Device-free Human Localization and Activity Recognition for Supporting the Independent Living of the Elderly

Wenjie Ruan

School of Computer Science

The University of Adelaide

This dissertation is submitted for the degree of

Doctor of Philosophy

Supervisors: Prof. Michael Sheng, A/Prof. Nickolas J.G. Falkner

Dr. Lina Yao and Prof. Xue Li

December 2017

To my mother and father,

my wife and my newborn baby,

my brother,

who made all of this possible,

for their endless encouragement and patience.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Wenjie Ruan

December 2017

Acknowledgements

I would like to thank all the people who have ever helped, supported and advised me in one way or another. Without them, this thesis would never be possible.

First of all, my biggest gratitude goes to my PhD supervisor Prof. Michael Sheng. Without his encouragement three years ago, I would not have had a chance to pursue my PhD at the University of Adelaide. Without his continuous and selfless guidance, I would never have accomplished my PhD research within three years. It is my great honor and luck to have Michael being my PhD supervisor. He has always been patient, passionate, encouraging throughout the whole journey of my PhD study. His many insightful suggestions and comments on my research have significantly improved the work in this thesis.

Secondly, I would like to greatly thank A/Prof. Nickolas J.G. Falkner, who has provided me countless valuable suggestions in academic writing. His many insights regarding how to deliver a native and coherent research idea have greatly increased the quality of this thesis. I also would like to sincerely thank Dr. Lina Yao. During the first year of my PhD journey, her enormous passion on research strongly inspired and encouraged me, leading me to the "door" of academic research. Her supervision with decisiveness and assertiveness has greatly sped up the progress of my PhD project. Moreover, I would like to give my enormous and sincere thanks to Prof. Xue Li, who has provided me many valuable suggestions and strong supports in both my PhD research and job hunting.

Thirdly, I would like to sincerely thank A/Prof. Tao Gu, who has provided many valuable suggestions and assistance on stepping in the research area of pervasive computing. I also

want to express heartfelt thanks to Dr. Lei Yang, Dr. Longfei Shangguan and Ms. Peipei Xu for their generous help and valuable discussion in the research. I also want to express my thanks to Dr. Yongrui Qin, who has offered me so many assistance on living in Adelaide. Moreover, I would like to thank other members in our lab, including Dr. Xianzhi Wang, Dr. Ali Shemshadi, Abdullah Alfazi, Dr. Yihong Zhang, Dr. Wei Zhang, Xiu Fang, Tran Khoi Nguyen, and Zhigang Lu for the valuable discussions in group readings and their companionship throughout the PhD journey.

Fourthly, I would like to express my sincere appreciation to University of Adelaide and Australia Research Council for funding my PhD study. I also deeply thank Tsinghua National Lab for Information Science and Technology and Dr. Lei Yang for accepting me and supporting my six-month visiting there. Being a visiting PhD student there has significantly broaden my research horizon and contributed to this thesis.

Lastly, I would express my countless thanks to my parents and my little brother, for their love and support. I am also deeply grateful to my wife, who has accompanied me in University of Adelaide during last two years of my PhD journey. She not only looked after my living but also joined my research, providing me enormous assistance in experiments and data analysis. It is my fortune to have her in my life. Finally, this thesis is the best gift to my newborn baby, wishing he has a colorful and meaningful life.

Abstract

Given the continuous growth of the aging population, the cost of health care, and the preference that the elderly want to live independently and safely at their own homes, the demand on developing an innovative living-assistive system to facilitate the independent living for the elderly is becoming increasingly urgent. This novel system is envisioned to be *device-free*, *intelligent*, and *maintenance-free* as well as deployable in a residential environment. The key to realizing such envisioned system is to study low cost sensor technologies that are practical for device-free human indoor localization and activity recognition, particularly under a clustered residential home. By exploring the latest, low-cost and unobtrusive RFID sensor technology, this thesis intends to design a new device-free system for better supporting the independent living of the elderly. Arising from this live-assistive system, this thesis specifically targets the following *six* research problems.

Firstly, to deal with severe missing readings of passive RFID tags, this thesis proposes a novel tensor-based low-rank sensor reading recovery method, in which we formulate RFID sensor data as a high-dimensional tensor that can naturally preserve sensors' spatial and temporal information. Secondly, by purely using passive RFID hardware, we build a novel *data-driven* device-free localization and tracking system. We formulate human localization problem as finding a location with the maximum posterior probability given the observed RSSIs (Received Signal Strength Indicator) from passive RFID tags. For tracking a moving target, we mathematically model the task as searching a location sequence with the most likelihood under a Hidden Markov Model (HMM) framework. Thirdly, to tackle

the challenge that the tracking accuracy decreases in a cluttered residential environment, we propose to leverage the Human-Object Interaction (HOI) events to enhance the performance of the proposed RFID-based system. This idea is motivated by an intuition that HOI events, detected by pervasive sensors, can potentially reveal people's interleaved locations during daily living activities such as watching TV or opening the fridge door.

Furthermore, to recognize the resident's daily activities, we propose a device-free human activity recognition (HAR) system by deploying the passive RFID tags as an array attached on the wall. This HAR system operates by learning how RSSIs are distributed when a resident performs different activities. Moreover, considering that falls are among the leading causes of hospitalization for the elderly, we develop a fine-grained fall detection system that is capable of not only recognizing regular actions and fall events simultaneously, but also sensing the fine-grained fall orientations. Lastly, to remotely control the smart electronic appliances equipped in an intelligent environment, we design a device-free multi-modal hand gesture recognition (HGR) system that can accurately sense the hand's in-air speed, waving direction, moving range and duration around a mobile device. Our system transforms an electronic device into an active sonar system that transmits an inaudible audio signal via the speaker and decodes the echoes of the hand at its microphone.

To test the proposed systems and approaches, we conduct an intensive series of experiments in several real-world scenarios by multiple users. The experiments demonstrate that our RFID-based system can localize a resident with average 95% accuracy and recognize 12 activities with nearly 99% accuracy. The proposed fall detection approach can detect 90.8% falling events. The designed HGR system can recognize six hand gestures with an accuracy up to 96% and provide more fine-grained control commands by incorporating hand motion attributes.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Fiona's frail 77-year-old father lives alone in a small apartment. He is making a cup of tea and his kitchen knows it. Tiny sensors monitor his every move and track each tea-making step. If he pauses for too long, a nearby computer reminds him about what to do next. Later that day, Fiona accesses a secure website and scans a checklist, which was created from the computer in her father's apartment. She finds that her father took his medicine on schedule, ate normally, and continued to manage his daily activities on his own. This puts Fiona's mind at ease.

With recent developments in cheap sensor and networking technologies, it has become possible to develop a wide range of valuable applications such as the remote health monitoring and intervention depicted above. These applications offer the potential to enhance the quality of life for the elderly, afford them a greater sense of security, and facilitate independent living. For example, by monitoring the daily routines of a person with dementia, an elder assistant service can track how completely and consistently the daily routines are performed, and determine when the resident needs assistance.

Central to realizing these applications is the study of low cost sensor technologies that are practical for human indoor localization and activity recognition, particularly for the elderly. However, existing approaches either rely on body-worn sensors to detect human

locations and activities, or dense sensing where low cost sensors (*e.g.,* wireless transceivers) are attached to objects and people's activities can be indirectly inferred from their interactions with the objects. In the former approaches, battery powered sensors are normally bigger in size, expensive, and require maintenance and user involvement (*e.g.,* wearing the device). In the latter approaches, sensors are typically cheaper and maintenance free. However, user involvement is still needed (*e.g.,* wearing a bracelet to detect objects). All these technologies are not very practical, especially for monitoring aged people with dementia, or even just those with mild cognitive impairment.

As a result, to tackle this challenge, this Ph.D. thesis intends to develop a *device-free*, *intelligent*, and *maintenance-free* system to better support the *independent living* of the elderly. This system should bear at least the following three promising characteristics: *i)* Device-free - it does not require the user to wear any devices or sensors at any circumstance; *ii)* Intelligent - it should automatically understand the user's daily living routines and activities, as well as timely and accurately recognize abnormal actions and provide useful assistance when necessary; and *iii)* maintenance-free - such a system should be light in both weight and size, as cheap as possible and require no human maintenance.

Recent advancement in low-cost passive Radio-Frequency Identification (RFID) tags makes device-free indoor localization and activity recognition possible. They are maintenance-free (no batteries in tags) and inexpensive (about 5 cents each and still dropping quickly). This thesis proposes a novel system for automated human indoor location and activity discovery and monitoring by deploying low-cost, unobtrusive passive RFID tags in a full-furnitured residential home. Also, by taking the recent advances of supervised machine learning and tensor theory, we first introduce a low-rank tensor completion method to deal with the reading loss of passive RFID tags and then propose a novel anomaly detection method to achieve a fine-grained fall detection that not only can recognize regular actions and fall events simultaneously but also distinguish different fall orientations. Moreover, to conveniently

control the smart electronic appliances in a smart home, we design a device-free multi-modal HGR system that can provide up to 162 control commands for various applications.

In the next section, we will first detail the system infrastructure and then identify the challenges of realizing this system. Furthermore, we illustrate how to deal with those challenges by decomposing this system into six research issues, as well as listing the research papers published for each part to demonstrate the effectiveness and novelty of our proposed approaches and models.

1.1 System Overview

As Fig. 1.1 shows, we propose a conceptual system infrastructure that is mainly built upon cheap and maintenance-free passive RFID hardware. The whole system consists of four main modules - Hardware Layer, Discovery Layer, Monitoring Layer and Application Layer.

- Hardware Layer: this layer is the hardware infrastructure of the whole system, in which we mainly deploy passive RFID tags in a residential environment, plus a few other commercialized sensors (*e.g.,* pressure sensor, proximity sensor and light sensor) on the domestic electronic appliances with a primary aim of detecting human-object interaction events.

- Discovery Layer: this layer is the key component of the system. Its main function is to automatically and accurately recognize and discover the user's locations and activities by using novel machine learning approaches to mine and analyze RFID and sensor readings collected from the Hardware Layer.

- Monitoring Layer: this module continuously records and tracks user's daily routines and activities, as well as performing context-aware, learning-based abnormal activity reasoning (*e.g.,*, falling down, lying-down or siting for an unusual long time) in a real-time manner.

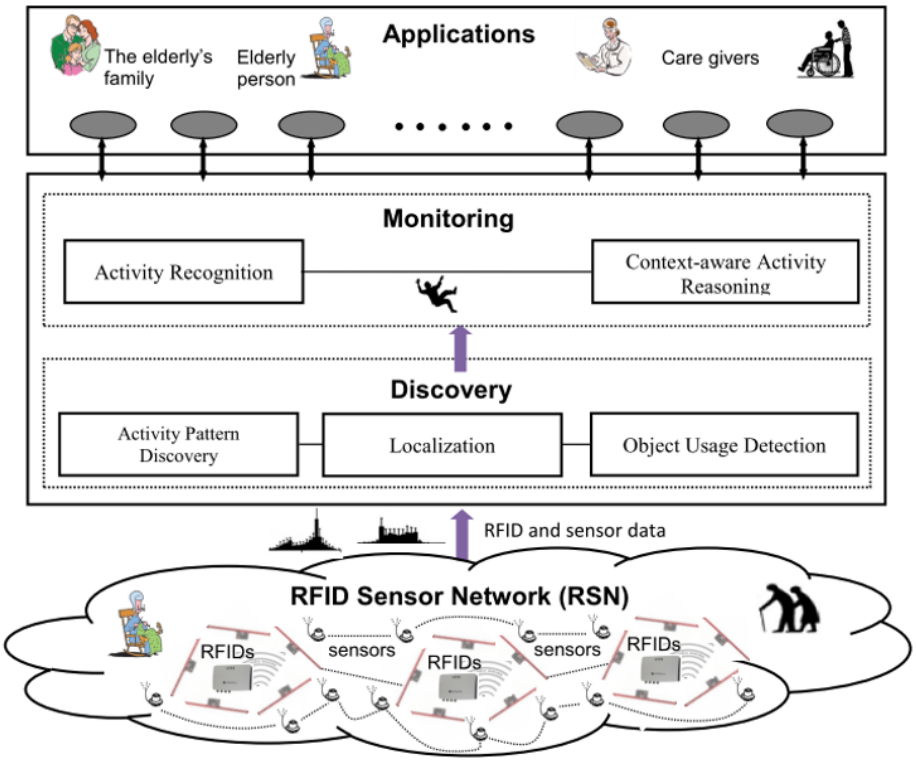


Fig. 1.1 The overall conceptual framework of the proposed system

- Application Layer: this layer provides useful knowledges or decision information for various kinds of real-world agents (*e.g.,* user's children, hospital, emergency service or aged caring institutes).

1.2 Challenges

However, transforming the above ideal system-concepts into a practical system that is workable in real-world residential environments requires us to deal with several non-trivial challenges.

First of all, to make our system lightweight, maintenance-free and as cheap as possible, we mainly use passive RFID tags (battery-free, extremely cheap, around 5 cents each; very tiny size, around 5cm $\times$ 1cm) in the Hardware Layer. However, since passive RFID tags can

only energized by harvesting the backscattered RF (Radio Frequency) signal, their signals are very weak and thus suffer significant reading loss and distortion. Those missing values will not only decrease accuracy of localization and activity recognition in the Discovery Layer but also compromise the real-time user-daily-routine monitoring and abnormality reasoning in the Monitoring Layer. As a result, how to efficiently yet accurately recover the missing sensor values is our first challenge. Secondly, given the weak RFID sensor readings, how to develop novel machine learning algorithms for accurately recognizing user's locations and activities also deserves a careful consideration, especially in a clustered residential environment where household furniture and electronic appliances strongly affect the sensor signals. More challengingly, how can we accurately yet robustly recognize abnormal activities of users in a real-time manner? In particular, we need to carefully deal with how to enable a fine-grained abnormality detection (*e.g.,* distinguishing different falling directions).

In the next section, we briefly introduce the key chapters of this thesis which deal with those challenges and realize the core functionalities of the proposed supporting system from six different research points of views.

1.3 Summaries of Key Chapters

In this thesis, we illustrate our solutions and methods from six research perspectives, detailed as follows:

1.3.1 Recovering Missing Readings for Corrupted Sensor Data via Low-Rank Tensor Completion

Passive RFID tags attached on the walls of a residential house usually generate RSSI readings with both time-stamps and geo-tags. Such type of data usually have shown complex spatio-temporal correlation and are easily missing in practice due to communication failure or

furniture obstruction. In Chapter 3, we aim to tackle the challenge – how to accurately and efficiently recover the missing values for corrupted spatio-temporal sensor data. In particular, we first formulate such sensor data as a high-dimensional tensor that can naturally preserve sensors' both geographical and time information, which we call a *spatio-temporal Tensor*. Then we model the sensor data recovery as a low-rank robust tensor completion problem by exploiting its latent low-rank structure and sparse noise property. To solve this optimization problem, we design a highly efficient optimization method that combines the alternating direction method of multipliers and accelerated proximal gradient to minimize the tensor's convex surrogate and noise's ℓ_1-norm. We test our proposed method by a synthetic dataset and a real-world sensor-array testbed built by passive RFID tags. The key research papers related with this part are listed as follows:

[C1] **W. Ruan**, P. Xu, Q. Z. Sheng, N. Falkner, X. Li, and W. E. Zhang, Recovering Missing Values from Corrupted Spatio-Temporal Sensory Data via Robust Low-Rank Tensor Completion, *The 22nd Int. Conference on Database Systems for Advanced Applications (DASFAA'17)*, Suzhou, China, Mar 27-30, 2017. [*ERA/CORE A, Full Research Paper, Acceptance Rate = 24.3%, Oral Presentation*]

[C2] **W. Ruan**, P. Xu, Q. Z. Sheng, N.K. Tran, N. Falkner, X. Li, and W.E. Zhang, When Sensor Meets Tensor: Filling Missing Sensor Values Through a Tensor Approach, *The 25th ACM Conference on Information and Knowledge Management (CIKM'16)*, Indianapolis, USA, Oct 24-28, 2016. [*ERA/CORE A, Acceptance Rate = 24%*]

1.3.2 Device-free Human Localization and Tracking Using Passive RFID Tags

Device-free Passive (DfP) human localization and tracking is one of the key components in the proposed system. It is promising in two aspects: *i)* it neither requires residents to wear any sensors or devices, *ii)* nor needs them to consciously cooperate during the localization. In Chapter 4, we build a novel *data-driven* DfP localization and tracking system upon a set

of commercial UHF (Ultra-High Frequency) passive RFID tags in an indoor environment. In particular, we formulate human localization problem as finding a location with the maximum posterior probability given the observed RSSIs. We propose a series of localization schemes to capture the posterior probability by taking the advance of supervised-learning models including Gaussian Mixture Model (GMM), k Nearest Neighbor (kNN) and Kernel-based Learning. For tracking a moving target, we mathematically model the task as searching a location sequence with the most likelihood, in which we first augment the probabilistic estimation learned in localization to construct the Emission Matrix and propose two human mobility models to approximate the Transmission Matrix in HMM. The proposed HMM-based tracking model is able to transfer the pattern learned in localization into tracking but also reduce the location-state candidates at each transmission iteration, which increases both the computation efficiency and tracking accuracy. The extensive experiments in two real-world scenarios reveal that our approach can achieve up to 94% localization accuracy and an average 0.64m tracking error, outperforming other state-of-the-art RFID-based indoor localization systems. The key research papers related with this part are listed as follows:

[C3] **W. Ruan**, L. Yao, Q. Z. Sheng, N. Falkner, and X. Li, TagTrack: Device-free Localization and Tracking Using Passive RFID Tags, *The 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous'14)*, London, UK, Dec 2-5, 2014. [*ERA/CORE A, Full Research Paper, Acceptance Rate = 18.1%, Oral Presentation; This work also won* ***Highly Commended Research Poster Award*** *in The 25th Australia Database Conference (ADC'14) PhD School in Big Data*]

[C4] L. Yao, **W. Ruan**, Q. Z. Sheng, X. Li, and N. Falkner, Exploring Tag-free RFID-based Passive Localization and Tracking via Learning-based Probabilistic Approaches, *The 23rd ACM International Conference on Information and Knowledge Management (CIKM'14)*, Shanghai, China, Nov. 3-7, 2014. [*ERA/CORE A, Acceptance Rate = 21.9%*]

[J1] **W. Ruan**, Q. Z. Sheng, L. Yao, X. Li, N. Falkner, *etc.*, Device-free Human Localization and Tracking with UHF Passive RFID Tags: A Data-driven Approach, *Journal of Network and Computer*

Applications (JNCA), Under 2nd revision.

[*ERA A, Impact Factor = 3.5, Extended Version of MobiQuitous'14*]

1.3.3 Enhanced Device-free RFID-based Indoor Localization and Tracking through Human-Object Interactions

In a cluttered environment such as a residential home, RSSIs are heavily obstructed by furniture or metallic appliances. Thus the tracking precision of the passive RFID-based system greatly deceases. However, on the other side, this residential environment is important to observe as human-object interaction (HOI) events, detected by pervasive sensors, can potentially reveal people's interleaved locations during daily living activities, such as watching TV or opening the fridge door. In Chapter 5, to deal with the accuracy degradation in a fully furnished environment, we propose a general Bayesian probabilistic framework to integrate both RSSI signals and HOI events to infer the most likely location and trajectory. By leveraging the HOI contexts, the proposed approach significantly enhances the localization and tracking accuracy of the original system. Experiments conducted in a residential house demonstrate the effectiveness of our proposed method, in which we can localize a resident with average 95% accuracy and track a moving subject with 0.58m mean error distance. The key research papers related with this part are listed as follows:

[C5] **W. Ruan**, Q. Z. Sheng, L. Yao, T. Gu, M. Ruta and L. Shangguan, Device-free Indoor Localization and Tracking through Human-Object Interactions, *The IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'16)*, Coimbra, Portugal, June 21-24, 2016. [*ERA/CORE A, Full Research Paper, Acceptance Rate = 19.5%, Oral Presentation*]

[C6] **W. Ruan**, Q. Z. Sheng, L. Yao, L. Yang and T. Gu, HOI-Loc: Towards Unobtrusive Human Localization with Probabilistic Multi-Sensor Fusion, *The 14th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'16)*, WiP Track, Sydney, Australia, March 14-18, 2016. [*ERA A, CORE A*, 1 of 4* ***Nominees for Best WiP Poster Award***]

1.3.4 Device-free Human Activity Recognition based on Passive RFID Tag-Array

Human activity recognition is another fundamental functionality in our proposed system. It usually requires an intelligent environment to successfully infer what a person is doing or attempting to do. In Chapter 6, we propose a device-free activity recognition approach by deploying the low cost, passive RFID tags as an array attached on the wall. HAR in our system is achieved by learning how RSSIs from the passive RFID tag-array are distributed when a person performs different daily activities. We also systematically explore the impacts of tag number and locations on the recognition accuracy. Furthermore, we propose a novel tag selection method to choose the optimal subset of RFID tags in the array. To deal with the uncertainty in RSSIs caused by the changes of different human activities, we propose the Dirichlet process Gaussian Mixture Model (DPGMM) based HMM to model the transition process from one activity to another activity. We conduct extensive experiments consisted by 12 orientation-sensitive activities and a series of activity sequences in a lab environment and a residential home. The experimental results demonstrate that our proposed approach can distinguish a series of orientation sensitive postures with high accuracy in both environments. The experimental results demonstrate the high accuracy of our RFID-based device-free HAR approach. The key research papers related with this part are listed as follows:

[C7] **W. Ruan**, L. Chea, Q. Z. Sheng, and L. Yao, Recognizing Daily Living Activity Using Embedded Sensors in Smartphones: A Data-Driven Approach, *The 17th International Conference on Advanced Data Mining and Applications, (ADMA'16)*, Gold Coast, Australia, Dec 12-15, 2016. [*ERA/CORE B, Spotlight Paper, Acceptance Rate = 17%, Oral Presentation,* ***Best Student Paper Runner-Up***]

[C8] **W. Ruan**, Unobtrusive Human Localization and Activity Recognition for Supporting Independent Living of the Elderly, *The 14th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'16)*, PhD Forum, Sydney, Australia, March 14-18, 2016.
[*ERA A, CORE A*, Oral Presentation*]

[C9] L. Yao, Q. Z. Sheng, **W. Ruan**, T. Gu, N. Falkner, X. Li and Z. Yang, RF-Care: Device-free Posture Monitoring of Elderly People Using a Passive RFID Tag Array, *The 12th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous'15)*, Coimbra, Portugal, July 22-24, 2015. [*ERA/CORE A, Full Research Paper, Acceptance Rate = 27.9%, Oral Presentation*]

[C10] L. Yao, Q. Z. Sheng, **W. Ruan**, X. Li, S. Wang, Z. Yang and W. Zou, Device-Free Posture Recognition via Online Learning of Multi-Dimensional RFID Received Signal Strength, *The 21st IEEE International Conference on Parallel and Distributed Systems (ICPADS'15)*, Melbourne, Australia, Dec 14 - 17, 2015. [*ERA/CORE B, Full Research Paper, Oral Presentation*]

[C11] L. Yao, Q. Z. Sheng, X. Li, S. Wang, T. Gu, **W. Ruan** and W. Zou, Freedom: Online Activity Recognition via Dictionarybased Sparse Representation of RFID Sensing Data, *IEEE Intl. Conference on Data Mining (ICDM'15)*, Atlantic, USA, Nov 14 - 17, 2015.
[*ERA A, CORE A*, Acceptance Rate = 18.2%, Oral Presentation*]

1.3.5 Enabling the Fine-grained Device-free Fall Detection

Falls are among the leading causes of hospitalization for the elderly and illness individuals. Considering that the elderly often live alone and receive only irregular visits, it is essential to develop such a system that can effectively detect a fall or abnormal activities. In Chapter 7, we propose a device-free, fine-grained fall detection approach based on pure passive ultra-high frequency RFID tags, which not only is capable of sensing regular actions and fall events simultaneously, but also provide caregivers the contexts of fall orientations. In particular, we first augment the Angle-based Outlier Detection Method (ABOD) to classify normal actions (*e.g.,* standing, sitting, lying and walking) and detect a fall event. Once a fall event is detected, we then segment a fix-length RSSI data stream generated by the fall and then utilize DTW based *k*NN to distinguish the falling direction. The experimental results demonstrate that our proposed approach can distinguish the normal daily activities before a fall, as well

as the fall orientations with more than 90% accuracy. The key research papers related with this part are listed as follows:

[C12] **W. Ruan**, L. Yao, Q. Z. Sheng, N. Falkner, X. Li, and T. Gu., TagFall: Towards Device-free, Fine-grained Fall Detection based on UHF Passive RFID Tags, *The 12th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous'15)*, Coimbra, Portugal, July 22-24, 2015. [*ERA/CORE A, Full Research Paper, Acceptance Rate = 27.9%, Oral Presentation; This work also won* ***Best Poster Award*** *in The 9th ACM International Workshop on IoT and Cloud Computing*]

1.3.6 Realizing Human-Machine Interactions Using Touch-free Hand Gestures

Another important issue in an intelligent residential home is how to accurately and conveniently control the domestic electronic appliances equipped (*e.g.,* automated window curtain, brightness-adjustable lamp, TV and air conditioner). For example, we enter a smart house and turn on the TV by simply waving a hand in the air, then we can use another hand gesture to turn on the Air Conditioner as well, furthermore, by several continuous up-and-down hand-waves, we can adjust the Air Conditioner into a comfortable temperature. To achieve this functionality, in Chapter 8, we present *AudioGest*, a device-free gesture recognition system that can accurately sense the hand in-air movement around user's mobile devices. Compared to the state-of-the-art, AudioGest is superior in using only one pair of built-in speaker and microphone, without model-training or any extra hardware or infrastructure support, to achieve a multi-modal hand detection. Our HRG system is not only able to accurately recognize various hand gestures, but also reliably estimate the hand in-air duration, average moving speed and waving range. We achieve this by transforming the device into an active sonar system that transmits inaudible audio signal and decodes the echoes of hand at its microphone. Our experimental results on four real-world scenarios show that AudioGest

detects six hand gestures with an accuracy up to 96%, and by distinguishing the gesture attributions, it can provide up to 162 control commands for the smart environment. The key research papers related with this part are listed as follows:

[C13] **W. Ruan**, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, AudioGest: Enabling Fine-Grained Hand Gesture Detection by Decoding Echo Signals, *The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*, Heidelberg, Germany, Sept 12-16, 2016. [*ERA A, CORE A*, Full Research Paper, Acceptance Rate = 23.7%, Oral Presentation*]

[J2] **W. Ruan**, Q. Z. Sheng, P. Xu, L. Yang, *etc.*, Making Sense of Doppler Effect for Multi-Modal Hand Motion Detection, *IEEE Transaction on Mobile Computing (TMC)*, To appear
[*ERA A*, Impact Factor = 3.822*]

1.4 Summary

In conclusion, this Ph.D. thesis attempts to develop a device-free, intelligent and maintenance-free supporting system that can enable a healthy, safe, cost-effective independent living for the elderly in a residential home. Recent advancement in low-cost passive Radio-Frequency Identification technology enables our envisioned system possible. We have systematically explored how to utilize low-cost, unobtrusive and battery-free passive RFID tags to realize this living-supporting system. In particular, we tackle this challenge from six research perspectives. For each part, we provide a novel, device-free and cost-effective solution by taking recent advances of sensor technologies and state-of-the-art machine learning techniques. Given the aging of the population, the cost of health care, and the importance that people want to remain independent and safe at their own homes, the demand on developing novel technologies such as the one in this thesis is becoming increasingly urgent. Our proposed innovative technologies can help the elderly live longer independently and safely in

their own homes, with minimal support from the decreasing number of individuals in the working-age population.

This thesis has been funded by Prof. Michael Sheng's Australian Research Council Discovery Project (ARC DP130104614).

Chapter 2

Literature Review

This chapter focuses on discussing and reviewing the state-of-the-art research works from five different aspects including missing sensor reading recovery, indoor localization and tracking, human activity recognition, fall detection and hand gesture recognition. It is specifically organized as follows, Section 2.1 discusses the latest sensor reading recovery techniques, especially compares the latest matrix completion and tensor completion methods. Then Section 2.2 intensively reviews the recent indoor localization and tracking systems from both wearable and device-free perspectives and further identifies the main pros and cons of existing RFID-based systems, as well as highlights the advantages of our system. Furthermore, Section 2.3 concentrates on discussing the state-of-the-art research efforts on human activity recognition and Section 2.4 reviews the latest fall detection systems, especially those device-free techniques. Finally, in Section 2.5, we extensively discuss the hand gesture recognition systems in terms of wearable device and device-free based technologies.

2.1 Missing Sensor Reading Recovery

Imputing/estimating the missing values from a partially observed data have attracted much interest in the past decades such as signal processing, data mining, computer vision [1, 2].

Generally, we categorize the techniques of recovering missing values into three types - *regression* based methods, *matrix completion*, and *tensor completion* based methods. In this section, we will concentrate on discussing the latter two categories that are more related to our work.

2.1.1 Matrix Completion Techniques

To capture the global information of a targeted dataset, the “rank" of the matrix is a powerful tool and many matrix completion/recovery based on the inherent low-rank structure assumption have drawn significant interest. Massive optimization models and efficient algorithms are proposed [3]. Some researchers [4] have shown that under some mild conditions, most low-rank matrices can be perfectly recovered from an incomplete set of entries by solving a simple convex optimization program, namely, solving $\min_M\{\text{rank}(M)|P_\Omega(X) = P_\Omega(T)\}$, where M indicates recovered data matrix and P_Ω means only entries in Ω are observed. Although *low-rank matrix completion* has drawn significant interest and has played an important role in missing data recovery, such methods cannot work or fail to recover the data matrix under some circumstances that a subset of its entries may be corrupted or polluted by various sparse noises [5].

As a result, many robust versions of matrix completion that can recover the low-rank matrix from both noisy and partial observations of data are proposed lately [6, 7]. For example, Chen *et al.* [8] investigate the problem of low-rank matrix completion where a large number of columns are arbitrarily corrupted. They show that only a small fraction of the entries are needed in order to recover the low-rank matrix with high probability, without any assumptions on the location nor the amplitude of the corrupted entries. Chen *et al.* [5] also deal with a harder problem that a constant fraction of the entries of the matrix are outliers. They exploit what conditions need to be imposed in order to exactly recover the such underlying low-rank matrix. Finally, Klopp *et al.* [9] study the optimal reconstruction

error in the case of matrix completion, where the observations are noisy and column-wise or element-wise corrupted and where the only piece of information needed is a bound on the matrix entries. Recently, a multi-view learning based method is proposed to capture both local and global information in terms of spatial and temporal perspective, achieving state-of-the-art performance [10]. It also demonstrates that both local and global spatial/temporal correlations play an important role in sensor data reconstruction.

2.1.2 Tensor Completion Techniques

Though promising of matrix-based models, the recovered dataset, in many practical applications, has complex multi-dimensional spatio-temporal correlations, which can be naturally treated as a tensor instead of a matrix [11, 12]. Therefore data recovery based on high-dimensional tensor or multi-way data analysis is becoming prevalent in recent several years.

