2009.

[148] Jürgen Steimle. Designing pen-and-paper user interfaces for interaction with documents. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 197–204. ACM, 2009.

[149] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 33–40, New York, NY, USA, 2006. ACM.

[150] Johannes A Stork, Luciano Spinello, Jens Silva, and Kai O Arras. Audio-based human activity recognition using non-markovian ensemble voting. In *RO-MAN, 2012 IEEE*, pages 509–514. IEEE, 2012.

[151] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[152] V. Cadenat T. Germa, F. Lerasle N. Ouadah. Vision and rfid data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding*, 114(Issue 6):Pages 641651, jun 2010.

[153] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.

[154] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *International Conference on Pervasive Computing*, pages 158–175. Springer, 2004.

[155] Alien Technology. ALN-9713 SIT Inlay. `http://www.alientechnology.com/wp-content/uploads/Alien-Technology-Higgs-4-ALN-9713-SIT.pdf`, 2014. [Online; accessed 3-May-2018].

[156] Alien Technology. ALN-9740 Squiggle Inlay (Higgs 4). http://www.alientechnology.com/wp-content/uploads/Alien-Technology-Higgs-4-ALN-9740-Squiggle.pdf, 2016. [Online; accessed 3-May-2018].

[157] Alien Technology. ALN-9768 Wonder Dog. http://www.alientechnology.com/wp-content/uploads/ALN-9768%20WD%20Higgs4%20(2016-07-05).pdf, 2016. [Online; accessed 3-May-2018].

[158] Alien Technology. Alien family of EPC Gen 2 RFID Inlays. http://www.alientechnology.com/wp-content/uploads/Tag%20Family%20Overview%20(2015-02-03b).pdf, 2018. [Online; accessed 3-May-2018].

[159] Roberto L Shinmoto Torres, Damith C Ranasinghe, Qinfeng Shi, and Alanson P Sample. Sensor enabled wearable rfid technology for mitigating the risk of falls near beds. In *RFID (RFID), 2013 IEEE International Conference on*, pages 191–198. IEEE, 2013.

[160] Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 1–9. ACM, 2008.

[161] Sarvesh Vishwakarma and Anupam Agrawal. A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, 29(10):983–1009, 2013.

[162] Shrenik A Vora, William M Mongan, Endla K Anday, Kapil R Dandekar, Genevieve Dion, Adam K Fontecchio, and Timothy P Kurzweg. On implementing an unconventional infant vital signs monitor with passive rfid tags. In *RFID (IEEE RFID), 2017 IEEE International Conference on*, pages 47–53. IEEE, 2017.

[163] Tomotaka Wada, Norie Uchitomi, Yuuki Ota, Toshihiro Hori, Kouichi Mutsuura, and Hiromi Okada. A novel localization scheme for passive rfid tags; communication range recognition (crr). In *RFID (IEEE RFID), 2009 IEEE International Conference on*, pages 163–169. IEEE, 2009.

[164] Edward J Wang, Tien-Jui Lee, Alex Mariakakis, Mayank Goel, Sidhant Gupta, and Shwetak N Patel. Magnifisense: Inferring device interaction using wrist-worn passive magneto-inductive sensors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 15–26. ACM, 2015.

[165] Jue Wang and Dina Katabi. Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 51–62. ACM, 2013.

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

[167] Limin Wang, Zhe Wang, Yu Qiao, and Luc Van Gool. Transferring deep object and scene representations for event recognition in still images. *International Journal of Computer Vision*, 126(2-4):390–409, 2018.

[168] Jamie A Ward, Paul Lukowicz, Gerhard Troster, and Thad E Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1553–1567, 2006.

[169] Mark Weiser. The computer for the twenty-first century (pp. 94-100). *Scientific American, September Issue*, 1991.

[170] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision*, pages 499–515. Springer, 2016.

[171] Karl DD Willis, Takaaki Shiratori, and Moshe Mahler. Hideout: mobile projector interaction with tangible objects and surfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, pages 331–338. ACM, 2013.

[172] Andrew D Wilson and Hrvoje Benko. Crossmotion: fusing device and image motion for user identification, tracking and device association. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 216–223. ACM, 2014.

[173] Paul Wilson, Daniel Prashanth, and Hamid Aghajan. Utilizing rfid signaling scheme for localization of stationary objects and speed estimation of mobile objects. In *RFID (IEEE RFID), 2007. IEEE International Conference on*, pages 94–99. IEEE, 2007.

[174] Chenshu Wu, Zheng Yang, Zimu Zhou, Xuefeng Liu, Yunhao Liu, and Jiannong Cao. Non-invasive detection of moving and stationary human with wifi. *IEEE Journal on Selected Areas in Communications*, 33(11):2329–2342, 2015.

[175] Lu Xia and JK Aggarwal. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2834–2841. IEEE, 2013.

[176] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In *NSDI*, pages 71–84, 2013.

[177] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 237–248, New York, NY, USA, 2014. ACM.

[178] Li Yang, A. Rida, R. Vyas, and M.M. Tentzeris. Rfid tag and rf structures on a paper substrate using inkjet-printing technology. *Microwave Theory and Techniques, IEEE Transactions on*, 55(12):2894–2901, Dec 2007.

[179] Li Yang, Amin Rida, Rushi Vyas, and Manos M Tentzeris. Rfid tag and rf structures on a paper substrate using inkjet-printing technology. *IEEE Transactions on Microwave Theory and Techniques*, 55(12):2894–2901, 2007.

[180] Koji Yatani and Khai N Truong. Bodyscope: a wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 341–350. ACM, 2012.

[181] Polle T. Zellweger, Niels Olof Bouvin, Henning Jehøj, and Jock D. Mackinlay. Fluid annotations in an open world. In *Proc. Hypertext 2001*, pages 9–18. ACM Press, 2001.

[182] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*, pages 197–205. IEEE, 2014.

[183] Hong Zhang, Jeremy Gummeson, Benjamin Ransford, and Kevin Fu. Moo: A batteryless computational rfid and sensing platform. *University of Massachusetts Computer Science Technical Report UM-CS-2011-020*, 2011.

[184] Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li. A review on human activity recognition using vision-based method. *Journal of healthcare engineering*, 2017, 2017.

[185] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108. ACM, 2016.

[186] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495, 2014.

[187] Yongpan Zou, Jiang Xiao, Jinsong Han, Kaishun Wu, Yun Li, and Lionel M Ni. Grfid: A device-free rfid-based gesture recognition system. *IEEE Transactions on Mobile Computing*, 16(2):381–393, 2017.