Generally, there are two state-of-the-art techniques used for tensor completion. One is the nuclear norm minimization, many pioneering similar works are emerged [13, 14] since Liu *et al.* [11] first extend the nuclear norm of matrix (*i.e.,* the sum of all the singular values) to tensor. Later on, Gandy *et al.* [13] and Signoretto *et al.* [15] consider a tractable and unconstrained optimization problem of low-n-rank tensor recovery and adopt the Douglas-Rachford splitting method and Alternating Direction Method of Multipliers (ADMM) method. Another popular technique is to utilize the tensor decomposition [16], *i.e.,* decomposing the Nth-order tensor into another smaller Nth-order tensor (*i.e.,* core tensor) and N factor matrices. Generally, Tucker and CANDECOMP/PARAFAC are the two most popular tensor decomposition frameworks [17]. For example, Acar *et al.* [18] develop an algorithm called CP-WOPT (CP Weighted OPTimization), which introduces a first-order optimization approach for dealing with missing values and has been testified to provide a good imputation performance. Alexeev *et al.* [19] however focus on exploring tensor rank lower and upper bounds, especially for the explicit tensors. More recently, Da Silva *et al.* [20] and Kressner *et*

al. [16] propose a nonlinear conjugate gradient method for Riemannian optimization based on the hierarchical Tucker decomposition and Tucker decomposition separately. However, those tensor completion methods are neither applied into recovering spatio-temporal sensory data, nor can deal with a circumstance that the known sensor readings are corrupted by noise. Our ADMM based robust tensor completion method, on the contrary, can fill both two gaps and recover the missing sensor values with a high accuracy and robustness.

2.2 Device-free Human Localization and Tracking

This section will review the related works regarding indoor localization and tracking. Generally, they can be categorized as *wearable-device based localization* and *device-free localization*. We will focus more on the device-free techniques that is more related to our system.

2.2.1 Wearable Devices based Techniques

Wearable device based systems normally require the user to carry or wear a device such as RF transceivers, smart-phones, RFID reader or tags. The very first indoor localization work is Criket [21] which is able to track a subject wearing an ultrasonic transmitter by measuring the ToA (time-of-arrival) of a short ultrasound pulse. Another very famous pioneering work, LANDMARC [22], first deploys dozens of active RFID tags in the indoor environment, and then match the RSSI from a tag carried by a subject with the profiled RSSI fingerprints to localize a target. Lately, Yang *et al.* [23] design a high-performance tracking system based on passive RFID hardware, which can real-time track a tagged object with a centimeter-level error. With the popularity of smart phones, Zhou *et al.* [24] present an activity sequence-based pedestrian indoor localization approach using smartphones. They first detect the activity sequence using activity detection algorithms and use HMM to match the activities in the

activity sequence to the corresponding nodes of the indoor road network. MaLoc [25] utilizes magnetic sensor and inertial sensor of smart-phones by a reliability-augmented particle filter to localize a subject, which does not impose any restriction on smart-phone's orientation. Currently, wearable device based localization is still a very active research area due to its high accuracy and robustness. However, the requirement of wearing a sensor or device may not be practical for some circumstances.

2.2.2 Device-free Techniques

Device-free techniques can relax wearing requirements for users. In 2007, the device-free localization challenge was first identified by Youssef *et al.* [26] who designed a preliminary WIFI-based Device-free Passive (DfP) localization system. Since then enormous DfP localization schemes have emerged. Basically, according to the type of hardware installed, device-free localization schemes can be generally classified into three categories: WIFI, RFID, and environmental sensors[1] based techniques. Environmental-sensor based category includes many types of sensors, which either cost too much or need some special deployment for facilities, or may be influenced by natural light or thermal source. Next, we will intensively review the device-free localization systems based on WIFI and RFID, which is more related to our system.

WIFI-based Device-free Localization

With the pervasiveness of WIFI, enormous device-free localization systems built upon wireless signals have emerged during the last decade [43]. The general intuition behind this technique is that, when a user moves in a monitored area, RSS and CSI abstracted from WIFI signals will embody different attenuation levels. WIFI-based schemes exploit various models to decode the signal variations in either Radio Signal Strength (RSS) or

[1]For simplicity, in this thesis, we generally treat camera-based techniques as one type of environmental sensors, including infrared sensors [27], light sensors [28], and varies kinds of cameras [29–31]

Table 2.1 Comparison of typical device-free localization systems

Comparison Systems	Measured Physical Quantity	Non-LoS Localization?	Hardware	Cost of Single Node/Device	Localization Accuracy	Maintenance (*e.g.,* replace battery *etc.*)	Training Overhead	Tested in a Residential Home?	Device -free?
TagArray[32]	RSS Threshold	NO	Active Tags	Medium	Medium	Medium	Low	NO	YES
TASA[33]	RSS Threshold	NO	Passive and Active Tags	Medium	Medium	Medium	Low	NO	YES
RTI[34]	RSS Attenuation	NO	Wireless Nodes	Medium	High	High	Low	NO	YES
CareLoc[35]	Swipe Event	NO	Passive RFID Tags	Low	High/Detecting Swipe Event	Low	Low	NO/Test in Hospital	NO
NUZZER[36]	RSS Changes	YES	Wireless Nodes	Medium	High	Medium	High	NO	YES
SCPL[37]	RSS Changes	YES	Wireless Nodes	Medium	Medium	Medium	Medium	NO	YES
ilight[28]	Light Strength	No	Light Sensors	High	Medium	Medium	Low	NO	YES
Ichnaea[38]	RSS changes	YES	Wireless Nodes	Medium	Medium	Medium	Low	NO	YES
Twins[39]	Critical State Jump	YES	Passive Tags	Low	High	Low	Low	NO	YES
VisualLoc[40]	Video Frame	NO	Wireless Visual Sensors	High	Very High	Medium	Medium	No	YES
WiTrack[41]	FMCW signal	Yes	USRP	Very High	Very High	Medium	Low	NO	YES
FlexibleTrack[42]	RSSI	YES	Smartphone and Wireless Nodes	Medium	Medium	Medium	Low	NO	NO
Ours	RSS Variance	YES	Passive Tags	Low	High	Low	Low	YES	YES

Channel State Information (CSI) for localization or tracking [44]. For example, RTI [34] proposes a radio tomographic imaging model to resolve the RSS attenuation caused by human motion within an area with dense-deployed wireless notes. By extending the fingerprint-based technique, Xu *et al.* [45] adopt various several discriminant analysis approaches to

classify a user's location. Furthermore, they design another localization system, SCPL [37], which is able to count and localize multiple residents. NUZZER, a large-scale indoor DfP tracking system, was developed by Seifeldin *et al.* [36]. This work first builds a passive RF map in an off-line manner and then utilizes a Bayesian model to find a location with maximum likelihood. Ichnaea [38] is another advanced WIFI-based device-free system in terms of training overhead and robustness. It combines anomaly detection method and particle filtering to robustly track a single subject in an area with wireless infrastructure. More recently, WiTrack, designed by Adib *et al.* [41], is able to track a human body even the subject is behind a wall or occluded by furniture. It requires the support of USRP and decodes the locations by analyzing the reflected specialized Frequency Modulated Continuous Wave from the human body.

RFID-based Device-free Localization

Undoubtedly, WIFI-based systems bear some promising characters such as moderate cost, tiny node size and elegant signal propagation models. However, they still require to be powered in a wire or battery style, which inevitably needs regular maintenance, *e.g.,* periodical replacement of batteries. On the contrary, RFID-based DfP localization systems have shown more attractive features such as significant cost-efficiency, zero maintenance (cheap passive tags) and good hardware scalability. Thus several pioneering device-free systems have been developed recently based upon either active or passive RFID hardwares. The very first RFID-based device-free localization system, TagArray, is proposed by Liu *et al.* [32] who placed active RFID tags as arrays on the ground localizing a subject by measuring if RSSI readings are higher than a threshold. TASA [33] is another similar device-free localization system but is more cost-efficient due to it utilizes both passive and active RFID tags. Both TagArray and TASA systems focus more on mining frequent trajectory patterns instead of tracking accuracy, and they only quantify the binary relation of RSSI readings

with human locations (*i.e.,* comparing RSSIs with thresholds). Later on, Wagner *et al.* [46] extend the RTI model from WIFI-based localization to RFID hardware platform that can track a single user in a small obstacle-free zone with dense passive tags deployed. Very recently, a new localization system built upon passive tags, Twins [39], was also proposed, which leverages an interference observation of two very-near tags to detect an intruder in a warehouse reaching $0.75m$ mean tracking error. More lately, Yang *et al.* [47] design a device-free, see-through-wall tracking system with high accuracy, in which they attached a group of passive RFID tags on the outer wall to track a moving subject by analyzing the reflected signals from human body.

Table 2.1 compares our system with other typical localization systems in a high-level view. Our work thoroughly mines the relations between the RSSI of tags and the impact brought by human motion to achieve high accuracy localization and tracking. Moreover, our RFID-based system is built *solely* upon passive tags, which is less costly and more convenient for a practical deployment (*e.g.,* tiny size and weight, battery-free feature). At the same time, our system does not contain any privacy information since it merely exploits RSSI signals from passive tags. More importantly, most existing localization systems based passive RFID tags are deployed and tested in an controlled/semi-controlled or cleared space (*i.e.,* a room or office equipped only with a few objects, lack of metal electronic appliances). However, by further leveraging the human-object interaction events in a residential home, our passive RFID based, device-free system can beyond the limits of current similar systems and achieve high-accuracy localization and tracking accuracy even in a clustered full-furniture house.

2.3 Human Activity Recognition

The goal of activity recognition is to detect human physical activities from the data collected through various sensors. There are generally two main research directions: *i)* to instrument

people, on whom sensors and RFID tags are attached, and *ii)* to instrument the environment, where sensors are deployed inside the environment and people do not have to carry them.

Wearable sensors such as accelerometers and gyroscopes are commonly used for recognizing activities [48–51]. For example, the authors in [52] design a network of three-axis accelerometers distributed over a user's body. Activities can then be inferred by learning information provided by accelerometers about the orientation and movement of the corresponding body parts. However, such approaches have obvious disadvantages including discomfort of wires attached to the body as well as the irritability that comes from wearing sensors for a long duration. More recently, researchers are exploring smart phones equipped with accelerometers and gyroscopes to recognize activities and gesture patterns [53, 54]. Krishnan *et al.* propose an activity inference approach based on motion sensors installed in a home environment [55].

Apart from sensors, RFID has been increasingly explored in the area of human activity recognition. Some research efforts propose to realize human activity recognition by combining RFID passive tags with traditional sensors (e.g., accelerometers). In this way, daily activities are inferred from the traces of object usage via various classification algorithms such as Hidden Markov Model, boosting and Bayesian networks [56, 57]. Recently, passive RFID techniques have been widely used in pervasive computing community. Thus some pioneering efforts are emerged to exploit the potential of using "pure" RFID techniques for activity recognition. For instance, Wang *et al.* [58] present a prototype RFID-based system to characterize human activity by extracting temporal and spatial features from radio frequency patterns. Asadzadeh *et al.* [59] propose to recognize gesture with passive tags by combining with multiple subtags to tackle uncertainty of the RFID readings. However, these research efforts require people to carry RFID tags or even readers (*e.g.,* wearing a bracelet).

More recently, similar to device-free localization and tracking, many research efforts concentrate on exploring device-free activity recognition. Such approaches generally exploit

radio transmitters installed in environment, and people are free from carrying any receiver or transmitter. Most device-free approaches focus on analyzing and learning distribution of received signal strength or radio links. Youssef *et al.* [60] propose to localize people by analyzing wireless signal strength moving average and variance. Zhang *et al.* [33] develop a tag-free sensing approach using RFID tag array. However, most of these efforts have been done on localization and tracking, not on activity recognition. Only very recently, the authors of [61] and [62] propose device-free activity recognition approaches using sensor arrays. Arising from this idea, we deploy passive RFID tags as an array attached on the wall of a residential home to achieve the device-free activity recognition. Compared to current device-free HAR works, our passive RFID-based approach has many advantages including in low cost and maintenance free, as well as light size and weight.

2.4 Fall Detection

Timely detection of a fall event can abbreviate the damage degree and reduce the mortality for the elderly. Fall detection for the elderly has been a hot topic in health-care industry and has attracted a lot of attention from academia in the past two decades. Since early 1990s, many fall detection systems have been proposed by researchers from different communities. In [63, 64], the hardware and methods used in existing fall detection systems have been thoroughly discussed and reviewed. Based on the hardware used by fall detection, current systems can be classified into four groups: wearable sensor based, smart-phone based, vision-based, and environmental sensor based techniques. From the point of obstructiveness, the former two categories can be regarded as device-free, the latter two are of intrusive in general.

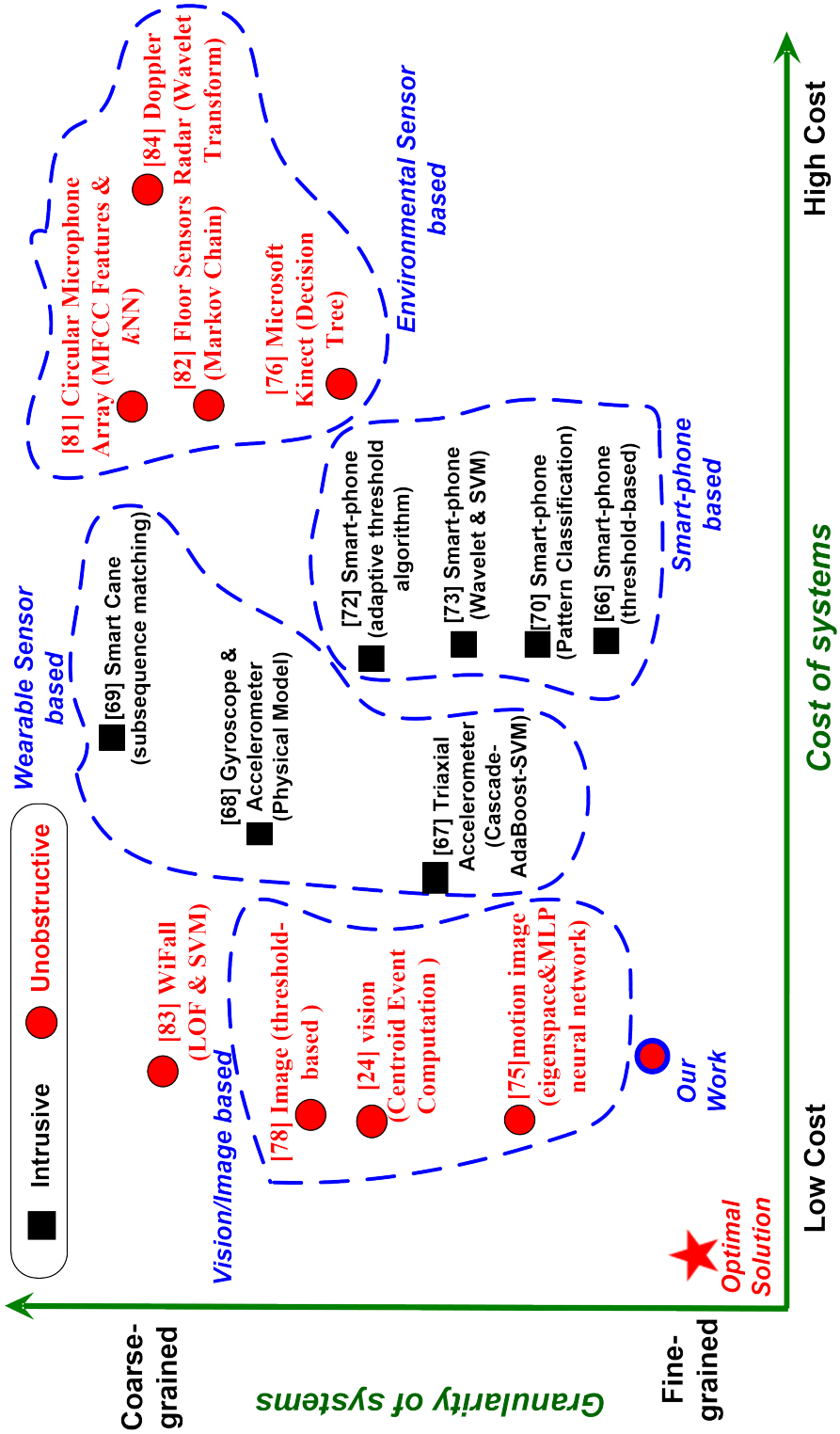


Fig. 2.1 Design Space: comparing to related fall detection systems

Wearable sensor based fall detection systems rely on sensors that are embedded in wearable items such as coat, belt and watch or be taken by hand, such as smart cane. The widely used sensors include inertial sensors [65], tri-axial accelerometers [66], gyroscopes [67] and smart cane [68]. Lee *et al.* [65] proposed a novel vertical velocity-based fall detection method to detect a fall event using a wearable inertial sensor. Cheng *et al.* [66] designed a cascade-AdaBoost-SVM classifier to realize a real-time fall detection method based on tri-axial accelerometers worn on the body. Li *et al.* [67] presented a fall detection system using both accelerometers and gyroscopes, in which linear acceleration and angular velocity are measured to determine whether motion transitions are intentional. In [68], Lan *et al.* present and design an automatic fall detection system by using a smart cane. These detection systems can only work on the premise that all the devices are worn by the subject and connected correctly to the human body. Such requirements give additional burden and interfere subjects' daily life, which are impractical for some applications.

Most modern smart-phones have built-in sensors that can measure motion, orientation, and various environmental conditions. These sensors are capable of providing raw data with high precision and accuracy. Thus, smart phone based fall detection is promising and with good potential [69], which can integrate all sensors into one single mobile device (e.g., inertial sensors [70], tri-axial accelerometers [71] and gyroscopes [72]). However, smart-phone based fall detection systems share the same mechanism as wearable sensors based techniques. They also have the same problem with wearable based methods. Most users may not take with their phones all the time, especially at home.

Much work has also been done in investigating the use of standard imaging sensors for fall detection. Approaches have ranged from single cameras mounted on the wall to multiple cameras placed around a room [73, 74], or to using a depth-camera Kinect [75, 76]. Lee [77] detected a fall by analyzing the shape and 2D velocity of the person. Rougier [78] used wall-mounted cameras to cover large areas and falls were detected using human shape

variation. Despite the considerable achievements that have accomplished in this field over the recent years, traditional camera-based systems still suffer from a number of limitations. The problem this method brings is that people may feel uncomfortable with a camera overhead, especially in bathroom. Besides the privacy intrusion, this method is also limited by line of sight problem and fails in darkness, where falls usually happen.

Device-free fall detection that use environmental sensors attempts to fuse ambient noise information including thermal distribution [79], audio [80], floor vibrational [81], Channel State Information (CSI) [82] data and microwave signal [83] produced by a fall for the detection purpose. The principle is based on the fact that human movements in a living setting will cause the signal variations of environmental sensors (e.g., pressure senors [81], acoustic sensors [80], thermal sensors [79] and wireless transceivers [82], radars [83]), which can be regarded as being less intrusive. For example, WiFall [82] employs the time variability and special diversity of Channel State Information (CSI) as the indicator of human activities to infer a fall event. However, current device-free fall detection systems focus more on detecting a fall event in some predefined areas and fail to provide fine-grained information such as status before falling and fall orientations, which may be valuable for rescuers. Figure 2.1 illustrates our device-free, fine-grained fall detection system based on pure UHF passive RFID tags in the design space of current FD systems. Compared to other hardware platforms, RFID is cost-effective (passive tags cost several cents each) and practical (e.g., no maintenance needs, no battery) and promising in identifying environmental changes [84, 85]. In the meantime, our FD system can provide fine-grained contextual information of a fall event, including what is people doing before falls and the falling orientation.

2.5 Hand Gesture Recognition

Hand gesture recognition is an active research area over last decades and has been widely used in many areas such as medical systems, human-machine interactions, and automotive

assistant systems. Existing HGR systems can be categorized into two general types: *wearable sensor/device based* gesture recognition and *device-free* gesture recognition.

2.5.1 Wearable Devices based Gesture Recognition

Wearable sensor/device based systems utilize various sensors (*i.e.,* 3-axis accelerometer [86], inertial sensor [87], gyroscope [88] or other smart devices [89] etc.) to sense the movement of hand or arm. For example, some researchers infer the hand movement by wearing a shaped magnet [90]. Humantenna [87] requires the user to wear a small Wireless Data Acquisition Unit enabling the human body as an antenna for sensing whole-body gestures. With the advanced built-in sensors in mobile device, the system in [88] transfers the acceleration recorded by a smartphone into a real-time hand moving trajectory.

Recently, Lu *et al.* [91] designed a wearable device to acquire acceleration and SEMG (Surface ElectroMyoGraphic) signals and adopted a DTW-based Bayesian classifier to recognize 19 predefined gestures. Singh *et al.* developed Inviz [92], a low-cost gesture recognition system using textile-based capacitive sensor arrays. It decodes hand gestures through a calculation-efficient hierarchical algorithm. More lately, some researchers adopt micro-radars to realize a series of gesture recognition applications. For instance, Li *et al.* proposed Tongue-n-Cheek [93], a contact-less tongue gestures recognition system by designing a head-wearable device containing three 24GHz micro-radars. By adopting a similar micro-radar array, Goel *et al.* designed a facial gesture recognition system, called Tongue-in-Cheek [94], which can differentiate 8 facial expressions. All these gesture recognition systems either require users to wear a device/sensor (e.g., magnet ring, smart bracket and SEMG sensors) or need to install extra hardware such as WDAU, micro-radar or capacitive plates, which might be impractical for some applications (*e.g.,* elderly people with dementia may forget to wear those devices or sensors) and add extra cost.

Beside those conventional gesture systems, some other research efforts focus on stroke-gesture recognition which enables smart-phones to accurately recognize the hand strokes on the screen. For example, Wobbrock *et al.* [95] develop a uni-stroke gestures recognition system, called $1 Recognizer, which can recognize 16 pen-gestures on the screen of a smartphone. Li *et al.* design Protractor [96], a fast and lightweight single-stroke gesture recognition system, which introduces a novel closed-form solution for calculating the similarity of hand strokes. However, these recognition systems are mainly for recognizing stroke-based gestures by touching the screen, which is different from our HGR system that focuses on in-air multi-modal hand gesture recognition without screen-touching.

2.5.2 Device-free Gesture Recognition

This category can be further classified into vision-based, environmental sensor based, RF-based, and sonar-based approaches.

Video-based hand-gesture recognition systems often do the hand-region segmentation using color and/or depth information, and sequences of features for dynamic gestures are used to train classifiers, such as Hidden Markov Models (HMM) [97], conditional random fields [98], SVM [99], DNN [100]. However, vision-based techniques are usually regarded as being privacy-invasive. They also require users within the LOS (line of sight) of cameras, fail to work in dimmed environments, and incur high computational cost. Some environmental sensor-based hand recognition systems have been emerged, such as Leap Motion that explores multiple channels of reflected infrared signals to identify hand gestures, Kinect [101] that uses depth sensor to enable in-air 3D skeleton tracking.

Recently, RF-based gesture recognition systems are also very popular due to its low-cost and being less intrusive [102, 103]. For example, WiVi [104, 105] uses ISAR technique to track the RF beam, enabling a through-wall gesture recognition. RF-Care [106] proposes to recognize human gestures and activities in a device-free manner based on a passive RFID

array. WiSee [107] can exploit the doppler shift in narrow bands in wide-band OFDM (Orthogonal Frequency Division Multiplexing) transmissions to recognize 9 different human gestures. WiGest [103] explores the effect of the in-air hand motion on the RSSI in WiFi to infer the hand moving directions as well as speeds. Melgarejo *et al.* [108] leverage the directional antenna and short-range wireless propagation properties to recognize 25 standard American Sign Language gestures. AllSee [109] designs a very power-efficient hardware that extracts gesture information from existing wireless signals.

SonarGest [110] is one of the pioneering audio-based hand recognition systems, which uses three ultrasonic receivers and one transmitter to recognize 8 hand gestures. The technique utilized is a supervised Gaussian Mixture Model that can capture the distribution of the feature vectors obtained from the Doppler signal of gestures. However, it needs to collect training data (potentially labour-intensive and time-consuming) and requires extra sonic hardware. SoundWave [111] is another pioneering HGR system by exploiting audio Doppler effect as well. It only utilizes the built-in speakers and microphones in computers and require no training. SoundWave designs a threshold-based dynamic peak tracking technique to effectively capture the Doppler shifts, thus can distinguish five different hand gestures.

Most recently, researchers are trying to transform Commercial off-the-shelf (COTS) speakers and microphones into a sonar system to detect human breath [112], track a finger movement [113], and sense user's presence [114]. Most of these systems adopt similar ideas from RF-based approaches, either decoding the echo of Frequency-Modulated Continuous-Wave Radar (FMCW) sound-wave to measure the human body, or utilizing the OFDM to achieve real-time finger tracking, or exploring the Doppler effect when human approaching or away from the microphone. However, such systems need two microphones or require specialized design of soundwave that is power-intensive. Motivated by, but different to, the previous works, our system only utilizes one speaker and one microphone by emitting single-tone audio to achieve a multi-modal gesture recognition. It can also decode the echo's

spectrogram into real-time hand waving velocity by thoroughly exploring the relations of hand motion and echo's frequency shifts.

2.6 Summary

In conclusion, this chapter intensively reviews state-of-the-art related works from five research facets, which substantially covers six research issues we intend to solve in this thesis. Concretely, in the hardware layer, we discuss the recent research efforts on missing data recovery, especially thoroughly compare the pros and cons between matrix completion and tensor completion techniques. In the discovery layer, we extensively review the indoor human localization and activity recognition approaches from wearable device and device-free based views. For the latter category, we further detail the latest advance by classifying it into WIFI-based, RFID-based techniques. In the monitoring layer, we primarily focus on reviewing the fall detection systems. In the application layer, we discuss the recent hand gesture recognition systems.

In this chapter, we provide contexts and literature reviews for the six research issues targeted by this thesis. Also, we illustrate how our ideas naturally arises from and advance those related works. From Chapter 3 to Chapter 8, we will present the technical details of our solutions or systems correspondingly.

Chapter 3

Recovering Missing Sensor Readings via Low-Rank Tensor Completion

With the booming of the Internet of Things, tremendous amount of sensors have been installed in different geographic locations, generating massive sensory data with both time-stamps and geo-tags. Such type of data usually have shown complex spatio-temporal correlation and are easily missing in practice due to communication failure or data corruption. In this chapter, we aim to tackle the challenge – how to accurately and efficiently recover the missing values for corrupted spatio-temporal sensory data. In particular, we first formulate such sensor data as a high-dimensional tensor that can naturally preserve sensors' both geographical and time information, which we call a *spatio-temporal Tensor*. Then we model the sensor data recovery as a low-rank robust tensor completion problem by exploiting its latent low-rank structure and sparse noise property. To solve this optimization problem, we design a highly efficient optimization method that combines the alternating direction method of multipliers and accelerated proximal gradient to minimize the tensor's convex surrogate and noise's ℓ_1-norm. In addition to testing our method by a synthetic dataset, we also use passive RFID (radio-frequency identification) sensors to build a real-world sensor-array testbed, which

generates overall 115,200 sensor readings for model evaluation. The experimental results demonstrate the accuracy and robustness of our approach.

3.1 Introduction

With the rapid development of sensor technology, enormous numbers of smart devices or sensors have been deployed in our planet and thus served as a basic yet essential component of IoT (Internet of Things) [115, 116]. Such tremendous smart devices enable ease of access, retrieval and monitoring of our surrounding environment in a real-time manner. For instance, fine-particles (*e.g.,* PM 2.5) sensors are deployed in different locations within a city to continuously and cooperatively monitor the air quality [117, 118]. Usually, many sensory data in real world share a common character that they are not only related to the time dimension (*e.g.,* time series data) but also have a two-dimensional (*e.g.,* sensors deployed in different latitudes and longitudes) or even three-dimensional spatial attribute (*e.g.,* sensors placed in various latitudes, longitudes and altitudes), which we thus call multi-dimensional spatio-temporal sensory data.

However, in practice, sensors easily experience an issue of missing readings due to unexpected hardware failures (*e.g.,* power outages) or communication interruptions [117, 119]. Those missing values will not only decrease the real-time monitoring performance, but also compromise the accuracy of back-end data analysis such as data predication, inference or visualization. Besides the data loss, the observed sensory data are also easily polluted by the environmental noise, making accurate data recovery even more difficult [117]. Therefore, accurately yet efficiently interpolating the missing sensory data is a non-trivial and challenging task, especially for the multi-dimensional spatio-temporal sensory data with noise pollution.

The key to tackle this challenge lies on how to accurately model the quantitative dependencies of the missing readings with the known ones. The most widely used and straightforward technique is various filtering or regression algorithms that estimate the missing values ac-

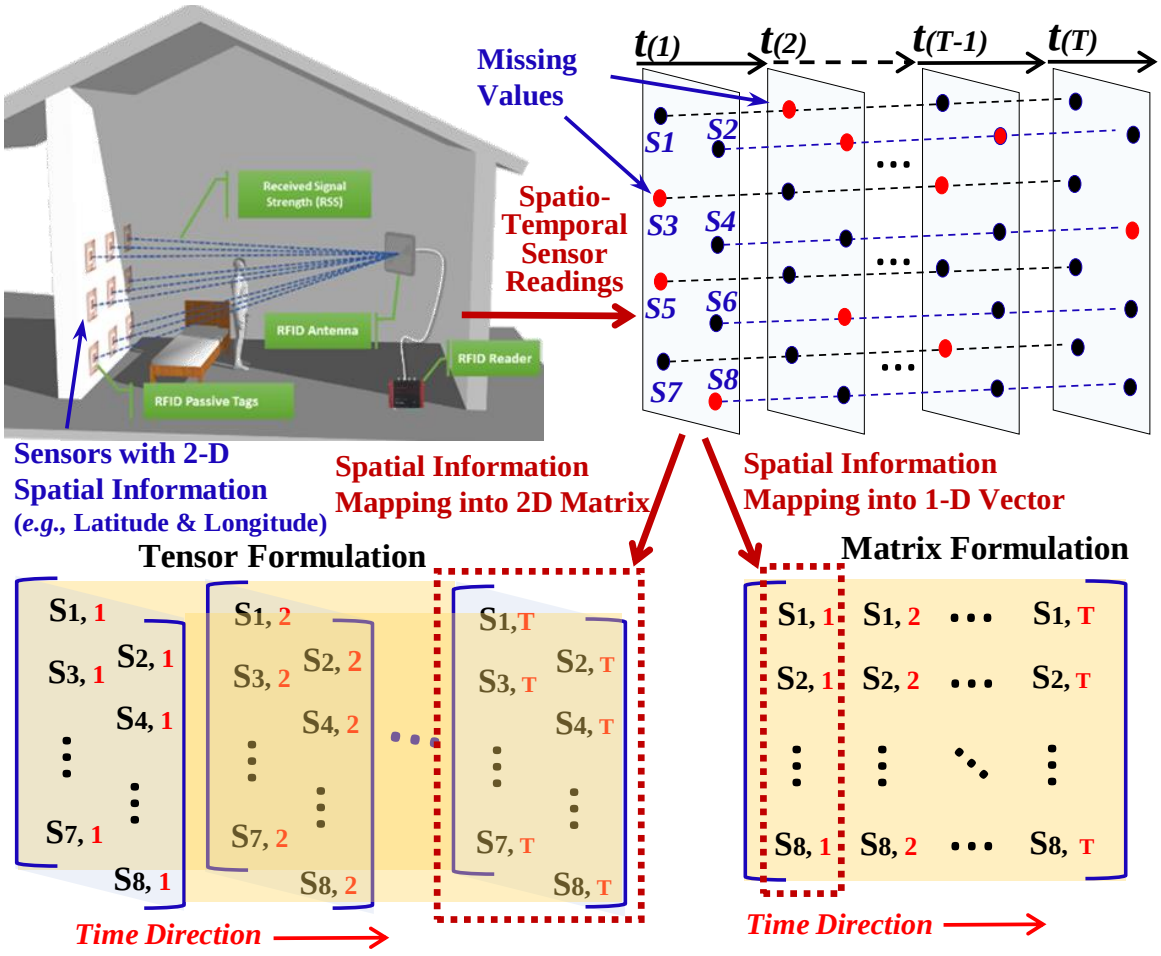


Fig. 3.1 Matrix formulation *vs* Tensor formulation

cording to their local temporal/spatial interdependence, such as median filtering, exponential moving averaging [120], Kriging, Kalman filtering [121], or regression methods with different complexities including ARMA/SARIMA (AutoRegressive Moving Average/Seasonal AutoRegressive Integrated Moving Average), SVR (Support Vector Regression) [122], kNN (k-Nearest Neighbors) *etc.* However, such intuitive approaches suffer from two shortcomings: *i)* it only learns either spatial or temporal dependencies among readings, and is hard to capture both features; *ii)* it unavoidably ignores the global correlations of data (*e.g.,* for some occasions, the missing readings may depend on some far-away entries instead of those nearby values), leading to an inaccurate estimation in some circumstances.

To solve this issue, some researchers treat the sensory data as a matrix and propose various matrix completion/recovery methods to estimate the missing values by capturing their inherent low-rank structure[1] [4, 123]. Those methods usually model time-dependent sensory data as a matrix $S = [s_1, s_2, ..., s_T]$ where vector $s_k \in \mathbb{R}^N$ represents readings of N different sensors at time-stamp k. In this regard, matrix completion based methods are, in principle, to recover the unknown entries by solving an optimization problem: $\min_M \{\text{rank}(M) | P_\Omega(M) = P_\Omega(S)\}$ where $M \in \mathbb{R}^{N \times T}$ indicates recovered data matrix and P_Ω is the project operator that means only entries in Ω are observed [123]. Also, several robust low-rank matrix completion methods are recently proposed to deal with a case that the observed data matrix S are polluted by noise [124]. Although matrix-based methods can well take advantage of the temporal information, still they are limited to capturing *one-dimensional* spatial structure due to a fact that, in matrix formulation, sensors with two-dimensional spatial coordinates are mapped into a one-dimensional vector, unavoidably resulting in the spatial information loss, as illustrates in Figure 3.1.

Recently, a multi-view learning based method has been proposed to capture both local and global information in terms of spatial and temporal perspective, achieving state-of-the-art performance [10]. It also demonstrates that both local and global spatial/temporal correlations play an important role in sensor data reconstruction. However, this method needs to carefully tune five different models, causing two issues: *i)* parameter tuning is not only labor-intensive but also requires some domain knowledge; and *ii)* the interpolation accuracy is sensitive to the parameters since the linear coefficients directly propagate to the model output. In addition, it cannot deal with a case that the known sensor readings are corrupted.

As a result, to resolve those unsatisfied issues, this chapter formates the spatio-temporal sensor data as a *multi-dimensional tensor* - a natural high-order extension of a matrix. Figure 3.1 illustrates our general idea, for spatio-temporal sensory data, compared to the

[1] The rank of a matrix is often linked to the order, complexity, or dimensionality of the underlying system, which tends to be much smaller than the data size.

matrix formulation that only preserves one-dimensional spatial similarity, a tensor-based method naturally captures the two-dimensional geographic dependency among sensors. Nevertheless, applying this high-level idea into the practical still requires to address several challenges. Similar to a matrix-based method [124], low-rank tensor-based data recovery can be formulated as an optimization problem: $\min_{\mathcal{M},\mathcal{N}}\{\text{rank}(\mathcal{M})+\lambda\|\mathcal{N}\|_0 \mid P_\Omega(\mathcal{M}+\mathcal{N}) = P_\Omega(\mathcal{S})\}$ where $\mathcal{M},\mathcal{N},\mathcal{S}$ represent the recovered data tensor, additive noise tensor and observed data tensor respectively, P_Ω means the known entries of tensor. To this end, the first challenge is that, the above optimization is a NP-hard untraceable problem [11]. For the matrix version, we can replace $\text{rank}(\mathcal{M})$ by its tightest convex surrogate (*i.e.,* trace norm), enabling it solvable (*i.e.,* a convex optimization problem). But how to define a convex surrogate for a tensor rank needs some careful design. Secondly, even we can define an effective convex surrogate and make the problem traceable, how to efficiently solve the convex optimization problem with a convergence guarantee also deserves an elaborative consideration.

To address the above challenges, we generalize the idea of trace norm in matrix completion into the tensor, replacing the rank regularization term by the sum of tensor unfoldings' trace norms under all modes (see details in Sec. 3.2). Moreover, to optimize the objective function, we first apply a variable-splitting trick by introducing auxiliary tensor variables to decouple the interdependency of different tensor-modes, then we design an efficient optimization method with a strict convergence guarantee by drawing upon recent advances of Alternating Direction Method of Multipliers (ADMM) [125] and Accelerate Proximal Gradient (APG) [3] (see details in Sec. 3.2 and Algorithm 1). In a nutshell, our main contributions are summarized as follows:

- We propose a robust low-rank tensor completion method to accurately recover the missing sensor readings under a circumstance of noise pollution by exploiting the latent spatio-temporal structures and sparse noise property. We also introduce an efficient

ADMM based optimization scheme to solve the robust tensor completion problem with a theoretical guarantee of convergence to a global optimum.

- We design a real-world sensor-array testbed consisting of 4×4 passive radio-frequency identification sensor-tags, generating overall 115,200 received signal strength indicator (RSSI) readings for the model evaluations. The experiments in both synthetic and real-world datasets demonstrate that our approach outperforms the state-of-the-art approaches in terms of accuracy and robustness.

The rest of the chapter is organized as follows. Sec. 3.2 introduces problem formulation and notations of our solution. We present our model and optimization method in Sec. 3.3 and Sec. 3.4 presents experimental results and analysis. In Sec. 3.5, we offers some concluding remarks.

3.2 Problem Formulation

First, we mathematically define our target problem. Assuming that we have $I_1 \times I_2$ sensors deployed in different spatial areas and collect (noisy) sensor readings[2] for T timestamps (see the example in Figure 3.1), we then can formulate it as a 3-order tensor $\mathscr{O} \in \mathbb{R}^{I_1 \times I_2 \times T}$ and $\mathscr{O} = \mathscr{M} + \mathscr{N}$ where $\mathscr{M}$ represents the true sensor readings (without noise) and $\mathscr{N}$ means the added noise. We use the projection operator $P_\Omega(\mathscr{O}) : \mathbb{R}^{I_1 \times I_2 \times T} \rightarrow \mathbb{R}^K$ that indicates the K observed sensor readings $o_{i,j,t}$ where the index $(i, j, t) \in \Omega$, mapping a tensor to a vector. Formally, this chapter therefore aims to solve the following *Corrupted Sensor Value Recovery* problem:

Problem 1 (Corrupted Sensor Value Recovery) *Given a partially observed data tensor $\mathscr{O}_\Omega$, our task is to accurately recover the true sensor readings $\mathscr{M}$ and additive noise $\mathscr{N}$, where $\mathscr{M}, \mathscr{N}, \mathscr{O} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$.*

[2]We assume the additive noises are sufficiently sparse relative to the data tensor $\mathscr{O}$.

Throughout this chapter, we represent scalars, vectors and matrices by lowercase letters *e.g., x*, bold lowercase letters such as **x**, and upper letters X. Tensors of d-order/dimension are written by calligraphic letters like $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$, whose elements are represented by $x_{i_1 \cdots i_k \cdots i_d} \in \mathbb{R}$ and $1 \leq i_k \leq I_k, 1 \leq k \leq d$. Thus a vector can be seen as a 1-order tensor and a matrix can be seen as a 2-order tensor.

Definition 1 *Unfolding Operator: the mode-k unfolding of* $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ *is denoted by* $unfold(\mathscr{X}, k) = X_{(k)} \in \mathbb{R}^{I_k \times \prod_{i \neq k} I_i}$, i.e., *the row of the matrix* $X_{(k)}$ *are determined by the k-th component of the tensor* $\mathscr{X}$*, whereas all the remaining components form its columns.*

This operation transforms a tensor into a matrix, i.e. matricization or flattening.

Definition 2 *Folding Operator: the mode-k folding of a matrix* $X_{(k)}$ *is defined as* $fold(X_{(k)}, k) = \mathscr{X}$.

Definition 3 *Inner Product of Tensor: the inner product of two tensors with identical size* $\mathscr{X}, \mathscr{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ *is computed by* $\langle \mathscr{X}, \mathscr{Y} \rangle := \sum_{i_1, i_2, \cdots, i_d} x_{i_1 i_2 \cdots i_d} y_{i_1 i_2 \cdots i_d}$.

Definition 4 *Frobenius Norm: Frobenius norm of* $\mathscr{X}$ *is defined as* $\|\mathscr{X}\|_F := (\sum_{i_1, i_2, \cdots, i_d} |x_{i_1 i_2 \cdots i_d}|^2)^{\frac{1}{2}}$.

Thus, for any $k \in \{1, ..., d\}$, we have $\|\mathscr{X}\|_F = \|X_{(k)}\|_F$, and $\langle \mathscr{X}, \mathscr{Y} \rangle = \langle X_{(k)}, Y_{(k)} \rangle$.

Definition 5 *Tensor-matrix multiplication: the multiplication of a d-order tensor* $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ *with a matrix* $A \in \mathbb{R}^{J \times I_k}$ *in mode-k is mathematically defined as* $\mathscr{X} \times_k A \in \mathbb{R}^{I_1 \times \cdots \times I_{k-1} \times J \times I_{k+1} \times \cdots \times I_d}$.

Definition 6 *Tucker decomposition: Given an input tensor, Tucker decomposition uses a smaller/core tensor multiplied by a matrix along each mode to describe the original tensor.*

A tensor $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ *decomposes as*

$$\begin{aligned} \mathscr{X} &= \mathscr{G} \times_1 A^{(1)} \times_2 A^{(2)} \cdots \times_d A^{(d)} = [\![\mathscr{G}, A^{(1)}, ..., A^{(d)}]\!] \\ &= \sum_{r_1=1}^{R_1} \sum_{R_2}^{r_2=1} \cdots \sum_{R_d}^{r_d=1} g_{r_1 r_2 \cdots r_d} a_{r_1}^{(1)} \circ \cdots \circ a_{r_d}^{(d)} \end{aligned} \tag{3.1}$$

where $\{A^{(k)}\}_{k=1}^{d} \in \mathbb{R}^{I_k \times R_k}$ are a set of factor matrices and $R_1, R_2, ..., R_d$ is defined as the Tucker rank. '$\circ$' denotes the outer product. When $d = 3$, Tucker decomposition of tensor is $\mathscr{X} = \mathscr{G} \times_1 A \times_2 B \times_3 C = [\![\mathscr{G}, A, B, C]\!]$.

3.3 Robust Low-Rank Spatio-Temporal Tensor Recovery

Being similar to matrix completion, Problem (1) can be formulated as solving a low-rank minimization problem.

$$\begin{aligned} \min_{\mathscr{M},\mathscr{N}} \quad & \text{rank}_{Tucker}(\mathscr{M}) + \lambda ||\mathscr{N}||_0 \\ \text{s.t. } & P_\Omega(\mathscr{M} + \mathscr{N}) = P_\Omega(\mathscr{O}) \end{aligned} \tag{3.2}$$

where $\text{rank}_{Tucker}(\mathscr{M})$ is the Tucker-rank of a tensor [126]. Similar to matrix completion, this problem is NP-hard. Thus, to make it tractable, we replace Tucker rank by its *convex surrogate* and use ℓ_1-norm instead of ℓ_0-norm as $\min_{\mathscr{M},\mathscr{N}}\{\text{ConSurro}(\mathscr{M}) + \lambda ||\mathscr{N}||_1 \mid \text{s.t. } P_\Omega(\mathscr{M} + \mathscr{N}) = P_\Omega(\mathscr{O})\}$.

Then the first issue is how to define the convex surrogate of a tensor. For a matrix, the trace norm $\|.\|_*$ is the tightest convex envelop of its rank, used as the convex surrogate. Thus, the idea can be generalized into the high-order tensor, defining its trace norm as the sum of the trace norms [11] of the mode-i unfolding in tensor $\mathscr{M}$, *i.e.,* $\text{ConSurro}(\mathscr{M}) = \sum_i \|M_{(i)}\|_*$. Eqn. (3.2) can be therefore transformed into a convex problem:

$$\begin{aligned} \min_{\mathscr{M},\mathscr{N}} \quad & \sum_{i=1}^{d} \|M_{(i)}\|_* + \lambda ||\mathscr{N}||_1 \\ \text{s.t.} \quad & P_\Omega(\mathscr{M}+\mathscr{N}) = P_\Omega(\mathscr{O}) \end{aligned} \tag{3.3}$$

To solve Eqn. (3.3), we introduce an Alternating Direction Method of Multipliers [125] that is very efficient in dealing with convex optimization problems by breaking them into smaller pieces, each of which is then easier to handle. However, the trace norm of each mode unfolding $\|M_{(i)}\|_*, (i = 1, ..., d)$ shares the same values in data tensor $\mathscr{M}$ and cannot be optimized independently so that existing ADMM cannot directly be applied to our problem. Hence, we split these interdependent terms by introducing auxiliary variables $\mathscr{M}_1, \mathscr{M}_2, ..., \mathscr{M}_d$, so that they can be solved independently. In particular, we reformulate Eqn. (3.3) as

$$\begin{aligned} \min_{\mathscr{M}_1,...,\mathscr{M}_d,\mathscr{N}} \quad & \sum_{i=1}^{d} \|M_{i,(i)}\|_* + \lambda \|\mathscr{N}\|_1 \\ \text{s.t.} \quad & P_\Omega(\mathscr{M}_i+\mathscr{N}) = P_\Omega(\mathscr{O}), \;\; i = 1, ..., d. \end{aligned} \tag{3.4}$$

We hence define its *augmented Lagrangian function* as

$$\begin{aligned} &\mathscr{L}_\mu(\mathscr{M}_1, ..., \mathscr{M}_d, \mathscr{N}, \mathscr{Y}_1, ..., \mathscr{Y}_d) = \sum_{i=1}^{d} (\frac{1}{2\mu}\|P_\Omega(\mathscr{M}_i+\mathscr{N}) - P_\Omega(\mathscr{O})\|^2 - \\ &\langle \mathscr{Y}_i, P_\Omega(\mathscr{M}_i+\mathscr{N}) - P_\Omega(\mathscr{O})\rangle) + \sum_{i=1}^{d} \|M_{i,(i)}\|_* + \lambda \|\mathscr{N}\|_1. \end{aligned} \tag{3.5}$$

According to ADMM, we first fix $\mathscr{N}$ to optimize $\mathscr{M}_i$ $(i = 1, ..., d)$ by solving

$$\begin{aligned} &\min_{\mathscr{M}_1,...,\mathscr{M}_d} \sum_{i=1}^{d} (\frac{1}{2\mu}\|P_\Omega(\mathscr{M}_i+\mathscr{N}) - P_\Omega(\mathscr{O})\|^2 - \langle \mathscr{Y}_i, P_\Omega(\mathscr{M}_i+\mathscr{N}) - \\ &P_\Omega(\mathscr{O})\rangle + \|M_{i,(i)}\|_*) \equiv \sum_{i=1}^{d} (\mu\|M_{i,(i)}\|_* + \frac{1}{2}\|P_\Omega(\mathscr{M}_i) - \mathscr{A}_i\|^2) \end{aligned} \tag{3.6}$$

where $\mathscr{A}_i = P_\Omega(\mathscr{O}) - P_\Omega(\mathscr{N}) + \mu\mathscr{Y}_i$. We define the function $f(\mathscr{M}_i) = \frac{1}{2}\|P_\Omega(\mathscr{M}_i) - \mathscr{A}_i\|^2$ and calculate the gradient $\nabla f(\mathscr{M}_i) = P^*_\Omega(P_\Omega(\mathscr{M}_i) - \mathscr{A}_i)$, where $P^*_\Omega(\cdot)$ means the adjoint operation of $P_\Omega(\cdot)$ such as $P^*_\Omega(\mathscr{O}) : \mathbb{R}^K \to \mathbb{R}^{I_1 \times I_2 \times \cdots \times T}$. According to Accelerated Proximal Gradient (APG) method [3], we can independently minimize $\mathscr{M}_i$ through iterative optimization to make the final sum minimal. In particular, we get optimal $\mathscr{M}_i^{(k+1)}$ given $\mathscr{M}_i^{(k)}$ until it converges by solving

$$\begin{aligned}
&\min_{\mathscr{M}_i^{(k+1)}} f(\mathscr{M}_i^{(k)}) + \nabla f(\mathscr{M}_i^{(k)})(\mathscr{M}_i^{(k+1)} - \mathscr{M}_i^{(k)}) + \frac{1}{2\eta}\|\mathscr{M}_i^{(k+1)} - \mathscr{M}_i^{(k)}\|^2 + \\
&\mu\|M_{i,(i)}^{(k+1)}\|_* = \frac{1}{2\eta}\|\mathscr{M}_i^{(k+1)} - \mathscr{M}_i^{(k)} + \eta\nabla f(\mathscr{M}_i^{(k)})\|^2 + \mu\|M_{i,(i)}^{(k+1)}\|_* \\
&\propto \frac{1}{2}\|\mathscr{M}_i^{(k+1)} - \mathscr{M}_i^{(k)} + \eta\nabla f(\mathscr{M}_i^{(k)})\|^2 + \eta\mu\|M_{i,(i)}^{(k+1)}\|_*
\end{aligned} \tag{3.7}$$

To solve Eqn. (3.7), we first need to define *singular value thresholding operator* for tensor.

Theorem 1 *For matrix, the singular value threshold operator is defined as* $\mathscr{T}_\mu(M) := Udiag(\bar{\sigma})V^\intercal$, *where* $M = Udiag(\sigma)V^T$ *is the singular value decomposition (SVD) and* $\bar{\sigma} := max(\sigma - \mu, 0)$.

Similarly, we define the singular value threshold [127] operator for tensor as $\mathscr{T}_{i,\mu}(\mathscr{M}) := \text{fold}(\mathscr{T}_\mu(M_{(i)}), i)$. We thus can calculate the closed-form solution of Eqn. (3.7) as follows:

$$\mathscr{M}_i^{(k+1)} = \mathscr{T}_{i,\eta\mu}(\mathscr{M}_i^{(k)} - \eta\nabla f(\mathscr{M}_i^{(k)})) \tag{3.8}$$

Next, we will optimize $\mathscr{N}$ when fixed $\mathscr{M}_i$ $(i = 1, ..., d)$ by solving the following problem:

$$\begin{aligned}
&\min_{\mathscr{N}} \ \mathscr{Y}\|\mathscr{N}\|_1 + \sum_{i=1}^{d}(\frac{1}{2\mu}\|P_\Omega(\mathscr{M}_i + \mathscr{N}) - P_\Omega(\mathscr{O}) - \mu\mathscr{Y}_i\|^2) \\
&\quad\propto \ \mu\lambda\|\mathscr{N}\|_1 + \frac{1}{2}\sum_{i=1}^{d}\|P_\Omega(\mathscr{N}) - \mathscr{B}_i\|^2
\end{aligned} \tag{3.9}$$

where $\mathscr{B}_i = P_\Omega(\mathscr{O}) + \mu\mathscr{Y}_i - P_\Omega(\mathscr{M}_i)$. To solve Eqn. (3.9), we define *Homogeneous Tensor Array* [126] by introducing an operator that combines the component tensors with the same size along the tensor mode-1 as:

$$\bar{\mathscr{M}} := \Big(\mathscr{M}_1, ..., \mathscr{M}_d\Big)^\mathsf{T} \in \mathbb{R}^{dI_1 \times I_2 \times \cdots \times I_d}, \tag{3.10}$$

which is written as $\text{TArray}(\mathscr{M}_1, ..., \mathscr{M}_d)$ and its linear operator $\mathscr{C} : \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d} \to \mathbb{R}^{dI_1 \times I_2 \times \cdots \times I_d}$, *i.e.,* $\bar{\mathscr{M}} = \mathscr{C}(\mathscr{M}) \in \mathbb{R}^{dI_1 \times I_2 \times \cdots \times I_d}$.

Then, we can attain its adjoint operator $\mathscr{C}^* : \mathbb{R}^{dI_1 \times I_2 \times \cdots \times I_d} \to \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$, such that $\mathscr{M} = \mathscr{C}^*(\bar{\mathscr{M}}) = \sum_{i=1}^d \mathscr{M}_i$.

According to this definition, we can rewrite Eqn. (3.9) as

$$\begin{aligned} &\min_{\mathscr{N}} \ \mu\lambda\|\mathscr{N}\|_1 + \frac{1}{2}\|\mathscr{C}(P_\Omega(\mathscr{N})) - \bar{\mathscr{B}}\|^2 \\ &\propto \frac{\mu\lambda}{d}\|\mathscr{N}\|_1 + \frac{1}{2}\|P_\Omega(\mathscr{N}) - \frac{\mathscr{C}^*(\bar{\mathscr{B}})}{d}\|^2 \\ &= \frac{\mu\lambda}{d}\|\mathscr{N}\|_1 + \frac{1}{2}\|P_\Omega(\mathscr{N}) - \frac{1}{d}\sum_{i=1}^d (P_\Omega(\mathscr{O}) + \mu\mathscr{Y}_i - P_\Omega(\mathscr{M}_i))\|^2 \end{aligned} \tag{3.11}$$

where $\bar{\mathscr{B}} = \Big(\mathscr{B}_1, ..., \mathscr{B}_d\Big)^\mathsf{T}$ and $\mathscr{C}(P_\Omega(\mathscr{N})) = \Big(P_\Omega(\mathscr{N}), ..., P_\Omega(\mathscr{N})\Big)^\mathsf{T}$.

Before solving the Eqn. (3.11), we need the following Theorem.

Theorem 2 *When* $\min_{\mathbf{y}}\{\frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \mu\|\mathbf{y}\|_1\}$*, it has a closed form solution* $\mathbf{y} = \mathscr{S}_\mu(\mathbf{x}) := sign(\mathbf{x})max(|\mathbf{x}| - \mu, 0)$*, where* $\mathscr{S}_\mu(\mathbf{x})$ *is the shrinkage operator, where all the operations are element-wise.*

According to the shrinkage thresholding operator [128], we can define $\mathscr{S}_\mu(\mathscr{M})$ on the $vec(\mathscr{M}) = \mathbf{m}$. As a result, we can solve Eqn. (3.11) to get $\mathscr{N} = \mathscr{S}_{\frac{\mu\lambda}{d}}(\frac{1}{d}\sum_{i=1}^d (P_\Omega(\mathscr{O}) + \mu\mathscr{Y}_i - P_\Omega(\mathscr{M}_i)))$. We thus have $\mathscr{C}^*\mathscr{C}(\mathscr{N}) = d\,\mathscr{N}$, then the optimal condition of Eqn. (3.11) is $0 \in d\,\mathscr{N} + \mu\lambda_1\partial\|\mathscr{N}\|_1 \Leftrightarrow 0 \in \mathscr{N} + \frac{\mathscr{C}^*(\bar{\mathscr{B}})}{d} + \frac{\mu\lambda_1}{d}\partial\|\mathscr{N}\|_1$.

Algorithm 1: ADMM based Robust Tensor Completion

Input: Observed Sensory Data Tensor: $\mathscr{O}$, Set of Missing Value Indexes: Ω
Model Parameters: $\lambda,\ \eta,\ \mu$
Initialization: $\mathscr{M}_i^{(0)} = \mathscr{N}^{(0)} = \mathscr{Y}_i^{(0)} = 0 \quad (i = 1, ..., d)$
Output: Recovered Sensory Data Tensor: $\mathscr{M}$, Estimated Noise Tensor: $\mathscr{N}$

while $k > 0$ **do**
  for $i = 1 : d$ **do**
    $\mathscr{M}_i^{(k+1)} = \mathscr{T}_{i,\eta\mu}(\mathscr{M}_i^{(k)} - \eta\nabla f(\mathscr{M}_i^{(k)}))$;
    `/* Optimize` $\mathscr{M}_i$ `using singular value threshold */`
  end
  $\mathscr{N}^{(k+1)} = \mathscr{S}_{\frac{\mu\lambda}{d}}(\frac{1}{d}\sum_{i=1}^{d}(P_\Omega(\mathscr{O}) + \mu\mathscr{Y}_i^{(k)} - P_\Omega(\mathscr{M}_i^{(k+1)})))$;
  `/* Optimize` $\mathscr{N}$ `using shrinkage thresholding operator */`
  for $i = 1 : d$ **do**
    $\mathscr{Y}_i^{(k+1)} = \mathscr{Y}_i^{(k)} - \frac{1}{\mu}(P_\Omega(\mathscr{M}_i^{(k+1)} + \mathscr{N}^{(k+1)}) - P_\Omega(\mathscr{O}))$;
    `/* Update Lagrangian multiplier parameters` $\mathscr{Y}_i$ `*/`
  end
  if *StoppingCondition* == *TRUE* **then**
    Break; `/* Ending loop when stop condition satisfied */`
  end
end
return $\mathscr{M} = \frac{1}{d}\sum_{i=1}^{d}\mathscr{M}_i^{(k+1)}$ and $\mathscr{N}^{(k+1)}$; `/* Return results */`

Finally, given $\mathscr{M}_i^{(k+1)}$ and $\mathscr{N}_i^{(k+1)}$, we can update the Lagrangian multiplier parameter by $\mathscr{Y}_i^{(k+1)} = \mathscr{Y}_i^{(k)} - \frac{1}{\mu}(P_\Omega(\mathscr{M}_i^{(k+1)} + \mathscr{N}^{(k+1)}) - P_\Omega(\mathscr{O}))$. Algorithm 1 shows the pseudo-code of our optimization method.

Essentially, when $d = 3$, Algorithm 1 alternatively optimizes two blocks of variables $\{\mathscr{M}_1, \mathscr{M}_2, \mathscr{M}_3\}$ and $\mathscr{N}$. By defining $f(\mathscr{M}_1, \mathscr{M}_2, \mathscr{M}_3) := \sum_{i=1}^{3}\|X_{i,(i)}\|_*$ and $g(\mathscr{N}) := \lambda_1\|\mathscr{N}\|_1$, it is easy to verify that Algorithm 1 meets the convergence condition of ADMM. Briefly, the sequence $\{\mathscr{M}_1^{(k)}, \mathscr{M}_2^{(k)}, \mathscr{M}_3^{(k)}, \mathscr{N}^{(k)}\}$ obtained from Algorithm 1 can converge to optimal tensors as $(\mathscr{M}_1^{(*)}, \mathscr{M}_2^{(*)}, \mathscr{M}_3^{(*)}, \mathscr{N}^{(*)})$ for Eqn. (3.4). Hence, the sequence $\{\frac{1}{3}(\sum_{i=1}^{3}\mathscr{M}_i^{(k)}), \mathscr{N}^{(k)}\}$

can reach optimal values. Due to the page limitation, we make the proof details available online[3].

3.4 Experiments

In this section, we first conduct a simulation experiment using synthetic data to compare with the state-of-the-art data recovery methods under different data loss percentages and additive noise ratios. Then we design a real-world experimental testbed using passive RFID (Radio-frequency identification) sensors to generate geo-tagged time-series RSSI readings, which are used to test the practical performance of our method. We run the experiments on a computer (CORE i7-4710HQ 2.50GHz CPU and 16GB RAM) using MATLAB R2015b. We use the Tensor Toolbox[4] and for tensor operations and decompositions and PROPACK Toolbox for SVD (Singular Value Decomposition) calculation[5].

Similar to other data recovery works [11, 123], we adopt the *relative error* $\frac{\|\mathscr{M}-\mathscr{M}_0\|}{\|\mathscr{M}_0\|}$, where $\mathscr{M}, \mathscr{M}_0$ mean the recovered and original data tensor respectively, to evaluate the recovery performance.

3.4.1 Comparison Methods

We compare our method with the following typical data recovery methods:

- *MAF* means the moving averaging interpolation that is the most widely-used method to fill in missing values in time-series sensory data[6].

[3] www.dropbox.com/s/mcqqpxc6m0b5jyn/Appendix.pdf?dl=0
[4] Available in www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html
[5] Available in http://sun.stanford.edu/~rmunk/PROPACK/
[6] Available in au.mathworks.com/help/curvefit/smoothing-data.html

- *IAL-MC* [123] represents the inexact augmented Lagrangian method that can recover a data matrix of being arbitrarily corrupted, it is a matrix-based robust completion method and greatly motivates our work[7].

- *LR-TC* [11] is the earliest yet very effective tensor completion method using block coordinate descent optimization but it cannot deal with the corrupted data (not robust version)[8].

- *ADMM-TC* [125] also utilizes ADMM for solving the tensor completion problem. It provides valuable intuitions for the optimization part in this chapter[9].

3.4.2 Evaluations on Synthetic Data

Similar to the works in [11, 125], we generate a $50 \times 50 \times 20$ data tensor with Tucker rank-$(5,5,5)$. We randomly choose a fraction ρ_n of the tensor entries that are polluted by an additive *i.i.d.* (*i.e.,* independent and identically distributed) noise following uniform distribution $\mathsf{unif}(-a,a)$. Then a fraction ρ_o of the corrupted tensor elements are randomly picked as observed values in $\mathscr{O}_\Omega$. In the experiments, we set μ and η as constants for simplicity, *i.e.,* $\mu = 5 \times \mathsf{std}(\mathsf{vec}(\mathscr{O}))$ and $\eta = 0.91$. We set parameter as $\lambda = \alpha r \lambda_*$, where $\lambda_* = 1, r = \frac{1}{\sqrt{max(I_1,I_2,T)}}$ and α are tuned in $1 < \alpha < 2$.

Recovery Accuracy

Fig. 3.2 compares the relative errors of different methods under an observation percentage from 5% to 100% and a noise ($\rho_n = 0.1$, $a = 1$). We can see that our method has a better recovery accuracy than the other three algorithms. Especially, during the interval between 30% and 60%, our method reveals significantly higher recovery capability. We also observe

[7] Available in www.cis.pku.edu.cn/faculty/vision/zlin/RPCA+MC_codes.zip
[8] Available in www.cs.rochester.edu/u/jliu/code/TensorCompletion.zip
[9] Available in https://github.com/ryotat/tensor

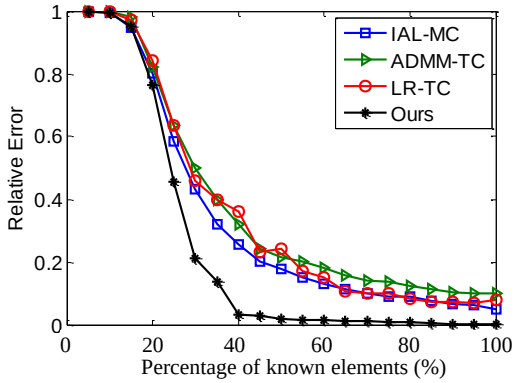


Fig. 3.2 Relative errors for different known elements ($\rho_n = 0.1$, $a = 1$)

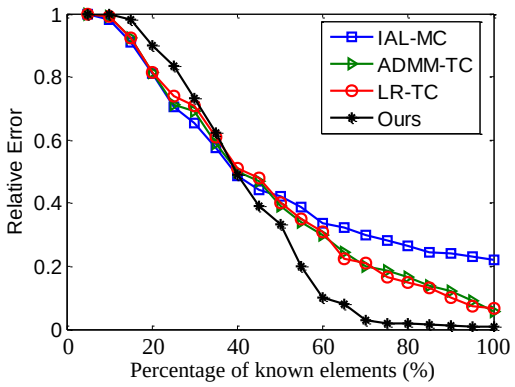


Fig. 3.3 Relative errors for different known elements ($\rho_n = 0.25$, $a = 1$)

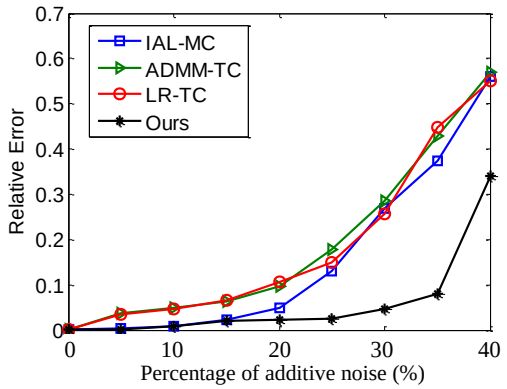


Fig. 3.4 Relative errors for different corruption percentages ($\rho_o = 1$, $a = 1$)

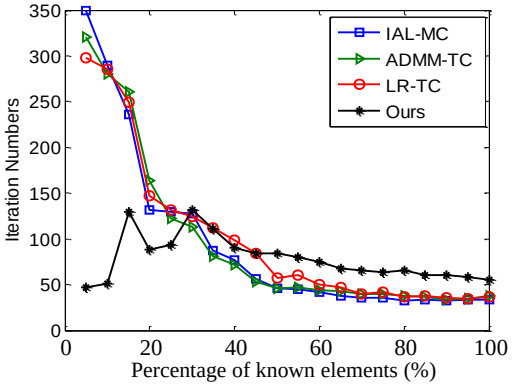


Fig. 3.5 Iteration numbers for different known elements ($\rho_n = 0.15$, $a = 1$)

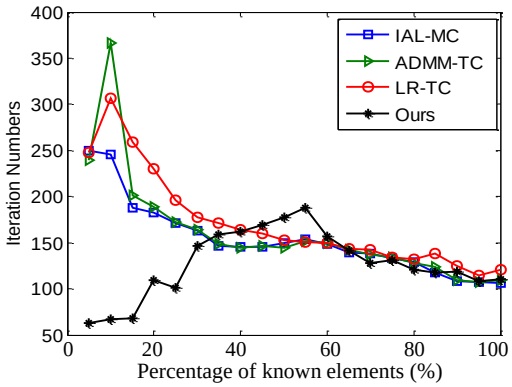


Fig. 3.6 Iteration numbers for different known elements ($\rho_n = 0.3$, $a = 1$)

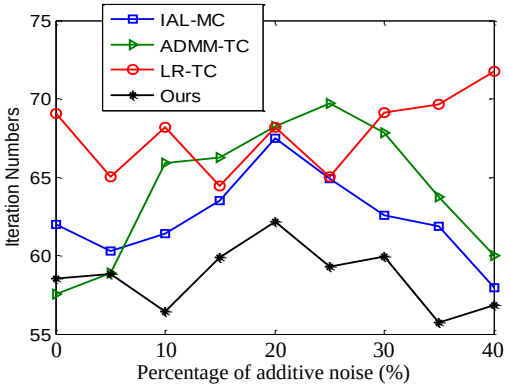


Fig. 3.7 Iteration numbers for different corruption percentages ($\rho_o = 1$, $a = 1$)

that, from 5% to 40% the recovery performance dramatically increases, while it does not show significant improvement from 40% to 100% observation. We then add more polluted

data ($\rho_n = 0.25$, $a = 1$) to test these approaches. As Fig. 3.3, all four methods perform similarly under a circumstance that the missing data are less than 40%, while the proposed method achieves a smaller relative error of the data recovery when more than 50% data are polluted. Combining both Fig. 3.2 and Fig. 3.3, our method appears an obvious "thresholding phenomena" that the recovery accuracy is continuously improved when the observed data increases below a certain threshold (*e.g.,* 40% in Fig. 3.2 and 70% in Fig. 3.3) and the threshold is bigger when adding more corrupted data. Next, we set the observation $\rho_o = 1$ and investigate the recovery performance under different corruption percentages (from 0% to 40%). The results are shown in Fig. 3.4. Similarly, our method shows a relatively higher recovery accuracy (*e.g.,* improving around 3 times under 30% noise pollution) and the other three methods reveal a similar performance. It is worth mentioning that the recovering performance greatly degenerates when more than 35% data are polluted.

Iteration Number

Fig. 3.5∼3.7 compare the computation time of different methods in terms of iteration numbers[10]. In details, Fig. 3.5 illustrates the iteration times needed for different percentages of known elements. We observe that the proposed method is super fast when a few data are observed (*e.g.,* from 5% to 30%) comparing to other solutions, however it requires a similar or slightly higher iteration times when observing more data (*e.g.,* from 40% to 100%). A similar result applies to Fig. 3.6 where the proposed method only needs around 50∼100 iteration times under 5% to 30% known data comparing to other methods that requires 180∼370 iterations. Fig. 3.6 shows the experimental results of iteration numbers for different data corruption percentages with no missing values. Our method overall reveals a slightly better computation efficiency.

[10]In each iteration, all the tensor completion based methods require to calculate 3 times SVD that is most time-consuming computation task so we use iteration number as evaluation metric for computation time.

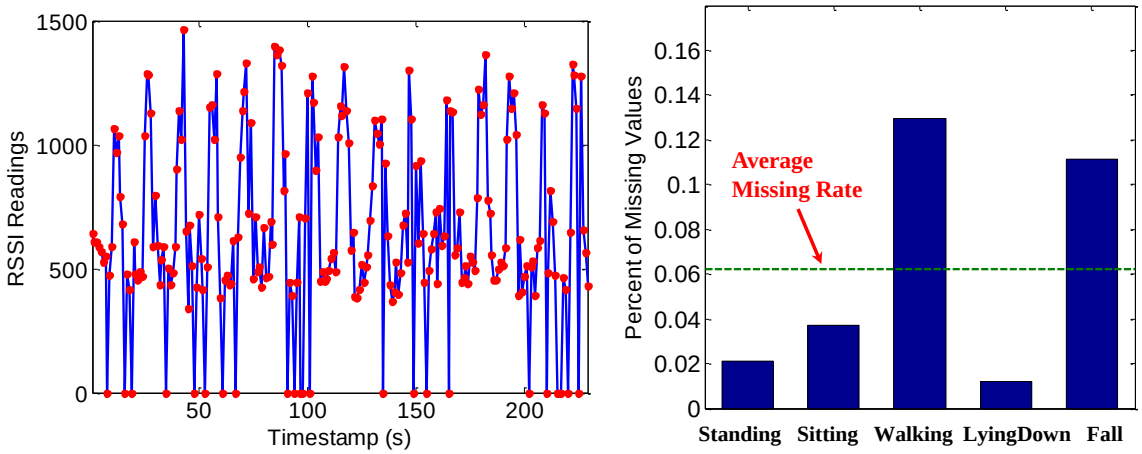


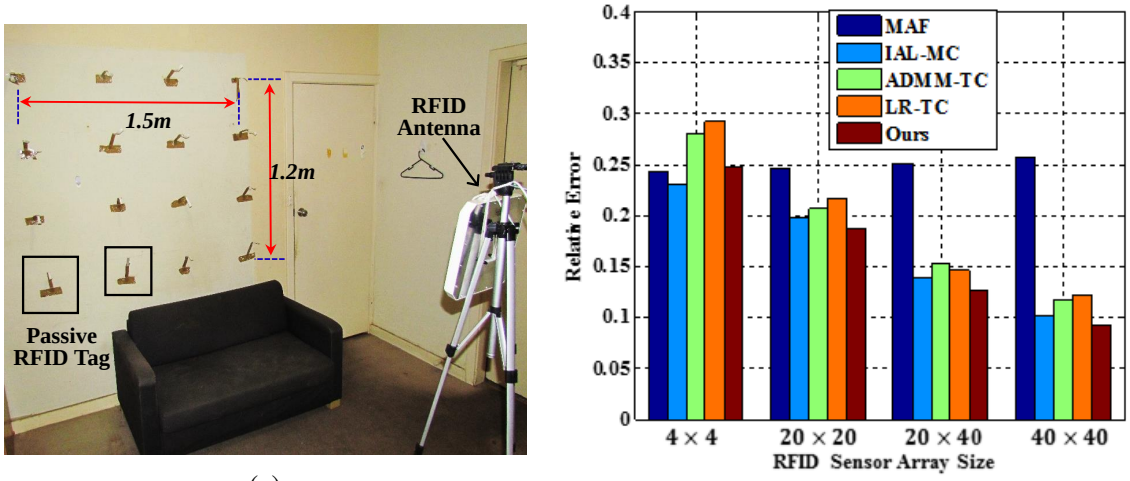
Fig. 3.8 Left: The phenomena of RSSI readings loss in passive RFID tags; Right: The missing rates of RSSI readings from a practical Human Activity Recognition system built upon a passive RFID tag-array

In summary, the experimental results on the synthetic data suggest that our model can achieve better recovery performance and computation efficiency with both partial and polluted observations, especially for the scenarios that only very limited data are known such as less than 30%.

3.4.3 Evaluations on RFID Sensory Data

Passive RFID tags are one of the most frequently-used sensors due to its cheap price (<5 cents each) and battery-free features [129–131]. It is widely used to identify and track objects through remotely accessing the electronically stored data. However, since passive RFID tags are powered by radio signals and deliver the data via the weak backscatter signal, they experience severe RSSI reading loss, especially with a high sampling rate or when tag/reader is moving [132]. As a result, how to accurately recover missing RSSI readings is still a challenge, especially for a large-scale RFID usage.

To deal with this practical issue as well as to test our method, we designed a testbed consisting geo-tagged 4×4 RFID sensor array (see Fig. 3.9a) and collect overall 115,200



(a) (b)

Fig. 3.9 (a) Experimental testbed of RFID sensor array; (b) Relative errors for different tag-array size with 20% missing values

RSSI readings[11]. Fig. 3.8 illustrates an example of the sensor readings loss in a practical RFID-based HAR system.

To be more practical, we formulate the readings of RFID sensor array it into a tensor with different sizes to simulate various real-world application scenarios (*e.g.,* 4×4, 20×20, 20×40 and 40×40 sensor array). Similarly, we add noise ($\rho_n = 0.1$, $a = 10$) with uniform distribution and randomly choose 20% elements as the unknown in our experiments. Fig. 3.9b shows the recovery results of our method and MAF (most frequently used in practical RFID system) as well as other matix-tensor completion methods. For small-scaled deployment (*i.e.,* 4×4 sensor array), our method achieves similar performance to MAF. However, the tensor-based methods perform significantly better than MAF in a large-scale deployment (*e.g.,* 20×20 sensor array). The lack of performance improvement in 4×4 sensor array mainly lies in the fact that the low-rank structure only exists in mode-3 of data tensor which conflicts our assumption that requires low-rank in all tensor modes.

[11]We collect over all one hour's RSS readings, the sampling rate is 2Hz. During the data collection, a participant is doing various activities between the RFID sensor-array and reader, including walking, sitting, standing, lying down as well as falling down etc. By doing so, the collected RSSI reading will reveal different patterns.

3.5 Conclusion

In summary, we propose a method for recovering the missing data by using the robust tensor completion. The proposed method can accurately recover the missing values given partial observed corrupted data.

Our whole system, especially the indoor localization, human activity recognition and fall detection parts, is built upon passive RFID tags, which suffer severe reading loss. Our proposed tensor-based data recovery method can significantly reduce the influence of missing data to the system performance. In the next chapter, we will illustrate how we purely utilize passive tags to achieve a high-accuracy indoor localization and tracking.

Chapter 4

Device-free Human Localization and Tracking Using Passive RFID Tags

Localizing and tracking human movement in a *Device-free and Passive* (DfP) manner is promising in two aspects: *i)* it neither requires users to wear any sensors or devices, *ii)* nor it needs them to consciously cooperate during the localization. Such a DfP indoor localization technique underpins many real-world applications such as shopping navigation, intruder detection, surveillance care of seniors. However, current passive localization techniques either need expensive/sophisticated hardware such as ultra-wideband radar or infrared sensors, or have an issue of invasion of privacy such as camera-based techniques, or need regular maintenance such as the replacement of batteries. In this chapter, we build a novel *data-driven* DfP localization and tracking system upon a set of commercial UHF (Ultra-High Frequency) passive RFID (Radio-Frequency IDentification) tags in an indoor environment. In particular, we formulate human localization problem as finding a location with the maximum posterior probability given the observed RSSIs (Received Signal Strength Indicator) from passive RFID tags. In this regard, we design a series of localization schemes to capture the posterior probability by taking the advance of supervised-learning models including Gaussian Mixture Model (GMM), k Nearest Neighbor (kNN) and Kernel-based Learning. For tracking

a moving target, we mathematically model the task as searching a location sequence with the most likelihood, in which we first augment the probabilistic estimation learned in localization to construct the Emission Matrix and propose two human mobility models to approximate the Transmission Matrix in the Hidden Markov Model (HMM). The proposed HMM-based tracking model is able to transfer the pattern learned in localization into tracking but also reduce the location-state candidates at each transmission iteration, which increases both the computation efficiency and tracking accuracy. The extensive experiments in two real-world scenarios reveal that our approach can achieve up to 94% localization accuracy and an average $0.64m$ tracking error, outperforming other state-of-the-art RFID-based indoor localization systems.

4.1 Introduction

With the increasing aging population, intelligent space that can better support the independent living of the elderly has been attracting the increasing attention both from industry and academia. One of the key preconditions for such a smart environment lies on an accurate and timely detection of users' locations and daily routines [43, 133], especially for an *indoor environment* that GPS (Global Position System) cannot handle [26]. To tackle this challenge, a wide range of indoor localization and tracking systems have been proposed for the last two decades, including but not limited to LANDMARC [22], WILL [134], Tagoram [23] and BackPos [135]. However most of the approaches are wearable-device based technique that inevitably requires the user to actively carry one or more devices such as various types of sensors, smart-phones, RFID tags/readers or other Radio Frequency (RF) transceivers, thus raising many inherent impractical issues in reality [44]. For example, the attached sensors/tags may be damaged or lost. It is also obstructive and inconvenient for the user

to wear devices all the time[1], especially considering that many electronic devices have a moderate size or weight.

To this end, *device-free* (also called *unobtrusive*) passive indoor localization has gained more attention lately and many promising approaches have been proposed [37, 33, 32, 47]. One popular device-free human tracking technique is built upon the recent advance of computer vision, which develops various models to capture human movement from images or videos by using RGB cameras [29, 30], or infrared sensors [27] or depth cameras (*e.g.,* Kinect) [31]. However computer vision based approaches require the tracked user within the line-of-sight[2] (LOS) of a camera, and usually fail to work in a dimmed environment [29]. Moreover, vision-based technique can also be considered to be privacy invasive [43]. Another DfP localization technique is to intensively exploit the radio-frequency signal, *e.g.,* localizing the target by analyzing the Received Signal Strength (RSS) variations [34, 38, 133] or Channel State Information (CSI) [136, 137] in WIFI, or tracking the user through a wall by decoding the radiowaves reflected of human movement [41]. Though promising, these systems often require specialized RF signals such as Frequency-Modulated Continuous Wave (FMCW) or build upon costly special-purpose devices such as USRP (universal software radio peripheral), or need to modify the low-level firmware such as abstracting CSI signals. Most importantly, they all require regular maintenance such as battery replacement, thus hindering their practical deployment in the real world [43, 44]. In this regard, device-free tracking systems built on COTS (commercial off-the-shelf) passive RFID tags are more promising in terms of deployment convenience (commercialized product without any hardware or firmware modification), maintenance effort (no batteries needed and purely harvesting the in-air backscattered energy) and cost efficiency ($\approx$5 cents each, still dropping quickly) [39, 138]. As a result, in this chapter, we design a DfP system that can unobtrusively localize, track a subject to high accuracy based on *pure* passive RFID tags.

[1]Deloitte Mobile Consumer Survey 2016: www2.deloitte.com/au/en/pages/technology-media-and-telecommunications/articles/mobile-consumer-survey-2016.html

[2]There are no barriers or blocks between the subject and camera.

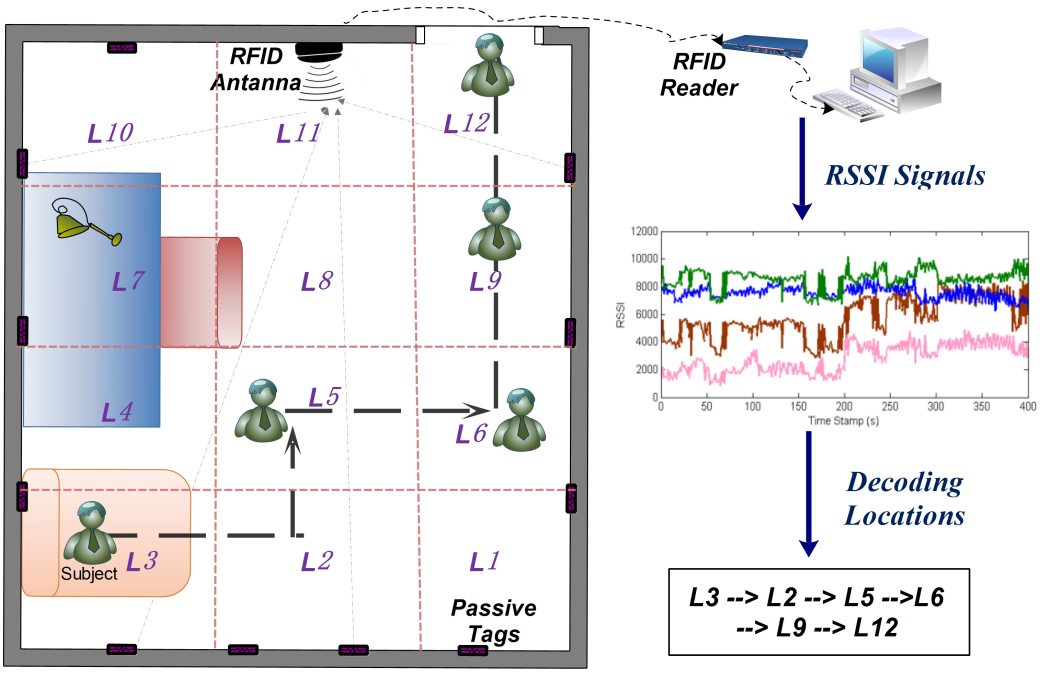


Fig. 4.1 The general idea of the proposed DfP localization and tracking system

However, applying this high-level idea into a practical indoor localization and tracking system is a non-trivial and challenging task. One key challenge lies on the fact that, in a practical residential environment, RSSI signal is quite complex and unstable because of the multipath effect, power source fluctuation and ambient noise disturbance. Unlike the theoretical analysis, the practical RSSI signal however does not strictly decrease along with tag-reader distance and exhibits significant nonlinearity, and it may be further corrupted when introducing human motion. Another challenging issue is how to model the localization and tracking problem from a data-driven point of view. Currently, most of existing RFID-based systems are built upon the signal propagation model or backscatter communication mechanism, thus there is no off-the-shelf learning-based localization model for us to use. Moreover, to reduce the learning burden, we intend to transfer the pattern learned in localizing a stationary person into tracking a moving subject. Thus how to effectively bridge the gap

between localization and tracking under a feasible mathematical framework also deserves a careful resolution.

To tackle the aforementioned challenges, we first need to enable the RSSI signal from passive tags to monitor the whole surveillance area in an efficient and unobtrusive manner. Thus we deploy a set of passive RFID tags and a reader (with antennas) to form a RSS field that can cover the whole monitored area. Fig. 4.1 outlines the general hardware deployment in our system. In particular, unlike other RFID-based systems that place the tags on the ground [32, 33], we attach the passive tags and antennas on the wall to *i)* make the RSSI signal face fewer obstacles and *ii)* not obstruct to user's routine activities, especially in a residential environment. Based upon our RFID infrastructure, some distinguishable patterns can be clearly observed in RSSI signals when a user appears in different locations of a room. In summary, our RFID-based system is intuitively based on two experimental observations:

Observation 1 *The RSSI vector illustrates differentiable changes when a user appears in an RSS-monitored area comparing to a non-subject scenario.*

Observation 2 *The RSSI vector reveals distinguishable fluctuation patterns when a user presents in different locations within an RSS-monitored zone.*

The above two observations substantially illustrate that distributions of a RSSI vector[3] are directly relevant with a user's indoor positions, and those distributions are differentiable for different locations. Motivated by these two experimental phenomena, we thus seek to decode human locations and motions by using data-driven approaches. In particular, to localize a stationary person, we mathematically formulate it as a classification problem, in which we first collect the RSSIs and associated location labels to train a location classifier that is then utilized to predict user's actual location according to the observed RSSI vector (see details in Sec. 4.4). For tracking a moving user, we first augment the traditional kNN with

[3]For example, in Fig. 4.1, we can formulate the RSSIs of all tags at a certain time-stamp as a vector containing 11 readings.

probabilistic information to quantify the likelihood of locations based on observed RSSIs, which then is utilized to construct the Emission Matrix in HMM. Furthermore, we calculate the Transmission Matrix by introducing two location transition strategies - *Constraint-Less Transition* (CLT) and *Constraint Transition* (CT). The latter transition strategy allows our system to largely narrow down the candidate locations at each state transmission in HMM, which turns out to only not minimize the computation overhead but also increase the tracking accuracy (see details in Sec. 4.5). At last, we use Viterbi Search to find the most likely path of the subject. We call this *k*NN-HMM. In a nutshell, we summarize the main contributions in the chapter as below:

- We design a device-free indoor localization and tracking system that utilizes COTS passive RFID tags and bears some promising characteristics in terms of hardware cost, deployment scalability and maintenance burden. To the best of our knowledge, the designed system, purely built upon passive RFID tags, is one of the device-free works that can not only localize a *stationary* user but also track a *moving* one with a high accuracy in a real-world *residential* environment.

- We introduce a *k*NN based HMM method to tracking a motion person by learning the underlying impacts of a non-moving human body to RSSIs for different locations, which to some extent bridges the gap of localization to tracking from a data-driven point of view.

- We conduct extensive in-situ experiments in a real-world residential house where participants unconstrainedly simulate a series of practical daily living routines. The experimental results demonstrate that our system achieves over 94% localization accuracy and 0.64m mean tracking error while largely reducing the training overhead to 2 minutes for a 17m^2 bedroom.

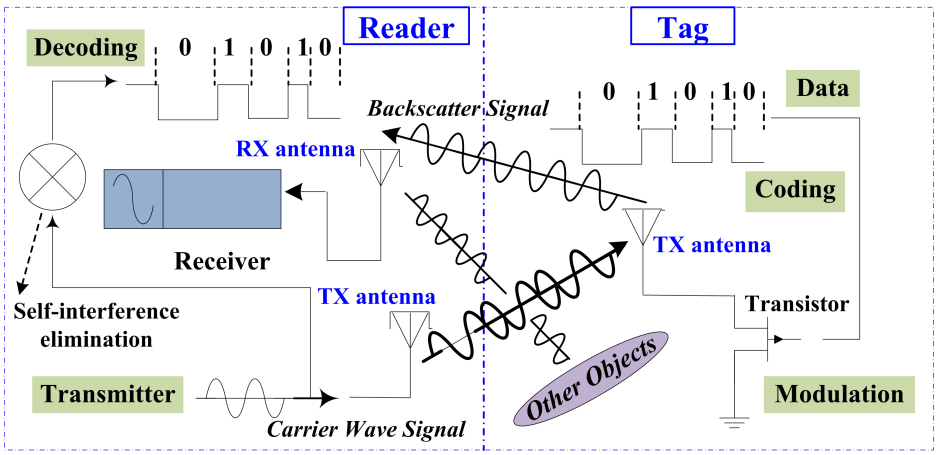


Fig. 4.2 Backscatter communication mechanism

We organize this chapter as follows. Sec. 4.2 illustrates our preliminary analysis and experiential observations. We then mathematically model our target localization and tracking problems in Sec. 4.3. We highlight the proposed solutions in Sec. 4.4 and Sec. 4.5. The experimental results are presented in Sec. 4.6. Finally, some discussions and concluding remarks are offered in Sec. 4.7.

4.2 Preliminary

In this section, we will theoretically analyze the RFID backscatter radio signal and then verify our system's capability to reach device-free localization and tracking.

4.2.1 Backscatter Radio Communication

RFID tags are widely applied in many industries, for example, an RFID tag attached to an automobile during production can be utilized to monitored its progress in the assembling, RFID-tagged containers can be tracked during the transportation [139, 140]. Unlike active RFID tags that are powered by batteries, passive RFID systems however communicate through the backscatter radio links due to that passive tags (no batteries powered) can only passively collect energy from the in-air backscattered radio signal. Fig. 4.2 illustrates a

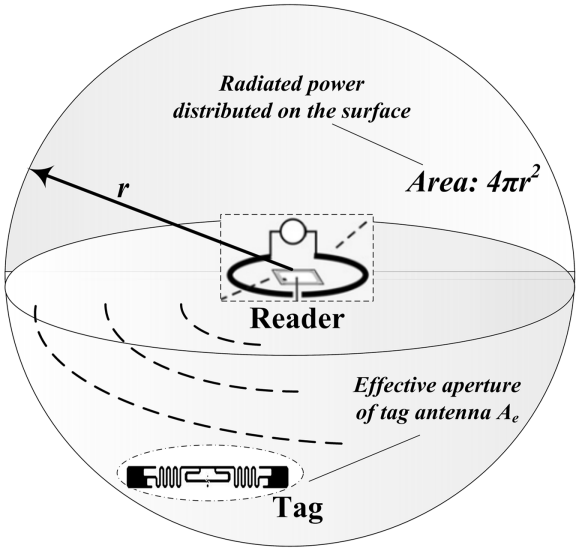


Fig. 4.3 Path loss illustration

conceptual diagram of the radio wave propagation between an RFID reader and a passive tag. In details, the current flow on a reader-antenna induces to a voltage on the tag-antenna (integrated in the circuit), further producing a radiation signal. The radiated wave then makes its way back to the reader-antenna, inducing a voltage, thus producing a signal that can be detected: a backscattered signal. In particular, the tag transmits “1" bit by changing the impedance on their antennas to reflect the reader’s signal and a “0" bit by remaining in their initial silence state [141], called ON-OFF keying. A typical UHF reader works in the frequency band from 860 MHz$\sim$950 MHz (*e.g.,* 902 $\sim$ 928MHz ISM band in US). Today’s commercialized RFID readers have an interrogation distance of about 10 *meter*, which is enough for a residential environment. More importantly, the electromagnetic field produced by RFID readers under no circumstance will harm the human body[4].

4.2.2 Received Signal Strength Indicator (RSSI)

RSSI measures the power of received radio signal between the tag-antenna and reader-antenna [141]. Shown as Fig. 4.3, *Path Loss* represents the power difference of signals from

[4]Is RFID Dangerous? www.inria.fr/en/centre/lille/news/is-rfid-dangerous

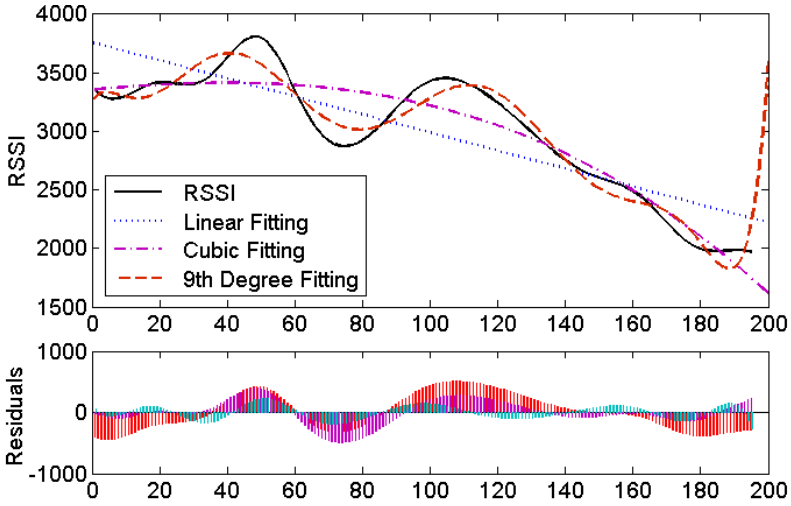


Fig. 4.4 RSSI variation with distance

the receiving antenna and the transmitting antenna. We assume the radiated power as being uniformly distributed over a spherical surface at given distance r from the reader-antenna. Then, only part of this power is received by a tag-antenna, represented as $P_{RX} = P_{TX}A_e/4\pi r^2$. Since the effective aperture of an antenna around a half-wavelength long corresponds to a square round a half-wavelength on a side, the path loss for the isotropic link can be estimated by $A_e = G\lambda^2/4\pi$ where G denotes the gain of an antenna. Thus we can calculate *Friis Equation* of the power from the transmission-antenna TX to the receiver-antenna RX [141].

$$P_{RX} = P_{TX}G_{TX}\frac{A_{e,RX}}{4\pi r^2} = P_{TX}G_{TX}G_{RX}(\frac{\lambda}{4\pi r})^2 \tag{4.1}$$

Then, we can mathematically model the backscatter signal prorogation as:

$$\begin{aligned} P_{RX,reader} &= P_{TX,tag}G_{tag}G_{reader}(\lambda/4\pi r)^2 \\ &= P_{TX,reader}T_bG_{tag}^2G_{reader}^2(\lambda/4\pi r)^4 \end{aligned} \tag{4.2}$$

where G_{tag} denotes the gain of the tag-antenna and T_b represents the loss of backscatter transmission. Thus, under an assumption that a wave directly leaves the antenna and strikes the tag (*i.e.,* interacting with no other objects), Eqn. 4.2 theoretically demonstrates that

the power received by the reader-antenna is inversely proportional to the fourth power of the reader-tag distance. Thereby, for a cleared or open space, RSSIs is capable of being a promising location indicator. However, our system targets to enable a device-free tracking in a cluttered environment. As Fig. 4.4 shows, the RSSI strength shows a uncertain nonlinearity with the distance in a residential room, which cannot be expressed by a *cubic* or even a *9th-degree polynomial* model. So how to model the RSSI-location relation for our application scenario is very challenging. Instead of developing delicate signal propagation models[5], this chapter intends to seek the answer from a data-driven point of view, *i.e.,* accurately learning the quantifying relation between the user's location and the interference of human body to RSSIs from the collected RSSI observations. We will elaborate the details in Sec. 4.4.

4.2.3 Intuitions Verification

In this section, we conduct several pilot experiments to demonstrate the localization potentials of our system. We first build a testbed consisted of one RFID reader and 4 UHF passive tags. The monitored area is divided into 9 virtual grids ($0.6m \times 0.6m$ each), representing 9 different zones $L_1, L_2, ..., L_9$. We want to verify whether the RSSI patterns reveal distinguishable differences when a user appears in different grids. Fig. 5 snapshots our pilot experimental results. At first, there is no user in the monitored area, then a person stands in L_5 and L_9. We observe that the measured four RSSI signals obviously vary due to the presence of a subject, so we can clearly discriminate whether there is a subject in the RSS field or not. We also find that the RSSI signal shows different fluctuation patterns when the subject stands in L_5 and L_9. We further cluster the RSSI data generated from these three scenarios (*i.e.,* no subject, L_5 and L_9) into a four-dimension space (illustrated by two 3-D scattering figures). It clearly shows the data clustering in three different subareas (revealing the number of locations the subject ever appeared) even without overlapping (can be learned to infer the exact human locations).

[5]This kind of models is also highly related to the furniture and room layout, thereby it is hard to design a physical localization model with satisfying robustness and accuracy.

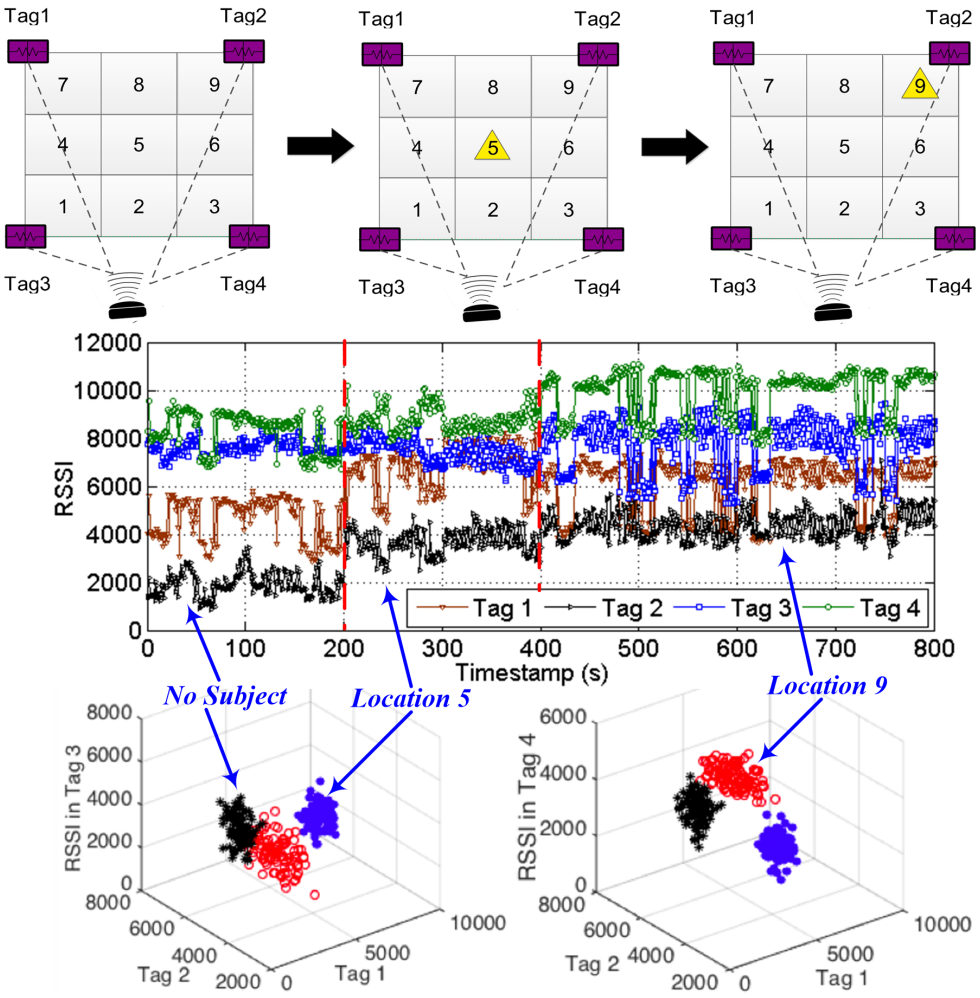


Fig. 4.5 The RSSI readings cluster in differentiable spaces when a person appears in different locations

In summary, the preliminary experiments reveal the intuitions and feasibility behind our system for solving the device-free localization. However, in a residential environment, how to accurately decode the accurate locations is still a non-trivial problem considering the complicated multi-path effect and the unstable backscattered RSSI propagation properties. We will elaborate it in Sec. 4.5.

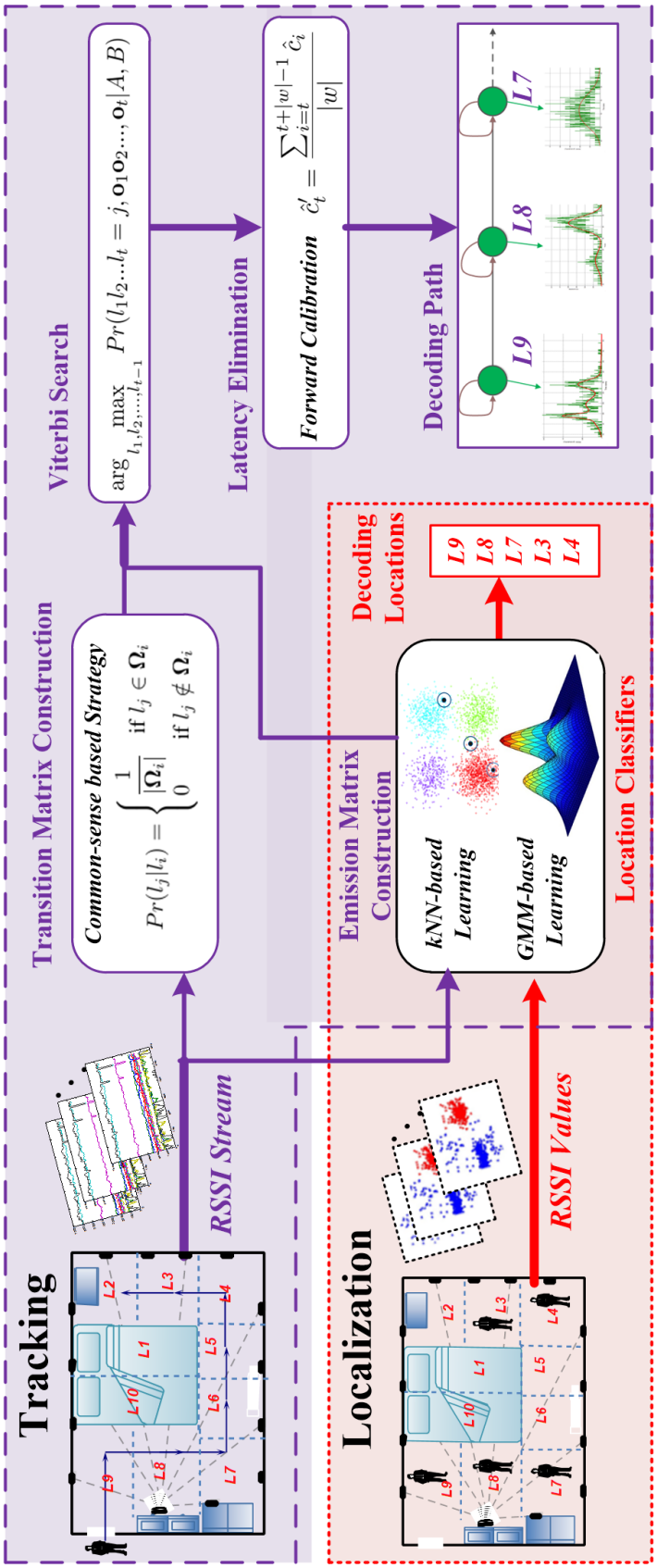


Fig. 4.6 The system architecture

4.3 Problem Formulation

As aforementioned, we intend to pinpoint the subject's locations and estimate its continuous trajectory based on the received RSSIs from a set of RFID tags. Thus we can formally define the two targeted problems - *localization* and *tracking* - in this chapter as follows.

Problem 2 (Localization) *In a monitored area covered by one or more RSS fields, can we accurately pinpoint the current location of a stationary user given a set of RSSI vectors?*

Problem 3 (Tracking) *In a monitored area covered by one or more RSS fields, can we continuously estimate the motion trajectory of a moving user with a moderate speed (less than $1m/s$) given a sequence of time-tagged RSSI vectors?*

Fig. 4.6 illustrates the pipeline of our solutions for the two problems. From a data-driven point of view, *Problem 1 - Localization* can substantially be reformulated as a location classification problem, in which we aim to accurately quantify the RSSI distributions for different geographical locations within the monitored area. In particular, assuming that D anchoring passive tags are deployed in a surveillance area which is divided into G small grids, we then can represent the locations as $\mathbf{l} = \{l_0, l_1, ..., l_G\}$ where l_i means the subject appears in location i and l_0 indicates the area is empty. Next, we collect profiling dataset in the following two steps: *i)* we record the RSSI readings of all anchoring tags when no human body in the monitored area; and *ii)* then a user appears in location $l_i, (i = 1, 2, ..., G)$ and collect the corresponding RSSI values. Then we build a training dataset $\mathscr{H} = \{\mathbf{S}_0, \mathbf{S}_1, ..., \mathbf{S}_G\}$, where $\mathbf{S}_i \in \mathbb{R}^{N \times D}$, N is the sample number in each grid. This dataset contains the latent information regarding how a human body influences the RSSIs' distribution for each location plus an empty environment. We further can quantify the underlying *RSSI-Location* relationship by training a classification model using $\mathscr{H}$. Finally, we construct a $(G+1)$-location classifier. During localization phase, a user randomly stands on any locations in the surveillance area,

and the corresponding RSSI vectors are collected and fed into the location classier. Then it will output location labels that associate with the subject's actual locations.

Assuming that the collected RSSI observation dataset is represented by $\mathscr{R} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_T\}$, Problem 2 is mathematically formulated as estimating the optimal posterior probability distribution $p(l_j|\mathbf{r}_i)$ given a RSSI observation sequence.

$$j^* = \arg\max_j Pr(l_j|\mathbf{r}_i) \tag{4.3}$$

In Sec. 4.4, we will give the technical details regarding how to solve the above optimization problem. Similarly, for *Problem 3-Tracking*, we can model it as estimating the joint probability distribution upon the RSSI observation sequence $R_{1:T}$ and the location labels $l_{1:T}$ where its location state at time-stamp t is denoted by l_t. We can further simplify the model by assuming that the dynamic motion is a Markov process which only depends on previous location state, represented by model $Pr(l_j|l_{j-1})$. In this end, we need to solve the following mathematical problem:

$$Pr(\mathbf{r}_{1:T}, l_{1:T}) = Pr(l_1)Pr(\mathbf{r}_1|l_1)\prod_{t=2}^{T} Pr(\mathbf{r}_t|l_t)Pr(l_t|l_{t-1}) \tag{4.4}$$

to estimate the expected location states $l_{1:T}$ with the maximum probability. We also need to train a marginal posterior $Pr(\mathbf{s}_i|l_{1:j})$ to estimate the expected value of l_j given observed RSSI readings. We will introduce the technical details in Sec. 4.5.

4.4 Localizing Stationary Subject

This section will introduce three location classifiers, *i.e., Multivariate Gaussian Mixture Model*, *k Nearest Neighbor*, and *Kernel-based Localization* for solving Problem 1 - estimating user's location given a set of RSSI vectors.

4.4.1 Gaussian Mixture Model based Localization

According to our previous analysis, the key part of localization is to model $Pr(l_j|\mathbf{r}_i)$, the probability distribution of locations given RSSI observation. This task is difficult since it needs to quantify the distribution of an underlying variable. However, the reversed distribution $Pr(\mathbf{r}_j|l_i)$ can be easily learned by observing how RSSIs distribute given the location of a user. Based on the *Bayes Theorem*, we thereby decompose the distribution $Pr(l|\mathbf{r})$ as follows[6] :

$$Pr(l|\mathbf{r}) = \frac{Pr(\mathbf{r}|l)Pr(l)}{Pr(\mathbf{r})} \propto Pr(\mathbf{r}|l) \cdot Pr(l) \tag{4.5}$$

where we assume $Pr(l) \sim 1/G$, denoting an uniform distribution at location l. The assumption lies on the fact that a user may appear in any locations with an equal probability. Next, we need to find an appropriate model that quantifies $Pr(\mathbf{r}|l)$ distribution. Then we can transfer Eqn. 4.3 as the following optimization problem.

$$l^* = \arg\max_{l \in \mathbf{l}} \; Pr(\mathbf{r}|l) \cdot Pr(l) \tag{4.6}$$

In our pilot experiment, we observe that RSSIs display a certain clustering pattern in the high-dimension space. When we take a close look at each cluster, it actually shows a multi-modal distribution that follows a Gaussian Mixture Model, as shown in Fig. 4.7. This RSSI distribution phenomenon in fact can be explained by the multi-path effect [141, 142]. Normally, several paths for the backscattered signal exist during the propagation from a tag to a reader. Among all the paths, the reader prefers to resolve the strongest signal path. When a human body blocks some propagation paths (*i.e.,* a subject appears in the RSS field), it will cause the propagation to jump among the multiple paths and lead to the strength migrating from one level to another. As a result, the signal strength exhibits multi-modal

[6] For simplicity, we drop i and j in the equation.

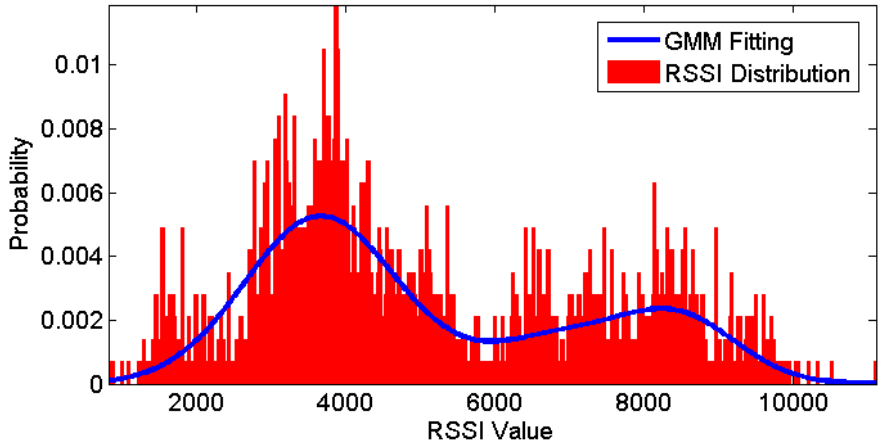


Fig. 4.7 RSSI distribution pattern and fitted by GMM

characteristics - the distribution is likely composed of multiple Gaussian models. Thus, we can utilize a GMM to capture the probability distribution when a user appears in each grid. In particular, assuming that a Gaussian Mixture Model has q_m^l components, mean μ_m^l and covariance matrix σ_m^l applies for location l, then we have

$$\begin{aligned} f_l(x) &= Pr(x|l) = \sum_{m=1}^{M} q_{l,m}\mathcal{N}(x|\mu_{l,m},\sigma_{l,m}) \\ &= \sum_{m=1}^{M} \frac{q_{l,m}}{\sqrt{(2\pi)^{\mathscr{D}}|\sigma_{l,m}|}} \exp(-\frac{1}{2}(x-\mu_{l,m})^T\sigma_{l,m}^{-1}(x-\mu_{l,m})) \end{aligned} \tag{4.7}$$

where $\phi_l = \{q_{l,m},\mu_{l,m},\sigma_{l,m}\}$ represents the model parameter set for location l, in which $q_{l,m}$ means the weighted factor for the m-th mixture component, $\mu_{l,m}$ and $\sigma_{l,m}$ denote the mean and covariance in the m-th Gaussian component. Furthermore, by using the maximum likelihood estimation, the optimal model parameters $\hat{\phi}_l$ can be learned through

$$\hat{\phi}_l = \arg\max_{\phi_l} Pr(x|l,\phi_l) = \arg\max_{\phi_l} \prod_{i=1}^{N} Pr(\mathbf{s}_i|l,\phi_l) \tag{4.8}$$

where $\mathbf{s} = \{\mathbf{s}_1,\mathbf{s}_2,...,\mathbf{s}_N\}$ denotes the training dataset.

To solve the optimization problem in Eqn. 4.8, we adopt Expectation Maximization (EM), which iteratively optimizes the object function by two steps - E-step (*Expectation step*) and M-step (*Maximization step*). Basically, the expectation step calculates the posterior probability $Pr(l|\mathbf{s})$ by using the training dataset $\mathbf{s}$. The Maximization step maximizes the log-likelihood expectation, which in turn enables us to re-calculate the parameters in the following iteration. We use cross validation to find an optimal value of GMM component number that maximize the localization accuracy. With a learned GMM location classier, we can first calculate all the probabilities for candidate locations $\mathbf{l}_{1:G}$ given an observed $\mathbf{r}$, and then we choose the maximal one as the predicted location of the user.

4.4.2 k Nearest Neighbor based Localization

Another way to build a location classier is to learn the Euclidean distances of RSSI vectors under a resident appearing on a certain candidate locations. In this regard, we introduce the k nearest neighbors (kNN) method that first measures the context-dependent Euclidean distances between a testing RSSI vector with the RSSI vectors of training dataset, and then use a majority vote among its nearest neighbors to assigns a location label. In particular, assuming that we have a training dataset $\mathbf{T} = \{(\mathbf{s}_1, y_1), (\mathbf{s}_2, y_2), ..., (\mathbf{s}_N, y_N)\}$ with N samples, where $\mathbf{s}_i \in \mathbb{R}^D$ is the RSSI vector, $y_i \in \mathbf{l} = \{l_1, ..., l_G\}$ is the corresponding location label. Then, given a distance measuring method and a testing RSSI vector $\mathbf{r}$, we can search its k nearest neighbors, represented by $N_k(\mathbf{r})$. Finally, the testing RSSI vector is given a most-common location label y^* among its k nearest neighbors by following equation.

$$y^* = \arg\max_{l_j} \sum_{\mathbf{s}_i \in N_k(\mathbf{r})} \mathbb{I}(y_i = l_j) \tag{4.9}$$

where $j = 1, 2, ..., G; i = 1, 2, ..., N$ and $\mathbb{I}$ denotes an indicator function which is 1 if $y_i = l_j$, otherwise 0.

4.4.3 Kernel-based Localization

From the point of probabilistic view, if two RSSI vectors have a stronger similarity, then they will be in a near or even same location with a higher probability. Based on this intuition, we thus can use a Kernel-based learning (KL) to resolve the posterior probability of candidate locations given an RSSI observation. By applying a kernel function in RSSIs, KL can directly construct possible non-Euclidean topologies that are underlaid implicitly in the RSSI vectors and locations. In particular, in the learning procedure, KL will assign the kernel with a probability mass for every RSSI vector of the training dataset. Then, for an observed RSSI vector, multiple density functions with equal weights will be utilized to estimate the probability. Mathematically, given the training data and corresponding location labels $\mathbf{S} = \{(\mathbf{s}_1, l_i), ..., (\mathbf{s}_n, l_n)\}$, the KL-based localization can be formulated as:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n}\xi_i \\ \text{S.t.} \quad & l_i(\mathbf{w}^T\theta(\mathbf{s}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \end{aligned} \tag{4.10}$$

where C means the error penalty, $\xi_i (i = 1, 2, ..., n)$ are slack variables and kernel function is represented by $K(\mathbf{s}_i, \mathbf{s}_j) = \theta(\mathbf{s}_i)^T\theta(\mathbf{s}_j)$. Based on the primal-dual relationship, we can optimize the model parameters by solving the following dual problem [143]:

$$\begin{aligned} \max_{\mu,\alpha}\min_{\mathbf{w},\xi,b} \quad & \mathbf{w}^T\mathbf{w} - \sum_{i=1}^{n}\alpha_i(l_i(\mathbf{w}^T\theta(\mathbf{s}_i) + b) - 1 + \xi_i) \\ & + C\sum_{i}^{n}\xi_i + \sum_{i=1}^{n}\mu_i\xi_i \end{aligned} \tag{4.11}$$

where $\alpha = (\alpha_1, ..., \alpha_n)^T$ and $\mu = (\mu_1, ..., \mu_n)^T$ are Lagrange multipliers. In the testing, we can feed the RSSI observations into the trained model and output the associated location labels. In this chapter, we adopt LibSVM [143] to realize the KL-based localization. The kernel function selection highly depends on the RSSIs' nonlinearity and noise caused path

loss, shadowing and multipath effects in localization. We intensively test the linear kernel, Gaussian kernel, polynomial kernel and radial basis function kernel, finding the linear kernel works better.

4.4.4 Discussion

To summarize, we introduce three different types of localization methods. GMM is motivated by the jumping property of backscattered RF signal from tags, which can be explained by the signal propagation mechanism. *k*NN is based on the similarity measurement of context Euclidean distance of observed RSSI readings. SVM (support vector machine) is an advanced classification method that are widely adopted by other localization systems. Actually, there exists other classification methods that can be applied into our localization system, such as Naive Bayes, Extreme Learning Machine (ELM). We conduct some pilot experiments to compare these methods. In particular, we first ask a subject to stand two minutes in each grids to collect the RSSI samples (the testbed is shown in Fig 5), then we randomly divide the dataset into training and testing datasets in different ratios (from 10% to 90%) to test the methods. As Fig. 4.8 shows, among all the classification methods, *k* Nearest Neighbors achieve the best result. Even with only 10% training data (12 seconds in each grid), it reaches 87.2% accuracy (greatly simplify the pre-calibration and relieve our training burden). It reveals that, with only a few labeled RSSI data, the context-dependent distance measurement can better interpret the fluctuation of RSSI signal caused by human body inference, which strongly motivates our *k*NN-HMM to tackle the tracking problem.

4.5 Tracking a Moving Subject

Comparing to localizing a relatively static user, human tracking is more challenging, especially considering the sudden and unpredictable RSSI changes caused by a moving human

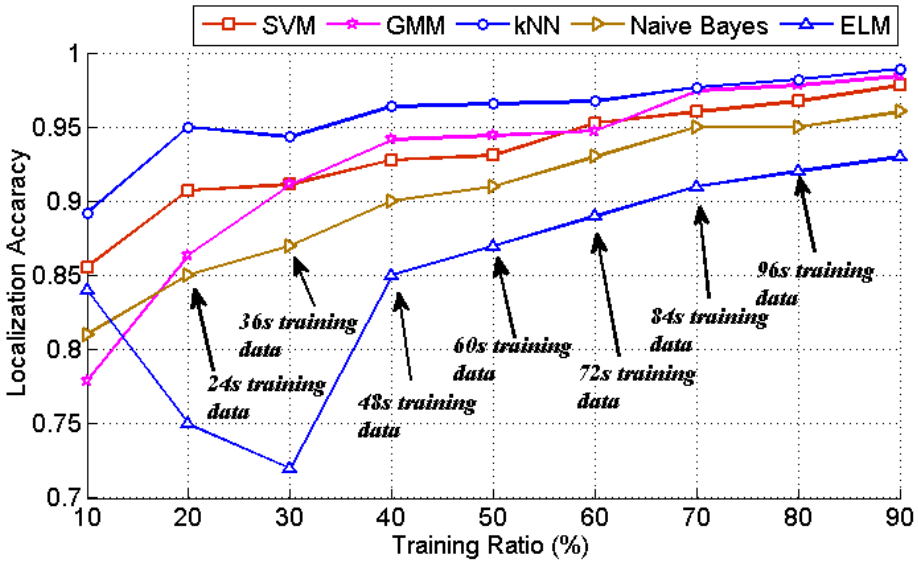


Fig. 4.8 Localization results of different methods

body, which makes the RSSI-Location mapping more difficult. However, on the other hand, within a sampling time, the next moving location will be near to the current location due to the human speed limitation ($\leq 1m/s$), which naturally narrows down the possible candidate locations. In other words, for tracking problem, we have one more evidence, namely *current location state*, that can help us to infer the possible locations besides the RSSI observations. In particular, we propose two HMM-based models, *kNN-HMM* and *GMM-HMM*, to decode the continuously time-stamped RSSIs into the subject's moving path by considering both patterns learned from localization model and the location transition constraints. Hidden Markov Model is widely applied in spatio-temporal pattern recognition such as handwriting recognition, proteins structure prediction and human activity recognition. It can be considered as a generalization of a mixture model where the latent variables, which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other. In this regard, HMM is perfectly fit the assumption of our tracking problem that the next moving location depends and only depends on present location, neither being totally independent nor related to the past location states. Another challenge in

tracking is the latency, namely the subject already moves to next location whiles the system is calculating the current location. To reduce this disturbing phenomenon, given the resulting continuous location points from HMM-based models, we further design a forward calibration mechanism that substantially takes account of a few past location estimations when resolving current location. Next, we will elaborate the details of kNN-HMM based and GMM-HMM based tracking methods as well as the forwarded calibration mechanism.

Assuming that $\mathbf{L}$ represents all candidate user's moving trajectories and $\mathbf{R}$ denotes the observed RSSI vector sequence, then our primary goal is to optimize a trajectory $\mathbf{L}^*$ with a maximum likelihood based on the following equation.

$$\mathbf{L}^* = \arg\max_{\mathbf{L}} Pr(\mathbf{L}|\mathbf{R}) \tag{4.12}$$

According to Bayesian Theorem, we transform optimizing the conditional distribution into finding an optimal joint probability distribution.

$$Pr(\mathbf{L}|\mathbf{R}) = \frac{Pr(\mathbf{L},\mathbf{R})}{Pr(\mathbf{R})} \propto Pr(\mathbf{L},\mathbf{R}) \tag{4.13}$$

Assuming that $\mathbf{R}$ is consisted of T time-tagged RSSI observations $\mathbf{r}_{1:T}$ and $\mathbf{L}$ contains T corresponding location states $l_{1:T}$, we can further decode Eqn. 4.13 as follows:

$$Pr(\mathbf{r}_{1:T}, l_{1:T}) = Pr(l_1)Pr(\mathbf{r}_1|l_1)\prod_{t=2}^{T}\underbrace{Pr(\mathbf{r}_t|l_t)}_{B}\underbrace{Pr(l_t|l_{t-1})}_{A} \tag{4.14}$$

Now we successfully model our tracking problem as a Hidden Markov Model. To solve the model, we first need to estimate *Transition Matrix A* and *Emission Matrix B* and then use *Viterbi Search* to find the optimal location trajectory.

- *Transition Matrix* captures state-transition probability of a user moving from a location-state l_{t-1} at time-stamp $t-1$ to a location-state l_t at time-stamp t. It can be represented via $Pr(l_t|l_{t-1})$.

- *Emission Matrix* models the probability of observing RSSI vector $\mathbf{r}_t$ given a location state l_t at time t, denoted by $Pr(\mathbf{r}_t|l_t)$.

- *Viterbi Searching* finds a location sequence $\{l_1, l_2, ...l_T\}$ that has a maximum likelihood given Transition Matrix A and Emission Matrix B.

4.5.1 Transition Matrix

First of all, we show how we build a transition matrix based on the location state constraint. Generally, the human motion can be seen as a state transition process that next moving location is solely dependent of current state but irrelevant to other states, which can be defined by a probability matrix $A_{ij} = Pr(a_t = l_i|a_{t-1} = l_j)$. To construct such a matrix, we define following two human motion patterns based on an intuition that a person is only able to move a limited distance during one sampling interval (*i.e.,* 0.5 *second* in our system) given the moving speed ($\leq 1m/s$) in an indoor environment.

- *Constraint-Less Transition (CLT):* The tracked user can move to any locations of the monitored area under a same likelihood, namely $l_t \in l_{0:G}$ with an equal probability.

- *Constraint Transition (CT):* The tracked user can only move to one-sampling-time reachable locations of the monitored area under a same likelihood and cannot reach other locations.

The second motion pattern greatly facilitates the tracking efficiency due to the fact that it can largely exclude some unlikely location states in each calculating iteration. For example, in Fig. 4.12, it is impossible for a resident to move from *L11* to *L64* within 0.5 second, so

we can eliminate *L64* from the next moving locations whilst user's current location is *L11*. In this chapter, we categorize the one-sampling-time reachable locations as those grids that are adjacent or equal to user's current location. Mathematically, we formulate these two transition patterns by one equation. We assume that the monitored area is divided into G locations and $l_i(i = 1,2,...,G)$ means the tracked user is in grid i. According to the proposed two motion patterns, we further define a location-state set ω_i including all feasible states that a user can move to given current state l_i, and use $|\omega_i|$ to denote the number of states. We then can construct a transition probability matrix as follows:

$$Pr(l_j|l_i) = \begin{cases} \frac{1}{|\omega_i|} & \text{if } l_j \in \omega_i \\ 0 & \text{if } l_j \notin \omega_i \end{cases} \tag{4.15}$$

4.5.2 Emission Matrix

As Eqn. 4.14 shows, $B_{ij} = Pr(\mathbf{r}_i|l_j)$ represents the emission matrix that essentially shares the same purpose as the localization problem - modeling the RSSI distributions for different location states. As a result, we can construct the emission matrix by taking advantage of aforementioned localization models.

GMM-based Emission Matrix

One straight-forward way is to construct the emission probability matrix based on the GMM model, which is capable of estimating emission probabilities given the RSSI observations. Similar to localization problem, we assume that the probability distribution of RSSI observations follows a multivariate Gaussian Mixture Model for each location state, and we thus are able to calculate the Emission Matrix using Eqn. 4.7.

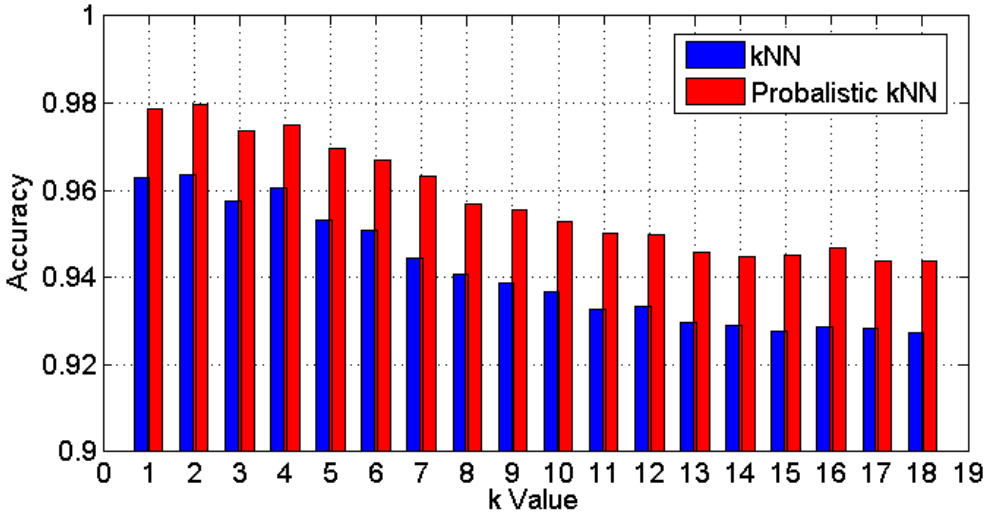


Fig. 4.9 Localization accuracy comparision with k changes

kNN-based Emission Matrix

Another way to construct the emission matrix is taking the merit of k nearest neighbor model which reveals a superiority in mapping the RSSI observations with the latent locations. To do so, we construct a kNN-based emission matrix by transforming a traditional kNN classier into a probabilistic style that can give an emission probability conditioning on the observed RSSIs.

In particular, the probabilistic kNN estimates the *Emission Matrix* as follows. We first search the top-k nearest neighbors $N(\mathbf{r}_j)$ in the profiling dataset for observed RSSI $\mathbf{r}_j$. Then we also mark these searched samples by its belonging locations, represented by $N^i(\mathbf{r}_j) = \{\mathbf{s}_k | \mathbf{s}_k \in \mathscr{N}(\mathbf{r}_j) \cap \mathbf{s}_k \in l_i\}$. Then the probabilistic kNN-based emission matrix is built as follows:

$$Pr(\mathbf{r}_j|l_i) = \frac{\sum_{\mathbf{s}_k \in \mathscr{N}^i(\mathbf{r}_j)}^{|\mathscr{N}^i(\mathbf{r}_j)|} \frac{1}{dis(\mathbf{r}_j, \mathbf{s}_k)}}{\sum_{\mathbf{s}_{k'} \in \mathscr{N}(\mathbf{r}_j)}^{|\mathscr{N}(\mathbf{r}_j)|} \frac{1}{dis(\mathbf{r}_j, \mathbf{s}_{k'})}} \tag{4.16}$$

where $dis(\mathbf{r}, \mathbf{s})$ represents two RSSI vectors' Euclidean distance.

We conduct a pilot experiment to compare probabilistic *k*NN and transitional *k*NN as well. We first collect 2 minutes training data in each grid, then use 40% as the training data and 60% as the testing data to test the methods. As Fig. 4.9 shows, the proposed probabilistic *k*NN method slightly outperforms traditional *k*NN in all k values. More importantly, the probabilistic *k*NN is capable to estimate the posterior possibilities by measuring the context distances. Overall its advantages lie in: *i)* it specifically gives the posterior distribution of each class rather than assigning a class-membership to the test sample; and *ii)* it assigns each neighbor a weight that is inverse-proportional to its distance with the test sample, which not only considers the number of its most-common neighbors but also measures their relative distances.

4.5.3 Viterbi Searching

Given a sequence of observations, Viterbi searching, one of the dynamic programming algorithms, can find an optimal sequence of hidden states with a maximum likelihood, especially being efficient in solving HMM. In particular, assuming that the length of time-stamped RSSI observations is t and the ending location state is l_j, Viterbi searching finds the most likely sequence of latent location states as following induction process.

$$V_j(t) = \arg \max_{l_1, l_2, \ldots, l_{t-1}} Pr(l_1 l_2 \ldots l_t = j, \mathbf{r}_1 \mathbf{r}_2 \ldots, \mathbf{r}_t | A, B) \tag{4.17}$$

where matrix A and B refer to Eqn. 4.14. By induction, we further obtain:

$$\begin{aligned} V_j(1) &= B_j(\mathbf{r}_1) \\ V_j(t+1) &= \arg \max_i V_i(t) A_{ij} B_j(\mathbf{r}_{t+1}) \end{aligned} \tag{4.18}$$

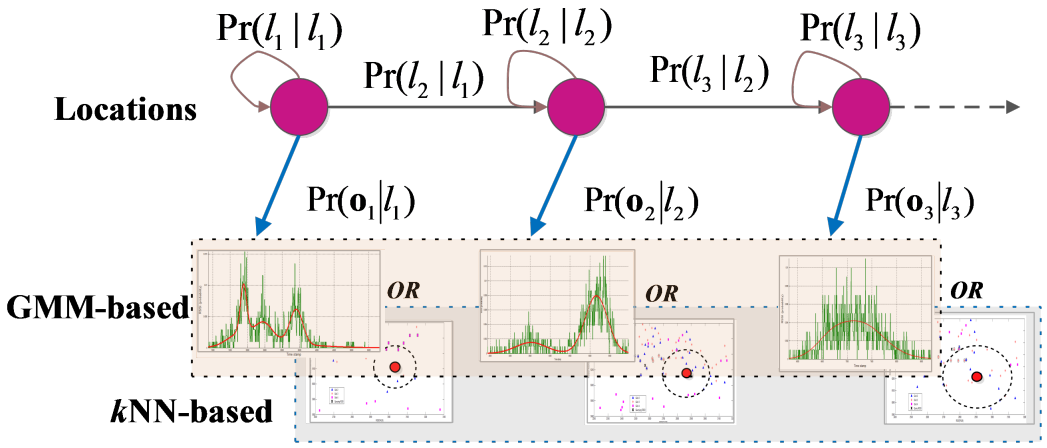


Fig. 4.10 HMM based methods

where $B_j(\mathbf{r}_1) = Pr(\mathbf{r}_1|l_j)$ and $A_{ij} = Pr(l_j|l_i)$. After the induction calculation, we finally can search an optimal moving trajectory for both GMM and *k*NN based HMM methods. Fig. 4.10 sketches these two HMM-based methods for dealing with *Tracking*.

4.5.4 Latency Reduction

As aforementioned, another challenge we need to deal with in tracking is the latency, which mainly results from the delay of RSSI collection and signals sending by passive tags [141]. As a result, we introduce a *forward calibration* mechanism to re-calibrate the walking trajectory outputted by the Viterbi searching to reduce the latency. In particular, we adopt a sliding window to average the latest several locations as follows:

$$\hat{c}'_t = \frac{\sum_{i=t}^{t+|w|-1} \hat{c}_i}{|w|} \tag{4.19}$$

where $\hat{c}'_t$ represents the calibrated coordinates of location l_t is the at time t, $|w|$ denotes the length of the sliding window, and $\hat{c}_i$ is raw coordinates of estimated grid's center at time i using Eqn. 4.17.

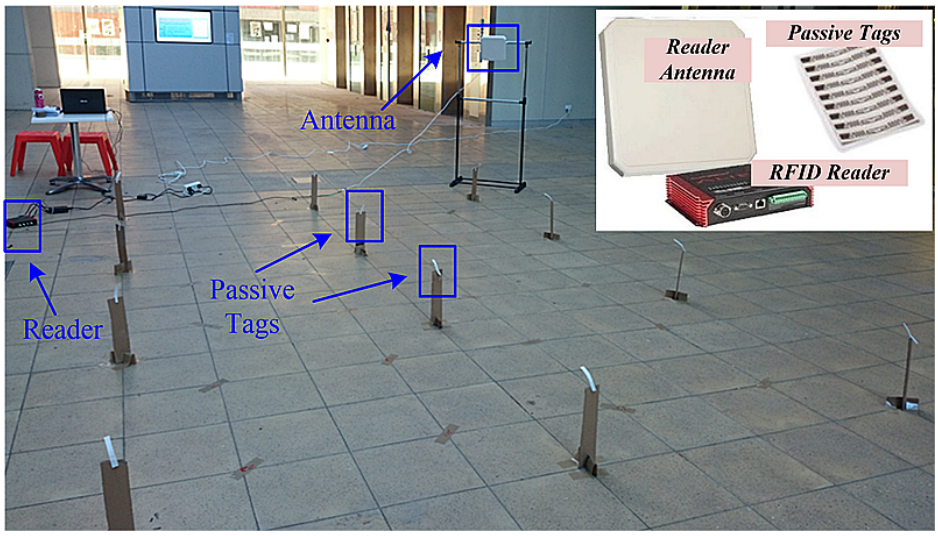


Fig. 4.11 Hardware deployment

4.6 Evaluation

We evaluate our approach through *i)* micro experiments in a $3.2m \times 4.8m$ testing area (stacked by 6 RSS fields); and *ii)* field experiments in a fully furnished house including two bedrooms and a kitchen (around $220m^2$ gross floor area).

4.6.1 Hardware Deployment

Ultra-low cost of UHF tags (5∼10 cents each) become the preferred choice of many industry applications. Following the common practices, we adopt passive UHF tags in this chapter. As Fig. 4.11 shows, our system is built upon commercial off-the-shelf RFID products without any hardware or firmware modification. In particular, we use an Alien ALR-9900+ RFID reader, several reader-antennas (Model: ; Size: $20cm \times 20cm \times 3cm$) and dozens of UHF passive tags (Model: squiggle Higgs-4; Size: $1cm \times 10cm$). The operation frequency of the reader is 840 to 960MHz and the sampling rate is 2Hz. Each collected RSSI readings includes a TAG-ID, RSSI and TIME. Our system runs in a laptop computer (CPU: I7-3537U 2.5GHz; RAM: 8G; OS: Win7).

4.6.2 Evaluation Metrics

Similar to other localization and tracking systems, we adopt the following two evaluation metrics, *Accuracy* and *Error Distance*, to measure the localization accuracy and tracking error respectively.

$$Acc. = \frac{\sum_i^N \mathbb{I}(\hat{l}_i, l_i)}{N} \tag{4.20}$$

where $\hat{l}_i$ and l_i respectively denote the estimated and actual location of a user, the indicator function $\mathbb{I}(a,b)$ equals to 1 if $a = b$, otherwise 0, and N denotes the tested RSSI numbers. The tracking error distance is defined by

$$D_{error} = \frac{\sum_i^{|T|} dis(\hat{c}_i, c_i)}{|T|} \tag{4.21}$$

The error distance depicted above actually measures the averaging accumulated error distance for each moving trajectory. In particular, c_i and $\hat{c}_i$ mean the actual and predicted coordinates of a subject at time i, and $dis(\hat{c}_i, c_i)$ denotes the Euclidean distance between them. $|T|$ is the number of all observed RSSIs of a moving trajectory.

4.6.3 Micro Experiments

We first conduct several micro experiments to test our methods. Before evaluating our approaches, we need to decide how to choose the optimal size for each virtual grid. According to our experiments, a small grid size brings more indistinguishable patterns due to the RSSIs' overlapping in adjacent locations, as a result, we need more profiling data to resolve such overlapping. In this chapter, a very high location resolution is not our primary goal. For example, caregivers normally more concern about the elderly resident locating on which area or room of a house or apartment instead of an extremely fine-grained location point. Based on this intuition, we set up our experiments as Fig. 4.12, in which each virtual grid is $0.8m \times 0.8m$, locating people in a $0.64m^2$ resolution.

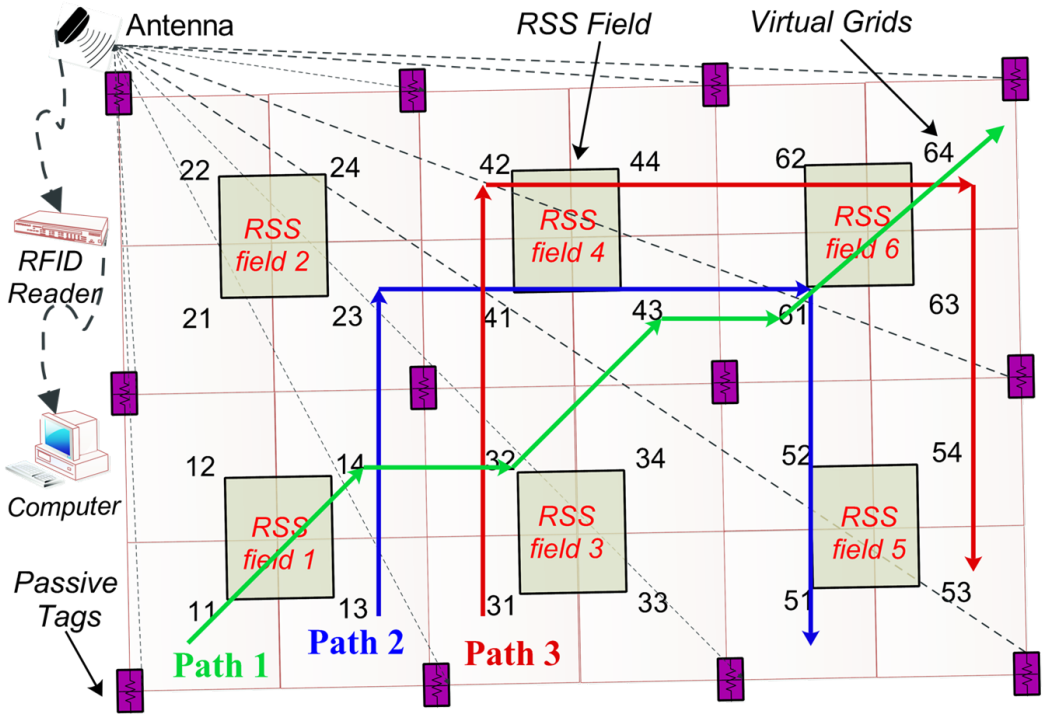


Fig. 4.12 Multiple RSS fields and testing paths

Experimental Settings

As Fig. 4.11 shows, one reader-antenna is placed at 1.55m height and faces to passive tags from 25° $\sim$ 75° angle[7]. The tags are attached on paperboard-holders placed 30cm above the ground. Considering that our model aims to learn the RSSI-Location mapping, those passive tags can be flexibly put as any geometric shape. For simplicity, we deploy the passive tags as a square array with around 1.6m distance. Another issue is that, the reader may lose some RSSI readings due to the human body occlusion during localization or tracking. As a result, to make the received RSSI vector with same number of readings, we fill in those missing values as 0 in each sampling time.

[7]The antenna angles or height can be set up arbitrarily as long as it is able to capture all the readings of all tags in an empty environment.

Table 4.1 Localization accuracies of different methods by using different ratios of training data

Scena.	Ratio (%)	10	20	30	40	50	60	70	80
1	*k*NN	**0.946**	**0.954**	**0.958**	**0.958**	**0.960**	**0.961**	**0.962**	**0.963**
	GMM	0.927	0.935	0.938	0.940	0.939	0.943	0.940	0.941
	SVM	0.707	0.756	0.823	0.851	0.897	0.912	0.919	0.928
	ELM	0.664	0.764	0.719	0.771	0.881	0.898	0.904	0.904
	NaiveBayes	0.883	0.887	0.913	0.930	0.938	0.944	0.943	0.946
2	*k*NN	**0.810**	**0.823**	**0.833**	**0.844**	**0.869**	**0.902**	**0.913**	**0.931**
	GMM	0.751	0.777	0.783	0.793	0.838	0.884	0.894	0.902
	SVM	0.656	0.717	0.775	0.797	0.819	0.832	0.846	0.857
	ELM	0.680	0.538	0.614	0.701	0.677	0.774	0.819	0.835
	NaiveBayes	0.741	0.777	0.793	0.844	0.872	0.890	0.903	0.914
3	*k*NN	**0.880**	**0.904**	**0.918**	**0.927**	**0.931**	**0.931**	**0.936**	**0.943**
	GMM	0.851	0.877	0.883	0.893	0.898	0.904	0.904	0.912
	SVM	0.715	0.746	0.774	0.826	0.840	0.854	0.876	0.881
	ELM	0.688	0.583	0.617	0.693	0.705	0.812	0.840	0.846
	NaiveBayes	0.768	0.789	0.855	0.889	0.918	0.921	0.928	0.929

Localization

To test the localization capability, we define three scenarios to simulate the possible real-world daily routines.

Scenario 1 (Stationary) *A person stands or sits statically in a certain location of monitored area, mimicking that a resident may talk with someone or watch TV.*

Scenario 2 (Dynamic) *A person moves around and does several activities within a certain small zone, mimicking a resident may cook in the kitchen or do morning exercise.*

Scenario 3 (Mixed) *A subject performs both activities defined in Scenario 1 and 2 within a certain location.*

Accordingly, we test our system based on the above three scenarios: *i)* a participant appears in each location for 120*s*; *ii)* a participant walks around and performs some activities in each grid for 120*s*; and *iii)* a participant does the above activities for 240*s* per grid.

Overall we collect 276,480 RSSI readings in the localization experiments. We randomly split it into testing and training datasets based on different ratios (in each ratio, we run the methods twenty times to calculate the average localization accuracy). Table 1 compares our experimental results of five localization methods with different training ratios. We carefully tune the parameters for each method - we set $k = 2$ for kNN and GMM component number as 4, and choose termination criterion and C in SVM with a linear kernal as 0.01 and 1 respectively [143]. For a stationary scenario, all five methods can localize the subject with a decent accuracy. Among all, kNN classifier achieves a 94.6% localization accuracy in particular with $12s/grid$ training data, which significantly outperforms other methods especially the SVM and ELM. For a challenging dynamic localization scenario, kNN still achieves a better performance with 93.1% accuracy using 80% training data. It is also noted that, under a *dynamic* scenario, the localization accuracy is more relevant to the training data size. A larger training dataset is able to provide more informative RSSI patterns for this case. In Scenario 3, our system is able to reach a high accuracy of 94.3%. In summary, under a circumstance of limited training data (*e.g.,* 10% training data), kNN based localization reveals a better and robust performance. It is worth to mention that, to achieve a similar accuracy, the shortest collection time of training data is of *minutes-level* in past localization systems [45]. On the contrary, our system only requires a *seconds-level* collection time to get a comparable localization performance. We also observe that, with more training data (*e.g.,* 80% training data in Table 1), other methods are also able to get good accuracy but more sensitive to the training data size.

Tracking

In the tracking experiments, we evaluate our HMM based models on three moving trajectories under the proposed two transition strategies, illustrated in Fig. 4.12. Two persons with various

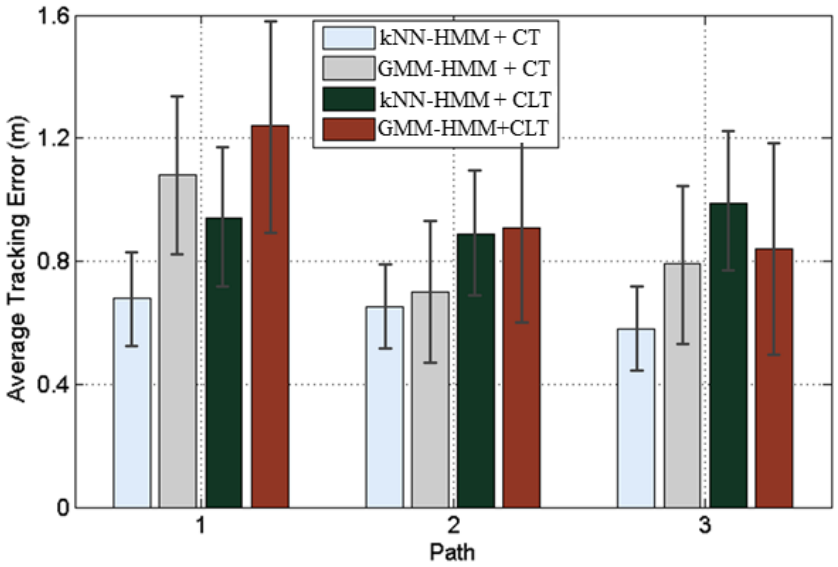


Fig. 4.13 Tracking errors on three paths (*CT: Constraint Transition; CLT: Constraint-Less Transition*)

weights and heights participate our experiments and every path is tested for 20 times[8]. As Fig. 4.13 illustrates, *k*NN-HMM with Constraint Transition (*i.e.,* kNN-HMM + CT) is able to track a subject with $0.64m$ mean error, achieving the best result among all the methods. This may lie in the fact that *k*NN-HMM + CT feasibly narrows down the candidate locations (excluding the invalid location candidates), thus can better quantify the mapping relation from RSSI sequence to moving trajectories. We also compare our system with other popular RFID-based localization works, as shown in Fig. 4.14.

LANDMARC [22] is the very first RFID-based localization system that tracks a tagged subject by measuring its weighted average locations of its nearest four tags. It needs the target attached with tags and know the reference tags' locations. In our experimental testbed, it achieves average tracking error $1.64m$ (*i.e.,* LANDMARC-1: 3×4 reference tags with $1.6m$ interval), and $1.11m$ (*i.e.,* LANDMARC-2: 5×7 reference tags with $0.8m$ interval).

TagArray [32] is one of the first RFID-based systems that can localize a subject in a device-free manner. Generally, TagArray detects a person by comparing the variation of

[8]We mainly focus on tracking a resident with a moderate moving speed ($\approx 0.6m/s$) due to that fast moving in an indoor environment is not practical.

RSSI readings with a pre-learned threshold. However it is built upon active RFID tags and requires a high tag tensity as a tag array. It reaches 1.7m (*i.e.,* TagArracy-1: 3×4 reference tags with 1.6m interval) and 1.15m (*i.e.,* TagArray-2: 5×7 reference tags with 0.8m interval) mean tracking error in our testbed.

TASA [33] is another device-free RFID-based localization system, which adopts both passive and active tags. Thus it is less costly than TagArray. But still, it requires to calibrate all tags' coordinates. It gives 1.53m (*i.e.,* TASA-1: 3×4 reference tags with 1.6m interval) and 1.05m (*i.e.,* TASA-2: 5×7 reference tags with 0.8m interval) mean tracking error.

SCPL [37] is one of the advanced wireless-based device-free localization systems. It utilizes a Gaussian model based Conditional Random Field (GM-CRF) method to track a moving person. It is very similar to our GMM-HMM method (utilizing Gaussian Mixture Model). We implement the GM-CRF method in our RFID dataset and get a mean 0.98m tracking error.

Twins [39] is a very recent RFID-based system purely built upon passive tags, which utilizes a interference phenomenon (called state jumping) of two passive RFID tags to do the motion detection. It gives a mean 0.75m tracking distance error in an open warehouse. Twins also needs to carefully calibrate the positions of the reference tags.

BackPros [135] is one of the recent RFID-based positioning systems, which is able to track a passive tag with a decimeter-level accuracy. However, BackPro aims to track an object instead of tracking a human body by exploring the phase differences of backscatter signals to infer the tag's location. It needs to carefully calibrate the positions of four antennas beforehand and the tracked object need to be attached with a passive tag.

Different to the baseline methods, our system does not need to calibrate the tags' locations[9] and achieves 0.64m average tracking error in our testbed. It offers about 2.56×, 2.66×, 2.39× and 1.53× improvement compared with LANDMARC [22], TagArray [32],

[9]Although we put tags in a square array in Fig. 4.12, we actually do not use any coordinates of the tags. Because we target to learn the mapping model, the tags can be placed in other geometric locations.

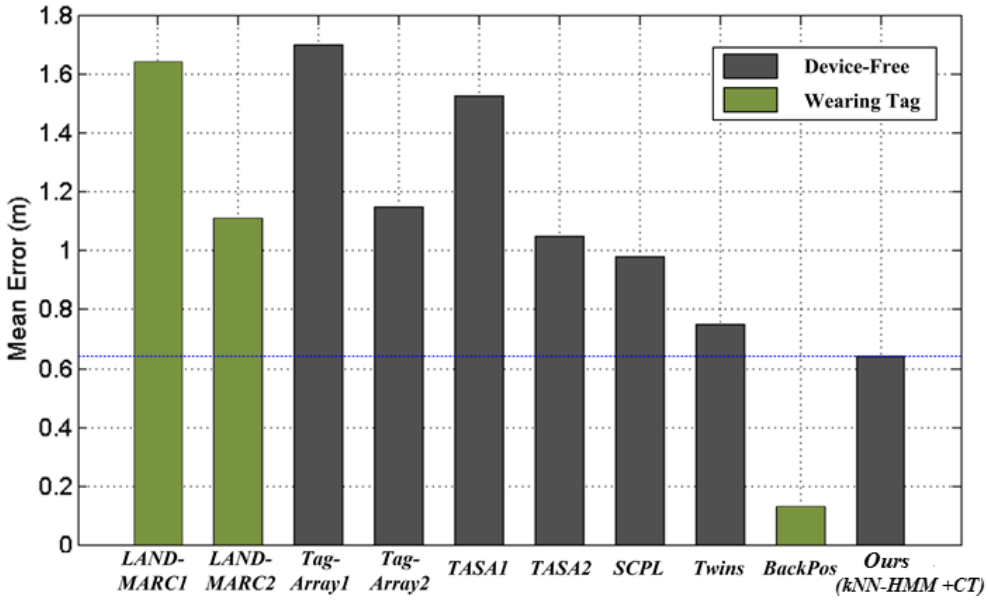


Fig. 4.14 Average tracking errors

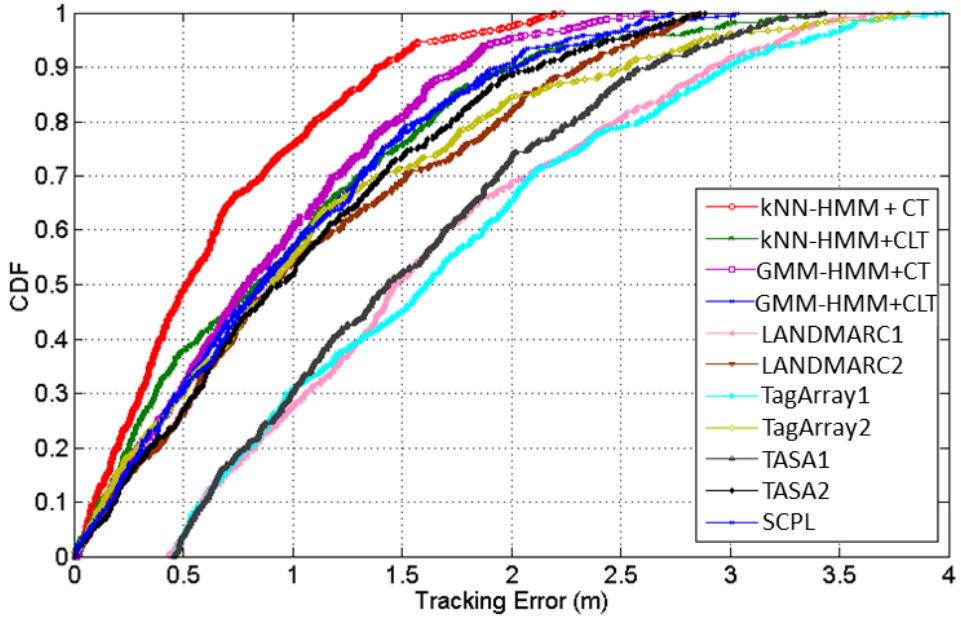


Fig. 4.15 Tracking error CDF

TASA [33], SCPL [37] (see Fig. 4.14) using the same number of tags. Fig. 4.15 shows the CDF (cumulative distribution function) curves of tracking error for different methods. The *k*NN based HMM with CT achieve a better result, nearly 76% tracking errors are below $1m$.

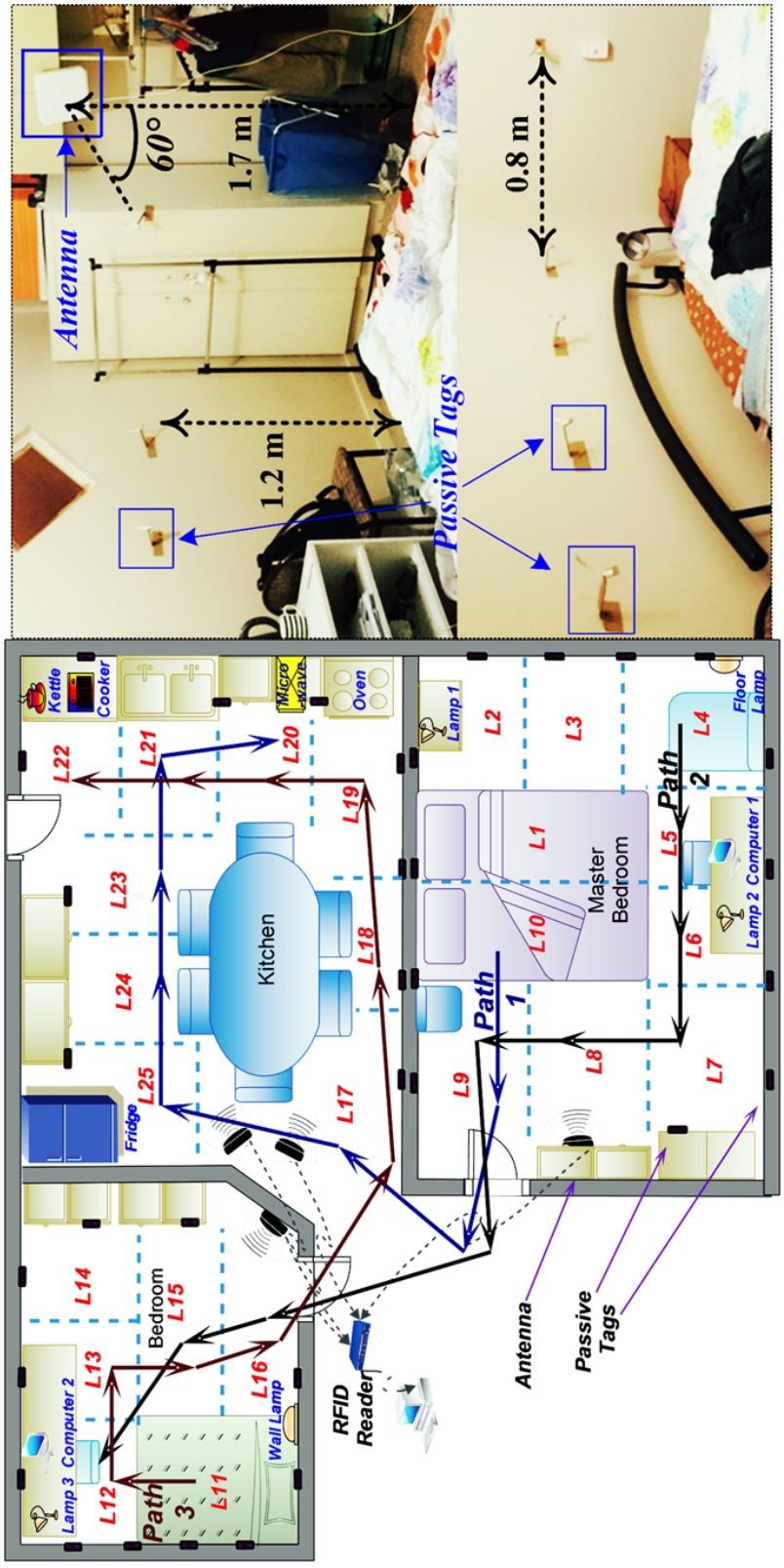


Fig. 4.16 House layout and tracking paths

4.6.4 Field Experiments

This section delivers the experimental results in a residential house that contains 2 bedrooms (*i.e.,*a home office and a master room) and a kitchen, as shown in Fig. 4.16. The reader-antennas are deployed around 1.7 *meters* vertical distance to the ground and the facing angle to the passive tags is around 60°, which is capable of capturing all RSSI readings under a non-resident environment. Overall we virtually divide the monitored area into 25 grids, and use 34 passive RFID tags and one reader with three antennas. We attach those passive tags on the room-walls with about $0.8m$ interval.

Localization

Similarly, we design three localization scenarios in our field experiments - *Stationary*, *Dynamic* and *Mixed*. Accordingly, three types of data are collected to train and test the location classifiers[10].

Figure 4.17∼4.19 show the results of localizing a subject using five different location classifiers varying training ratios (from 5% to 90%)[11]. In the *stationary scenario*, the localization accuracy of *k*NN is as high as 93.8% with 90% training ratio. More importantly, only with 6 seconds training data (5% training ratio) for each grid, it can achieve an accuracy over 85% in a residential house, revealing its advantage than other location classifiers. For Scenario 2, the performances of all methods are degenerated due to the unstable human inference, and the results among different methods are more close to each other. We also observe that more training data can significantly enhance the localization accuracy, which means, for the challenging *dynamic scenario*, collecting more training data can more accurately capture the human inference to RSSI signals. For Scenario 3, the best

[10] *i)* a person appears in each grid for 120*s*, *ii)* a person contentiously moves round in a grid for 120*s*, and *iii)* a participant does the above stationary and dynamic activities respectively for 120*s*. For L1, L10, L11, we only collect the data people lying down for all scenarios. Overall, we collect 848,640 RSSI readings, forming 24,960 RSSI vectors.

[11] We randomly choose the training dataset and testing dataset, and conduct each experiments 20 times, reporting the average accuracies

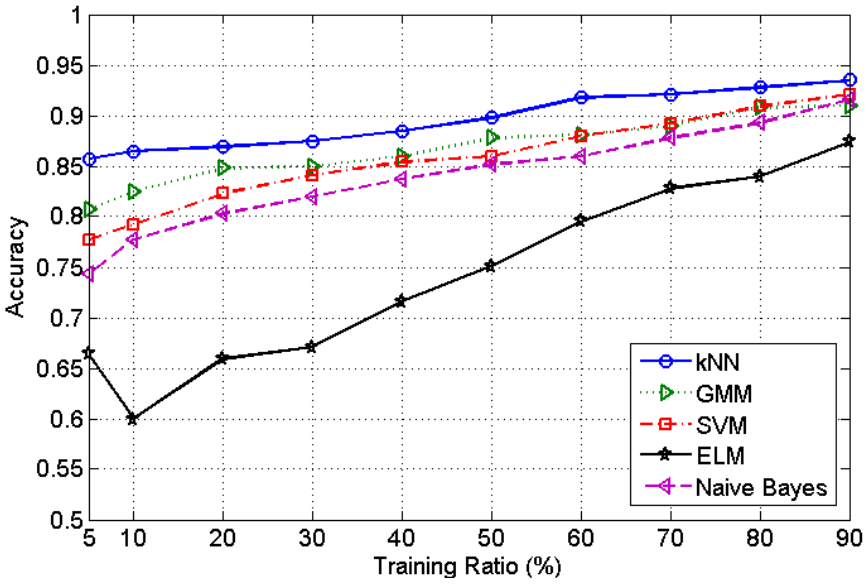


Fig. 4.17 Localization accuracy in Senario 1

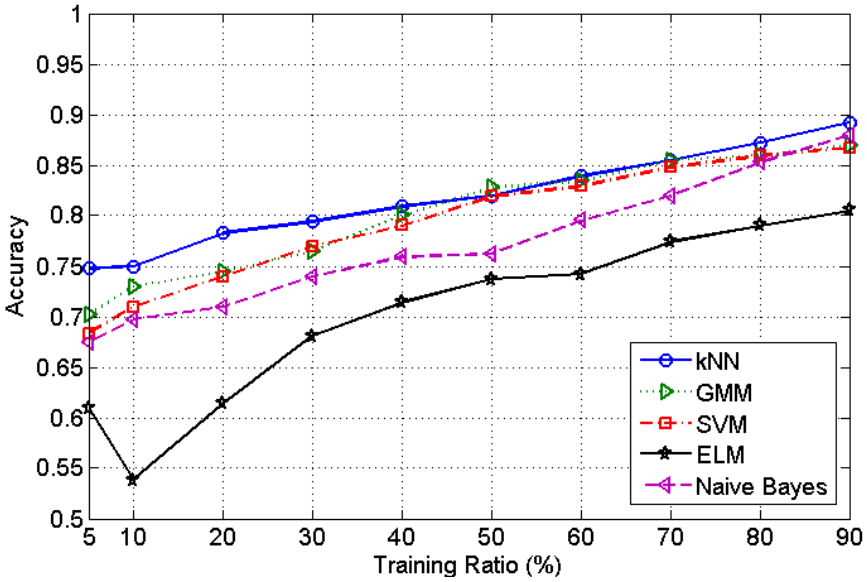


Fig. 4.18 Localization accuracy in Senario 2

perfermance is achieved by *k*NN using 90% training data, and the overall performance is between *stationery scenario* and *dynamic scenario*. In summary, *k*NN shows its superiority in RFID-based device-free localization, considering its simplicity, light computation overhead and relaxing requirement of training data.

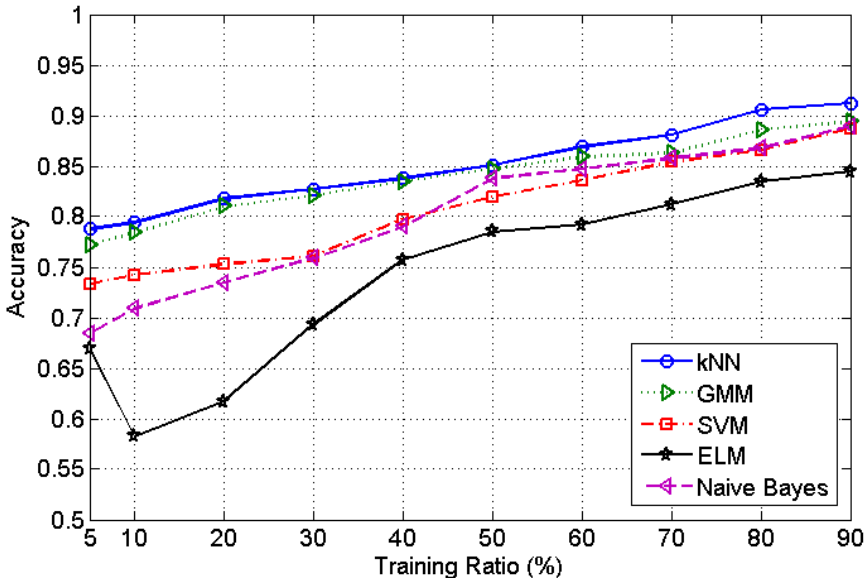


Fig. 4.19 Localization accuracy in Senario 3

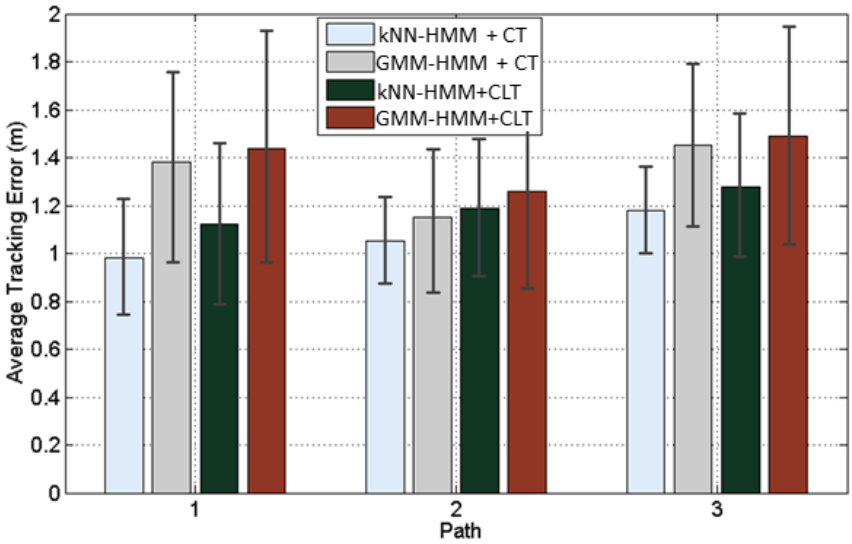


Fig. 4.20 Tracking errors on three paths

Tracking

We also test our tracking methods on three daily routines, as shown in Fig. 4.16.

Path 1: $L10 \rightarrow L9 \rightarrow L17 \rightarrow L25 \rightarrow L24 \rightarrow L23 \rightarrow L21 \rightarrow L20$ represents that, a resident gets up from the master room and does some cooking in the kitchen.

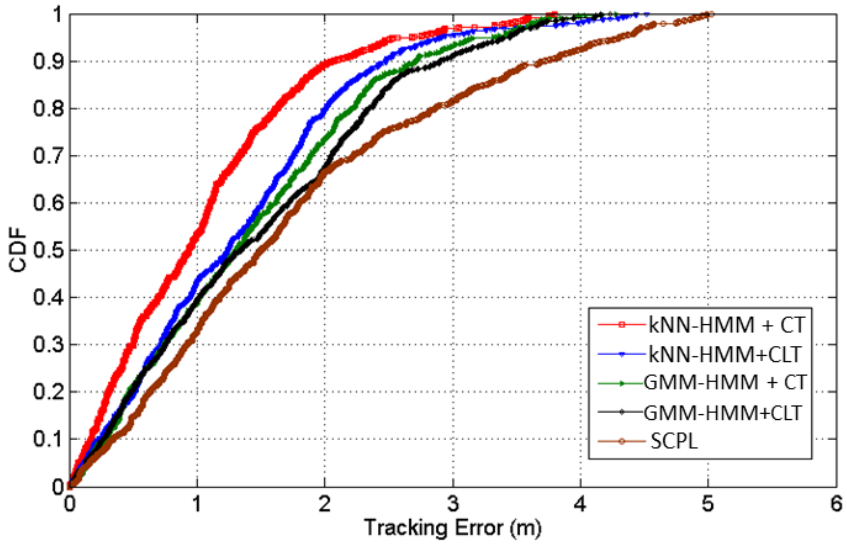


Fig. 4.21 Tracking error CDF

Path 2: $L4 \rightarrow L5 \rightarrow L6 \rightarrow L7 \rightarrow L8 \rightarrow L9 \rightarrow L16 \rightarrow L15 \rightarrow L12$ mimics that a resident gets up from the sofa $L4$ of the master room, and then goes to work at the desk $L12$ of the home office (*i.e.,* the room in the upper-left of Fig. 4.16).

Path 3: $L11 \rightarrow L12 \rightarrow L15 \rightarrow L16 \rightarrow L17 \rightarrow L18 \rightarrow L19 \rightarrow L20 \rightarrow L21 \rightarrow L22$ indicates that, a resident gets up from the bedroom and goes to the kitchen using the kettle.

Overall three subjects join the experiments and each path test is repeated 20 times. As Fig. 4.20 depicts, our proposed kNN-HMM with Constraint Transition illustrates a better result (with 1.07m mean tracking error) comparing to other HMM based models. It is noted that, in Path 3 - a more complex path of daily routine, our method obtains a larger tracking error (nearly 1.2m). The reason may be due to the fact that Path 3 involves walking through a narrow hall with many electronic appliances in the kitchen, which block or absorb the energy of backscattered signal from an antenna. Thereby the tracking accuracy decays for this application scenario. In general, our proposed method outperforms other methods by intensively learning the mapping relation between RSSI readings and human mobilities under a transition constraint. It is noted that SCPL achieves 1.66m mean error, 1.55 times large than our method. In the field experiment, we only compare our system with the proposed

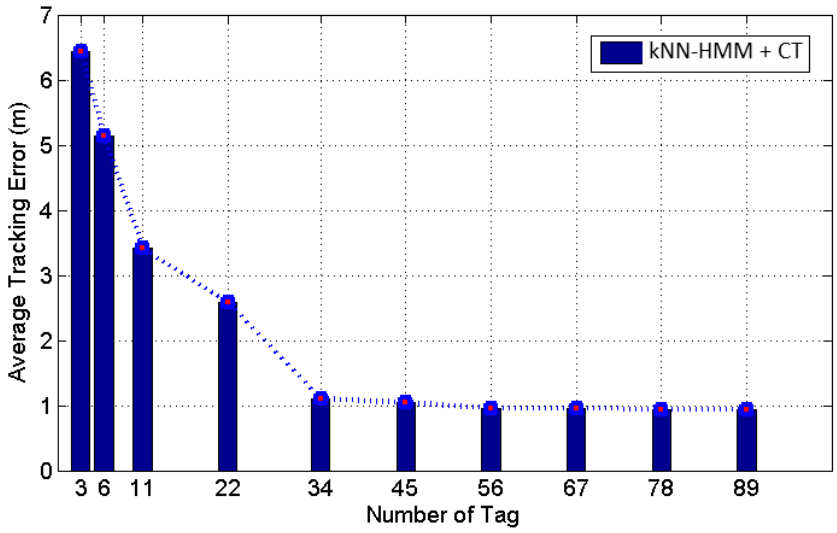


Fig. 4.22 Tracking errors with tag numbers

method in SCPL since the LANDMARC, TagArray and TASA place the RFID tags as arrays on the ground. Such deployments are impractical and obtrusive for a residential environment. Firstly, the reader even cannot catch the readings from passive tags that are deployed in a carpet ground since signals are blocked by furnitures around and absorbed by the carpet. Secondly, tag-arrays that densely deployed on ground in a residential environment strongly obstructs the mobility of the resident, causing uncomfortable and inconvenient. In our system, the passive tags are attached on the wall which is more practical and considered as less intrusion. As a result, the localization systems proposed in LANDMARC, TagArray and TASA are no longer capable for the residential application scenario.

4.6.5 Parameters Selection

In this section, we will discuss the factors that have impact on the tracking accuracy.

Tag Density

Tag density is an important influential factor to the tracking performance. As Fig. 4.22 shows, we investigate the impact of tag density by deploying different numbers of tags in the testing

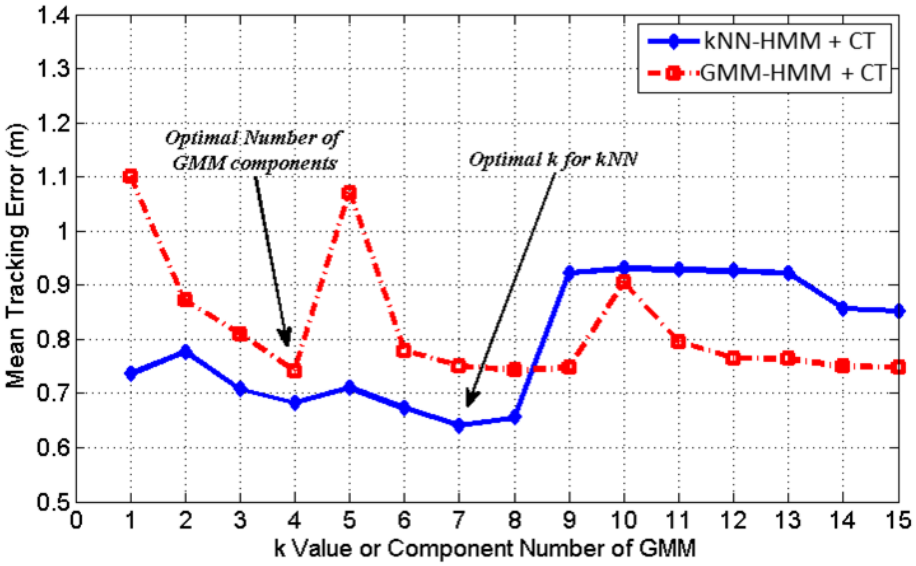


Fig. 4.23 k value and GMM component number

rooms. The experiments reveal that a sparse tag density (*e.g.*,2 tags/room) will reduce the tracking performance. On the other side, continuously using more passive tags does not improve the tracking accuracy significantly. For example, in our experiments, the tracking error does not decrease obviously when increasing the tag number from 34 to 89. Such a phenomenon lies in a fact that it is difficult for an antenna to probe a large number of passive tags and thus resulting in severe reading loss. It is noted that, comparing to TagArray and TASA that require a high density of tags, our system is able to achieve a comparable tracking accuracy using less passive tags.

k Value and GMM Component Number

There are two key parameters in our HMM-based models, one is k value in Emission Matrix of kNN-HMM, another one is the *component number* (CN) of GMM in GMM-HMM. We investigate these two parameters in our micro experiment testbed. Fig. 4.23 illustrates that, the tracking error reaches the lowest when $k = 7$, which thus is chosen as the optimal value in our tracking system. However, GMM-HMM achieves a better tracking accuracy at CN =

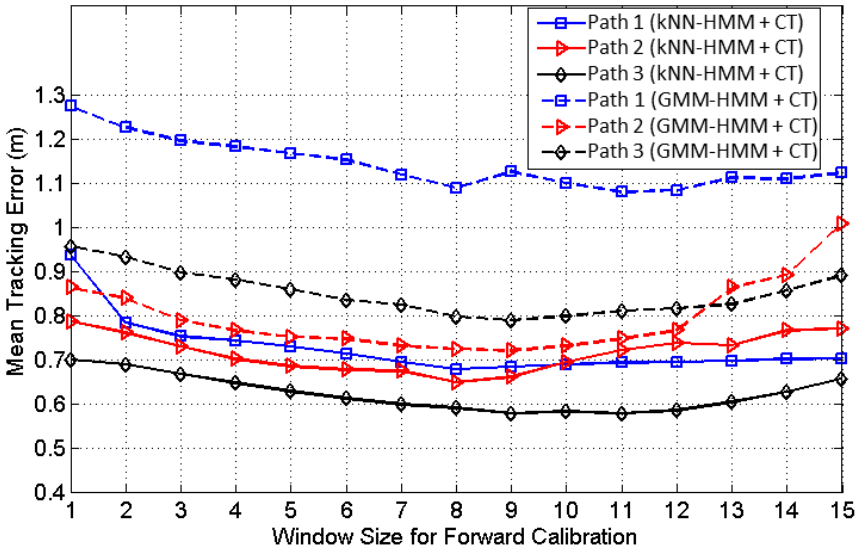


Fig. 4.24 Window size in forward calibration

4, 8, 9 and 15. Considering that a larger CN may potentially cause a model over-fitting and requires more computation overhead, we choose CN = 4 in this chapter.

Window Length

For a localization system, dealing with the latency is also a concerning issue [144, 43]. In this chapter, we introduce a simple yet efficient forward calibration to reduce the latency, laying on the fact that previous human motion has an impact on current location prediction. One of key parts is to decide the length of *previous motion*, *i.e.,* the smoothing window length. Fig. 4.24 shows the relevance between the window size of forward calibration and the tracking error in different paths using two HMM based methods. We observe that, when the window length ranges from 8 to 11, our system achieves a less tracking error. Thus, we select 8 as the optimal length in our system considering both the computational burden and accuracy.

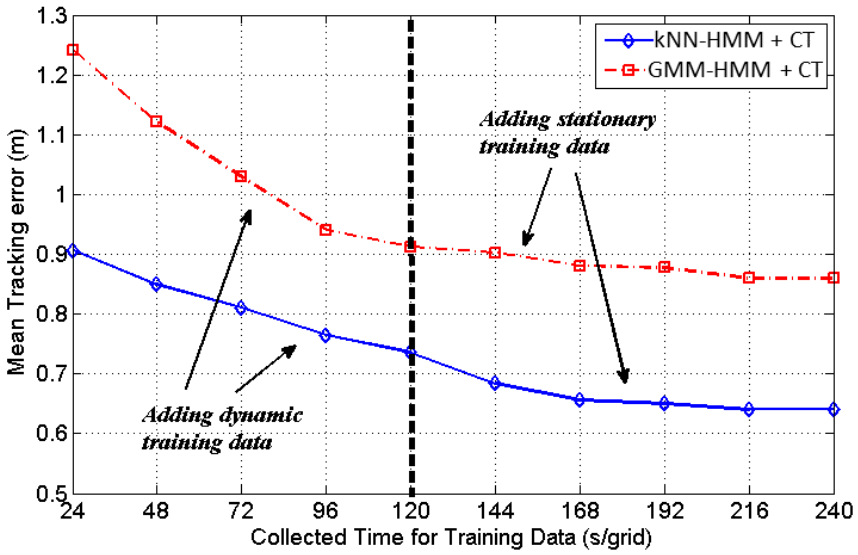


Fig. 4.25 Stationary data *vs* dynamic data

Stationary Data *vs* **Dynamic Data**

As mentioned before, we put two kinds of training data into the HMM based methods - stationary data (Scenario 1) and dynamic data (Scenario 2). In order to analyze which type of training data plays a key role in tracking, we first add 120 seconds *dynamic* training data (before black dot line, *the First-stage Training*), then we add another 120*s* stationary data for training (after black dot line, *the Second-stage Training*), shown as Fig. 4.25. Overall, we observe that the tracking error decreases as adding more training data. In details, the error diminishes rapidly in the first stage, but just slightly reduces in the second stage. Actually, the last 72 seconds stationary data does not make much contribution to improving the performance. It reveals that more dynamic data substantially provide richer anchoring RSSI information regarding the human motion, and a few stationary training data (*e.g.,* collecting 24 seconds training data) nearly provide all the essential statical information for tracking. In other words, we can add more dynamic training data to improve the system's tracking performance.

4.7 Conclusion

Indoor localization and tracking systems built upon passive RFID hardware have shown attractive potential of passive tags due to the cheap price, low-maintenance and battery-free character. Those promising features strongly motivate this chapter, in which we design, implement and evaluate an RFID-based DfP indoor localization and tracking system built upon passive tags. By taking advantage of supervised classification methods, we introduce a series of data-driven models to quantify the RSSI distributions when a user appears at various locations within a monitored area. These approaches enable our system to localize a subject by maximizing the posteriori probability given RSSI observations. To transfer the pattern learned in localization into tracking, we further propose the multivariate GMM-based HMM and *k*NN-based HMM methods, in which we utilize the probabilistic estimation learned in localization to construct the emission matrix and introduce two human mobility strategies to approximate the transmission matrix under the hidden Markov assumption. The intensive experimental results verify the effectiveness and accuracy of our system.

However, our system aims to accurately localizing and tracking resident in a clustered, fully-furnitured environment, in which the proposed method can only achieve around 1 meter mean tracking error by purely adopting passive RFID tags. In the next chapter, we will show how to incorporate the human-object interaction events to further enhance the performances of indoor localization and tracking in a residential house.

Chapter 5

Enhancing RFID-based Device-free Indoor Localization and Tracking through Human-Object Interactions

Device-free indoor localization aims to localize people without requiring them to carry any devices or being actively involved in the localizing process. It underpins a wide range of applications including older people surveillance, intruder detection and indoor navigation. However, in a cluttered environment such as a residential home, the Received Signal Strength Indicator (RSSI) is heavily obstructed by furniture or metallic appliances, thus reducing the localization accuracy. This environment is important to observe as human-object interaction (HOI) events, detected by pervasive sensors, can potentially reveal people's interleaved locations during daily living activities, such as watching TV, opening the fridge door. This chapter aims to enhance the performance of commercial off-the-shelf (COTS) RFID-based localization system by leveraging HOI contexts in a furnished home. In particular, we propose a general Bayesian probabilistic framework to integrate both RSSI signals and HOI events to infer the most likely location and trajectory. Unlike other RFID-based localization systems, which are limited to deployment and testing in clear/semi-clear spacial areas, our

system can work in a furnished environment. Experiments conducted in a residential house demonstrate the effectiveness of our proposed method, in which we can localize a resident with average 95% accuracy and track a moving subject with 0.58m mean error distance.

5.1 Introduction

Ambient intelligence has been drawing a growing attention as it enables a smart environment that can respond to people's locations and behaviors using various wireless signals, sensors, or Radio-Frequency Identification (RFID). Many attractive applications can be realized in these smart environments that will have huge impact on our daily lives, such as aged care, surveillance and indoor navigation. A crucial prerequisite of all these applications is to accurately localize and track people in a cluttered living environment [26, 32, 145]. To tackle this challenge, many state-of-the-art indoor localization systems have been developed over last decades such as LANDMARC [22], WILL [134], Tagoram [23]. Most of these techniques, however, require the target to either carry sensors/smartphones/tags or be actively involved in the localizing process, which has several limitations in practice. The attached sensor/smart phone/tag may be lost or damaged or elderly people with dementia may forget to carry the device. As a result, *device-free* (or *unobstructive*) indoor localization has gained significant momentum recently and several approaches have been proposed [43, 44, 47].

One popular device-free technique category is based on computer vision, such as using RGB camera [146], depth camera [31], or infrared sensors [147], however they are usually regarded as being privacy-invasive and causes uncomfortable feeling to the residents. Vision-based systems also require the tracked target within the LOS (line of sight) of cameras (*i.e.,* no barriers or blocks between the subject and camera), and fail to work in dimmed or dark environments. Another technique category is based on RF (radio frequency) signals, *e.g.,* detecting human locations by measuring RSS (received signal strength) or CSI (channel state information) in WLANs (wireless local area networks) [43, 44, 137], or tracking a target

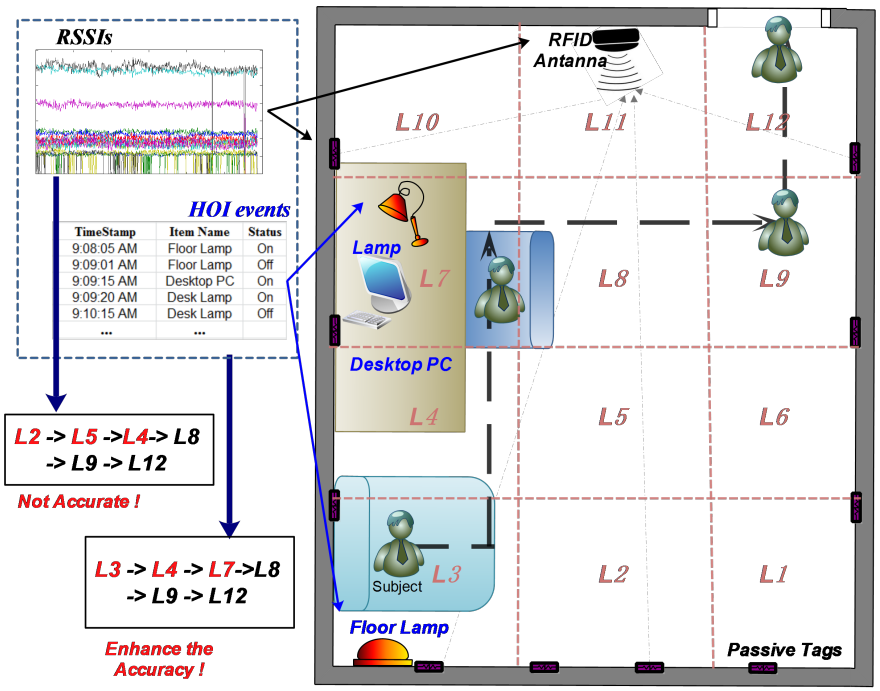


Fig. 5.1 Intuition of *HOI-Loc*

through a wall based on the RF signal reflected from human body [41]. However, most of such past systems often require regular maintenance (*e.g.,* replace the batteries regularly), or need specialized WIFI signals, *e.g.,* Frequency-Modulated Continuous-Wave (FMCW), or special-purpose devices, *e.g.,* Universal Software Radio Peripheral (USRP), which hinder their wide application and deployment in reality [148, 138, 47].

Thus, device-free localization based on passive RFID tags has attracted much attention recently due to its low-cost (5∼10 cents each, still dropping quickly) and maintenance-free (no need batteries) nature [148]. However, existing RFID-based device-free techniques usually work in clear or semi-clear spaces (*i.e.,* empty spaces or spaces with very few objects), and none of them are actually tested in clustered residential environments, especially in a multiple-room scenario. In addition, most RFID-based localization techniques are based on the assumption that knowing the tags' coordinates in advance, which is impractical in real-world applications (accurately locating the tag's position is a time-consuming and chal-

lenging task itself). Besides, many state-of-the-art RFID-based systems (*e.g.,* [148, 47, 23]) heavily rely on ideal propagation models of RF phase or RSSI, which may not be feasible in a full-furnitured residential room where rich multi-path reflections and frequent electromagnet interference exist (*e.g.,* turning on/off electronic appliances in a kitchen) [149, 33]. In addition, the residents usually move around in the space, and it creates fast changing communication environments such as loss of RSSI readings and varied backscatter prorogation paths, further introducing unpredictable disturbing factors. To tackle these challenges, in this chapter, we design *HOI-Loc*, an RFID-based device-free localization and tracking system to achieve *high accuracy* in *clustered* living environments using *Human-Object Interactions*.

With the booming of IoT (Internet of Things), human-object interaction has been advocated as an essential component of Cyber Physical System (*e.g.,* smart homes, intelligent space, and home automation). According to the report [150], there are more than 1.9 billion devices launched into the market each week that can connect to the residential home, and there will be rapidly increased into 9 billion by 2018, roughly equal to the number of smart phones, smart TVs, tablets, wearable computers, and PCs combined. With such tremendous smart devices, we can easily access, retrieve and monitor HOI events in our daily lives [151]. For instance, a smart home equipped with various sensors (see Fig. 5.1) is capable of reporting the operating conditions of the floor lamp, desktop computer and desk light [152]. Moreover, we observe that the locations causing severe signal decay are usually full of furniture or electrical appliances, and such locations are exactly where HOI frequently occurs. Whereas, from another perspective, we can substantially improve the localization accuracy by utilizing such interaction events. For example, localizing a person in the kitchen (equipped with rich electrical appliances) purely based on RSSI is difficult since the signals are severely interfered by electrical devices made of metals (*i.e.,* microwave oven, fridge or cooker). However, we can offset such signal disturbance and improve accuracy by using HOI, such as opening a *fridge*, turning on a *kettle* or a *microwave oven*. Inspired by this intuition,

we propose to incorporate HOI into existing RSSI based methods to improve localization accuracy in clustered indoor spaces.

Transforming the use of HOI into a practical system, however, requires addressing a number of challenges. First, localization from weak RSSI signals of passive RFID tags in a clustered environment is difficult. Unlike active RFID tags or wireless sensors that have their own power supplies, passive tags can only obtain energy from the interrogating field, which can easily be obstructed by furniture and metal appliances (*e.g.,* RSSI reading loss, RSSI jumps due to on-and-off of electronic appliances). In particular, this task is typically accomplished using COTS RFID readers, which currently do not support any low-level signal access or modification. In addition, HOI contexts are discrete events which occur from time to time, but RSSI readings are continuous signal (can be sampled as high as 10 times per second). How to feasibly incorporate the discrete HOI events with continuous RSSI signal under rigid mathematical reasoning is a challenging task. Moreover, the inherent signal diversity of passive tags caused by human mobility would introduce many unknown effects on the RSSI attenuation and reading disturbance, leading to unpredictable tracking errors.

To address these issues, in the *HOI-Loc* system, we first set up several RSS fields formed by passive RFID tags attached on the bedroom's walls[1] to continuously generate RSSI signals, and then deploy various kinds of sensors (*e.g.,* infrared sensor, touch sensor and light sensor etc.) to detect the resident's interaction events with electrical appliances. We propose three main techniques to tackle the aforementioned challenges. First, we propose a Probabilistic Polyhedron Isolation (PPI) method to model the likelihood of the target's locations by measuring the Euclidean distance of testing RSSI readings with each isolated high-dimension *polyhedron*, which is robust to the signal attenuation and jumping (see Section 5.4). Second, we develop a rigid Bayesian probabilistic framework to fuse the discrete HOI events (*i.e.,* indicating where and when people interact with objects) with continuous RSSI signals. In

[1]Unlike other device-free RFID systems (*e.g.,* LANDMARC [22], TagArray [32], TASA [33] and Tadar [47]), we do not need to know the locations of passive tags, meaning tags can be attached on the wall in an arbitrary shape

particular, we first estimate the *RSSI probability*, then update the likelihood by computing the *HOI probability*, and finally optimize a location with largest confidence (see Section 5.4). To track a moving subject, we introduce a Hidden Markov Model (HMM) to quantify the continuous location transition process to eliminate the negative impact caused by human mobility. In particular, we first approximate the *Emission Matrix* by a probabilistic scheme that considers both evidence of the RSSI sequence and HOI event stream based on *Bayesian Inference*, then propose a practical but efficient strategy to estimate the *Transition Matrix*, finally use the *Viterbi Search* to recover the target's trajectory (see Section 5.5). In a nutshell, our main contributions are summarized as follows.

- We introduce an approach that utilizes HOI events to facilitate device-free localization based on passive COTS RFID tags. Our experiments demonstrate the feasibility and accuracy of *HOI-Loc* in a furnished, clustered living environment. To the best of our knowledge, the proposed system is a very first effort to do so.

- We propose a general Bayesian-based probabilistic framework that provides a way to feasibly fuse HOI events with RSSI signals to enhance the tracking performance. In particular, for a multiple-room scenario (including two bedrooms and a kitchen), *HOI-Loc* can achieve average 95% localization accuracy and 58*cm* tracking error, offering about 1.3×, 1.86× and 2.86× improvement compared with Twins [148], TagTrack [138] and SCPL [37].

- *HOI-Loc* can accurately track up to three residents with average 85*cm* error distance in a non-concurrent case, and it is capable of decoding four basic living postures with overall 94.7% accuracy.

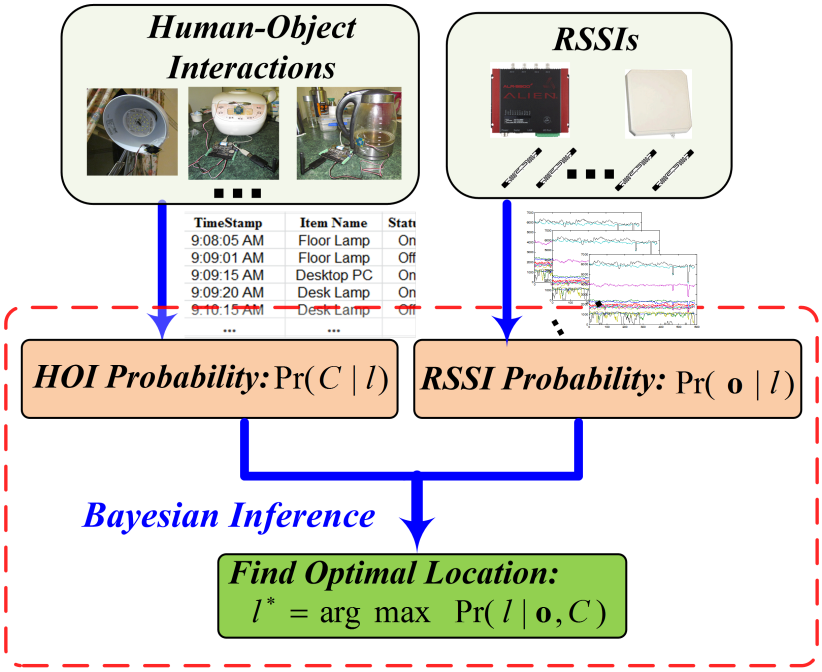


Fig. 5.2 *HOI-Loc* system overview

5.2 Preliminary

In this section, we will theoretically analyze the RFID backscatter communication mechanism and verify *HOI-Loc*'s capability to reach device-free localization and tracking.

5.2.1 Received Signal Strength Indicator (RSSI)

Passive RFID system communicates based on the backscatter radio link since the passive tags (no batteries powered) can purely harvest energy from the antenna's signal. RSSI measures the power of received radio signal between the tag and reader antenna [141]. We detail the characteristics of RSSIs from passive RFID tags in Chapter 4 (please see details in Sec. 4.2.2).

5.2.2 Human-Object Interactions (HOI)

Human-Object Interactions, study the interactions between human and the surrounding smart objects, facilitating the booming of context-aware computing [153]. Currently, many

researchers from computer vision community investigate how to utilize the HOI for object tracking/recognition [154], action recognition [155] or human postures detection [156]. In our daily lives, we also observe that resident's interactions with surrounding devices can be very helpful to reveal her locations in a home environment. Considering the following scenarios, when the door of a *microwave oven* is open, it is very likely the person is near the oven; if the *desk lamp* goes from on to off, or from off to on, we can almost be certain that the subject currently is in her home office. Thus, inspired by the observation, some HOI contexts can be very valuable to infer the target's possible locations. Then, the only question left is how we can monitor the HOI events in real-time, which although is not a challenging issue, still needs to be well-designed. In our system, we use the products of Phidgets Inc.[2] (i.e. Single Board Computer PhidgetSBC3, Phidget light sensor, Phidget touch sensor, Phidget motion sensor etc.). The sensors mounted in electrical appliances communicate with the PC through WiFi. We use the Microsoft .NET framework and SQL Server 2012 to manage the interaction events, which can be easily aggregated and visualized. Fig. 5.10 shows the hardware deployment in a bedroom. The whole detailed experimental settings can be seen from Fig. 5.9.

5.3 HOI-Loc Overview

Ultra-low cost of UHF tags (5$\sim$10 cents each) become the preferred choice of many industry applications. Following the common practice, we focus on device-free localization based on passive UHF tags in this chapter. Today's COTS RFID readers have an operating range of around $10m$, which is enough for a residential room. We also focus on locating and tracking residents that are not moving at a high speed ($< 1m/s$) since moving in a high speed in a residential room is unlikely.

[2]http://www.phidgets.com/

5.3.1 Problem Definition

We consider the target resident moving within a surveillance house. For each monitored house with D passive tags deployed, we divide it into J zones, denoted by $\mathscr{L} = \{L_1, L_2, ..., L_J\}$. When a subject appears in zone L_i, we collect N sample data $\mathbf{S}_i = \{\mathbf{s}_{i1}, \mathbf{s}_{i2}, ..., \mathbf{s}_{iN}\}$, where $\mathbf{s}_{ij} \in \mathbb{R}^D$ means data collected in jth sampling period. As a result, when going through all the zones, we can obtain a dataset $\mathscr{S} = \{\mathbf{S}_0, \mathbf{S}_1, ..., \mathbf{S}_J\}$, quantifying how a subject affects RSSIs from each zone. Here the environmental RSSIs without a subject is represented by $\mathbf{S}_0$. Similarly, for modeling HOI events, we assume that we overall have M different objects $C = \{I_1, I_2, ..., I_M\}$ available (e.g. electrical kettle, fridge, microwave oven etc.). Then we represent the interaction events in a binary way, *i.e.,* $I_i = 1$ means an interacting event happens. For example, if I_1 represents *fridge door*, then $I_1 = 1$ means the *fridge door* has opened from closed, or closed from opened (interacted with by a resident), otherwise $I_1 = 0$. Formally, given both signal available, this chapter targets the following two problems.

Problem 4 (Localization) *Can we locate a monitored subject by learning the RSSI patterns and interaction events of human with objects? Formally, given an RSSI vector and interaction events, we need to correctly estimate the subject's location.*

Problem 5 (Tracking) *Can we track a moving subject by learning RSSI changes and HOI event streams? Formally, given a continuous RSSI sequence and interaction event stream, we need to accurately estimate the subject's trajectory.*

5.3.2 Solution

Localization: Mathematically, *Problem 1* can be formulated as modeling the posterior distribution $Pr(l|\mathbf{o}, C)$ for each possible location. In particular, given observed RSSI signals $\mathbf{o}$ and corresponding interaction events $C = \{I_1, I_2, ..., I_M\}$, we find the most likely location

by using

$$l^* = \arg\max_{l \in \mathscr{L}} Pr(l|\mathbf{o}, C) \tag{5.1}$$

which is essentially a classification task. We need to model how RSSIs are distributed in different geographical areas based on a sample of measurements collected at several known locations and how to feasibly update the posterior probability of the classifier based on the contexts of HOI. We present our method in §5.4.

Tracking: When a resident walks in random zones, we can collect T continuous RSSI vectors $\mathscr{O} = \{\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_T\}$. Then, mathematically, *Problem 2* can be formulated as modeling the posterior distribution:

$$Pr(l_{1:T}|\mathscr{O}, \mathscr{C}) = Pr(l_{1:T}|\mathbf{o}_{1:T}, C_{1:T}), l_{1:T} \in \mathscr{L} \tag{5.2}$$

Then, given observed continuous RSSI vector sequence $\mathbf{o}_{1:T}$ and interaction event stream $C_{1:T}$, we need to find the location sequence with largest likelihood.

$$l^*_{1:T} = \arg\max_{l_{1:T} \in \mathscr{L}} Pr(l_{1:T}|\mathbf{o}_{1:T}, C_{1:T}) \tag{5.3}$$

The tracking problem can be regarded as given a continuous RSSI stream and a HOI event sequence, how we can recover the underlying location sequence which is as accurate as possible to the true location trajectory. We elaborate our solution in §5.5.

5.4 Localization

As aforementioned in Eqn.5.1, for localizing a static resident, we need to model the posterior distribution $Pr(l|\mathbf{o}, C)$ given RSSI signal and HOI events. However, it is difficult to measure the posterior likelihood since we cannot know the RSSI signal patterns and HOI events before the resident appearing in the monitored area. But we can observe what happened (*e.g.,* the

RSSI changes, the operating status of electronic appliance) when the resident located in a certain location during our experiments. So how to bridge the gap? Let us first do a more thoroughly analysis to Eqn. 5.1. Based on the *Bayesian Inference Theorem*, we can decode the equation as

$$\begin{aligned} Pr(l|\mathbf{o},C) &= \frac{Pr(l)Pr(\mathbf{o},C|l)}{Pr(\mathbf{o},C)} = \frac{Pr(l)Pr(\mathbf{o}|l)Pr(C|l,\mathbf{o})}{Pr(\mathbf{o},C)} \\ &\propto Pr(l)Pr(\mathbf{o}|l)Pr(C|l,\mathbf{o}) \end{aligned} \tag{5.4}$$

Since RSSI signals and HOI events are from independent sensor sources, we also have $Pr(C|l,\mathbf{o}) = Pr(C|l)$. Thus, we can model the posterior probabilities of candidate locations as

$$Pr(l|\mathbf{o},C) \propto Pr(l)Pr(\mathbf{o}|l)Pr(C|l) \tag{5.5}$$

where $Pr(l)$ is the prior probability distribution, which is set as $Pr(l) \sim 1/J$ without losing generality (means the target can be possible in any locations beforehand). So far, we successful find a way to model the posterior distribution $Pr(l|\mathbf{o},C)$. We give the following two definitions.

Definition 1 (RSSI Probability) *Given the resident appearing a certain location, RSSI Probability measure the probabilistic distribution $Pr(\mathbf{o}|l)$ of RSSI signals.*

Definition 2 (HOI Probability) *Given the resident interacting with objects in a certain location, HOI Probability measure the probabilistic distribution $Pr(C|l)$ of HOI events.*

Next, we need to deal with how to accurately measure $Pr(\mathbf{o}|l)$ and $Pr(C|l)$.

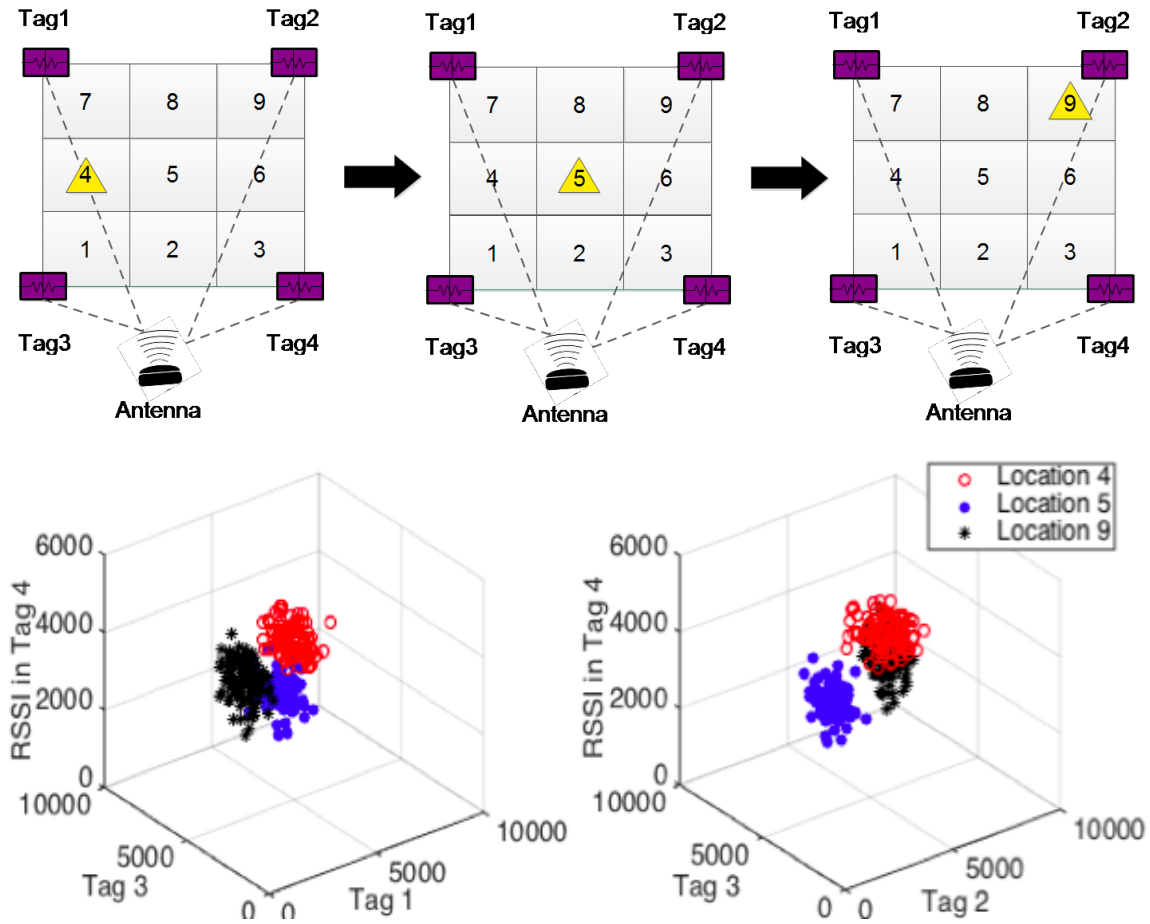


Fig. 5.3 RSSIs clustering in different HD spaces for subject in different locations

5.4.1 RSSI Probability

As elaborated in §5.2, mapping the RSSI signal to locations is very challenging under a clustered environment due to rich multi-path effect. Seeking solutions from backscatter propagation analysis is impractical in our case, laying in the facts: most backscatter communication models is depend on the assumption that the position of reference tag is accurately measured beforehand [47], which is not applicable in *HOI-Loc* (we relax the assumption, no need to know tag's coordinates). From Fig. 5.3, we can observe that the RSSI readings always cluster in a relatively same HD (high-dimension) space (treating one tag's signal as one dimension) when the resident appearing in a same location. Thus, based on this intuition, we propose a Probabilistic Polyhedron Isolation (PPI) method to efficiently locate

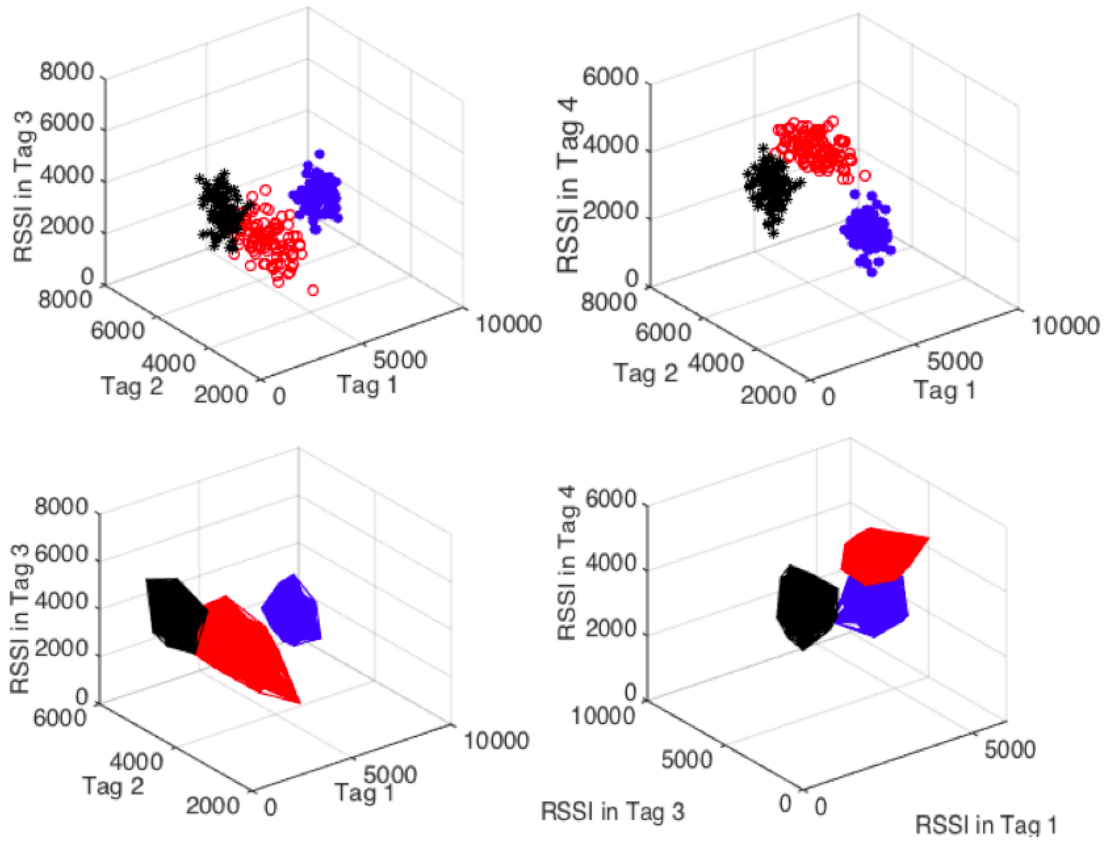


Fig. 5.4 RSSIs from different locations are bounded by isolated HD polyhedrons

the high-dimension space, inspired by kNN searching. Intuitively, kNN is based on distance estimation that assigns an Euclidean distance between any two RSSI samples. An observed RSSI sample is classified by a majority vote and assigned to the most common zone among its k nearest neighbors, which effectively eliminates the loss or jumping of signal reading by a *majority voting* mechanism. However, it only output a class-membership rather than offering probabilistic information, which is not capable for fusing with HOI events. Thus we propose the PPI method that works as follows. Assuming for each observation $\mathbf{o}$, we search its k nearest neighbors from the training set $\mathscr{S}$ in the high-dimension space, denoted as $N(\mathbf{o}) = \{\mathbf{s}_k | \mathbf{s}_k \in k\text{NN}(\mathbf{o})\}$. The training samples collected in location L_i among $N(\mathbf{o})$ is represented as $N^i(\mathbf{o}) = \{\mathbf{s}_k^i | \mathbf{s}_k^i \in N(\mathbf{o}) \cap \mathbf{s}_k^i \in \mathbf{S}_i\}$. In fact, $N^i(\mathbf{o})$ represent each isolated *HD-polyhedron*. Geometrically, the ith HD polyhedron (mapping to location L_i) is formed

by several high-dimension points (*e.g.,* RSSIs from all tags within a sampling time) in $N^i(\mathbf{o})$, illustrated as Fig. 5.4. Then, we can estimate *RSSI Probability* by measuring the Euclidean distance of testing RSSI readings with each isolated HD Polyhedron.

$$Pr(\mathbf{o}|l_i) = \begin{cases} \dfrac{\sum_{\mathbf{s}_k^i \in N(\mathbf{o})} \dfrac{1}{dis(\mathbf{o},\mathbf{s}_k^i)}}{\sum_{\mathbf{s}_k \in N(\mathbf{o})} \dfrac{1}{dis(\mathbf{o},\mathbf{s}_k)} + \sum \alpha}, & \text{if } |N^i(\mathbf{o})| \geq 1 \\ \dfrac{\alpha}{\sum_{\mathbf{s}_k \in N(\mathbf{o})} \dfrac{1}{dis(\mathbf{o},\mathbf{s}_k)} + \sum \alpha}, & \text{if } |N^i(\mathbf{o})| = 0 \end{cases} \tag{5.6}$$

where l_i indicates the target appears in location L_i, $(i = 1,..,J)$; $|N^i(\mathbf{o})|$ means the number of elements contained in $|N^i(\mathbf{o})|$, so does $|N(\mathbf{o})|$; α is a parameter with a very small value to avoid 0 probability for some locations where no training sample included in $|N(\mathbf{o})|$. In our case, it is chosen by

$$\alpha = 0.001 \max_{\mathbf{s}_k \in N(\mathbf{o})} \frac{1}{dis(\mathbf{o},\mathbf{s}_k)} \tag{5.7}$$

Eqn.5.6 gives the posterior distribution by finding its HD polyhedron and measuring its distance with the test sample. As Fig. 5.5 shown, we compare proposed PPI method with traditional kNN in a $1.8m \times 1.8m$ area (see Fig. 5.3, with 4 tags and 9 virtual grid). The PPI method outperforms kNN in all k values.

To conclude, it is superior in the following two ways: *i)* it specifically gives the posterior distribution of each locations by measuring the context distances, providing a way to fuse with HOI information; and *ii)* it utilizes HD polyhedron to code with signal distortion, *i.e.,* RSSI fluctuation is robustly tolerated by boundaries of a learned high-dimension space). However, merely based on RSSI Probability, we still cannot achieve satisfied localization accuracy in a clustered environment. We run a pilot experiment in a residential master-room (see Fig. 5.9, Area: $3.6m \times 4.8m$). As Fig. 5.6 shows, the average accuracy is around 80%,

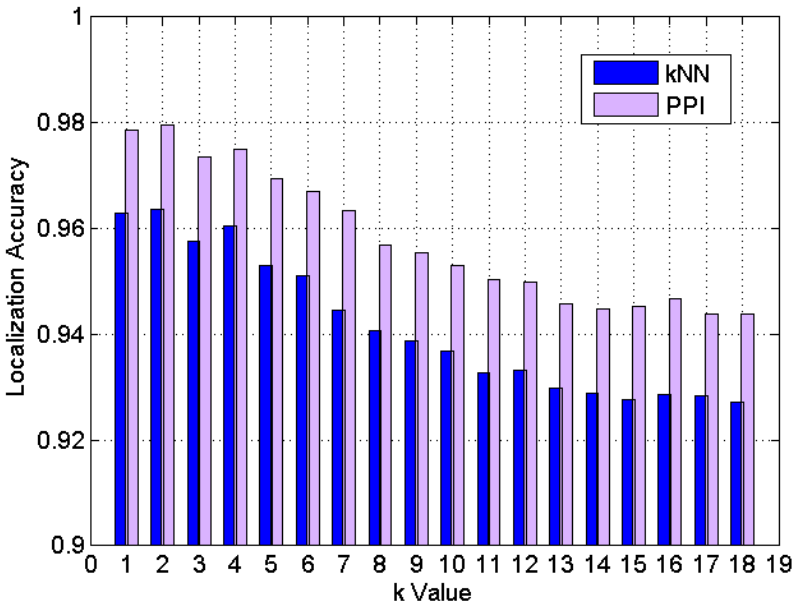


Fig. 5.5 Localization accuracy for proposed PPI and traditional *k*NN

Ground Truth	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
L1	178	0	1	0	0	0	0	0	0	1
L2	7	145	19	2	1	0	2	0	1	3
L3	2	17	153	5	2	0	0	0	0	1
L4	3	2	7	139	26	1	0	1	0	1
L5	2	0	4	17	148	7	1	0	0	1
L6	0	0	0	0	6	170	3	1	0	0
L7	0	0	0	0	0	5	171	4	0	0
L8	0	0	0	0	0	0	0	176	2	2
L9	0	0	0	0	0	0	1	1	176	2
L10	0	0	0	0	0	0	0	2	1	177

Fig. 5.6 Localization result based on RSSI signal (k=2)

and it mis-classifies the adjacent locations such as *L2* and *L3*, *L4* and *L5*. As a result, the unsatisfied localization performance motivates us to exploit the HOI events.

5.4.2 HOI Probability

HOI contexts basically reflects the interacting status of the resident with her environment at a particular point of time, which can be utilized to facilitate the localization. Based on the

problem definition in §5.3, for N continuous time slots, we can retrieve an interaction events data set $\mathscr{C} = \{C_1, C_2, ..., C_N\}$, where $C_i = \{I_1^i, I_2^i, ..., I_M^i\}$ represents statuses of M interacting events at ith time. We assume that, for each HOI event happening, there exists at least one candidate location, which is the criterion we choose HOI events. Thus, for each object I_i , its possible locations can be denoted as

$$L_{I_i} = [L_1^{I_i}, L_2^{I_i}, ..., L_J^{I_i}]^T \tag{5.8}$$

where $L_j^{I_i} = 1$ means L_j is the possible location of the subject regarding interaction event I_i; $L_j^{I_i} = 0$ means L_j is not the possible location. For overall M objects, we have

$$L_I = [L_{I_1}^T, L_{I_2}^T, ..., L_{I_M}^T]^T \tag{5.9}$$

Thus, given the interaction events with all objects $C = \{I_1, I_2, ..., I_M\}$, we can infer all the possible locations based on *HOI Matrix*, defined as:

Definition 3 (HOI Matrix) *HOI Matrix indicates all the possible locations for HOI events happen at a certain time, calculated by* $M_{HOI} = [I_1 L_{I_1}^T, I_2 L_{I_2}^T, ..., I_M L_{I_M}^T]^T$.

To avoid the cases that no available interaction events can be utilized to infer some certain candidate locations, we smooth the zero probability with adding a small value parameter β. Based on our numerical experiences, β does not affect the final estimation as long as it is small enough since it produces much smaller probability comparing to other cases. In this chapter, we choose $\beta = 0.001$. Then, we can estimate $Pr(C|l)$ for each possible locations based on *Algorithm 1*. In particular, for each timestamp, we receive a M_{HOI} to indicate current HOI status, then we feed it into *Algorithm 1* to get the *HOI Probability*.

In summary, based on Algorithm 1 and Eqn.5.6 and Eqn.5.5, we can conveniently integrate HOI events with RSSI signals under a *Bayesian Inference* probabilistic framework to estimate a subject's location with maximum likelihood. Through fusing these two signals,

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
L1	180	0	0	0	0	0	0	0	0	0
L2	0	179	1	0	0	0	0	0	0	0
L3	0	2	176	2	0	0	0	0	0	0
L4	0	0	2	175	3	0	0	0	0	0
L5	0	0	0	1	176	3	0	0	0	0
L6	1	0	0	0	3	172	2	0	0	2
L7	0	0	0	0	0	3	174	2	0	1
L8	0	0	0	0	0	0	1	176	3	0
L9	0	0	0	0	0	0	0	2	177	1
L10	0	0	0	0	0	0	0	0	1	179

Fig. 5.7 Localization result of fusing HOI events with RSSI signal (k=2)

HOI-Loc greatly increases the localization accuracy, illustrated by Fig. 5.7. With fusing HOI events, our method achieves overall more than 96% accuracy. We also make comparisons with methods that are frequently adopted by other fingerprinting-based localization systems, including SVM [143], EML [157], and Naive Bayes (see experiments in §5.6).

5.5 Tracking

Comparing to localize a relatively static resident, tracking a moving subject is more challenging, mainly because *i)* the moving subject causes the obvious RSSI signal latency (when estimating the current location, the subject already move to another place); and *ii)* the moving body introduces sudden, unpredictable RSSI signal pattern changes enabling us difficult to mapping the signals to locations. However, we observe that the next moving zone is usually adjacent and only adjacent to current locations[3]. So we can narrow down candidate locations as long as we estimate current resident's location. Intuitively, we introduce a Hidden Markov Model to model such location transition process. Then we need to deal with how to feasibly integrate both RSSI signal sequence and HOI event stream into a HMM framework.

[3] As mentioned before, we assume that the resident move naturally at residential room, less than $1m/s$. Under a $2Hz$ sampling frequency, moving distance is less than $0.5m/s$, within the range of one grid.

Algorithm 2: HOI Probability $Pr(C|l)$ Estimation

Input: HOI Matrix $M_{HOI} \in \mathbb{R}^{M \times J}$, β
Output: $Pr(C|l_j), l_j \in \mathscr{L}$

```
PossibleLocaSum = 0;
for i = 1 : M do
    for j = 1 : J do
        if M_HOI(i, j) == 1 then
            PossibleLocaSum=PossibleLocaSum+1;
        end
    end
end
for j = 1 : J do
    PossibleLocaSumj=0;
    for i = 1 : M do
        if M_HOI(i, j) == 1 then
            PossibleLocaSumj=PossibleLocaSumj+1;
        end
    end
    if PossibleLocaSumj ≠ 0 then
        Pr(C|l_j) = PossibleLocaSumj / PossibleLocaSum;
    end
    else
        Pr(C|l_j) = β / PossibleLocaSum;
    end
end
```

First, we revisit the definition of *Tracking Problem* in §5.3. Actually, we can decode the Eqn. 5.2 based on *Bayesian Inference* in the same way.

$$\begin{aligned} Pr(l_{1:T}|\mathbf{o}_{1:T}, C_{1:T}) &= \frac{Pr(l_{1:T}, \mathbf{o}_{1:T}, C_{1:T})}{Pr(\mathbf{o}_{1:T}, C_{1:T})} \\ &\propto Pr(l_{1:T}, \mathbf{o}_{1:T}, C_{1:T}) \end{aligned} \tag{5.10}$$

Similarly, since RSSI signal and HOI events are from independent sensor sources, and current state is only conditionally depend on previous one, we can further decode the above equation

as

$$
\begin{aligned}
&Pr(l_{1:T}, \mathbf{o}_{1:T}, C_{1:T}) \\
&= Pr(l_1)Pr(\mathbf{o}_1|l_1)Pr(C_1|l_1)\prod_{t=2}^{T}\underbrace{Pr(\mathbf{o}_t, C_t|l_t)}_{A}\underbrace{Pr(l_t|l_{t-1})}_{B} \\
&= Pr(l_1)Pr(\mathbf{o}_1|l_1)Pr(C_1|l_1)\prod_{t=2}^{T}\underbrace{Pr(\mathbf{o}_t|l_t)}_{A_1}\underbrace{Pr(C_t|l_t)}_{A_2}\underbrace{Pr(l_t|l_{t-1})}_{B}
\end{aligned}
\tag{5.11}
$$

So far, we successfully decompose our tracking problem into estimating two *Emission Matrix* A_1 and A_2, and *Transition Matrix* B. We observe that A_1 and A_2 are exactly the same forms (except the times-tamps) as the *RSSI Probability* and *HOI Probability*. As a result, for tracking problem, we can also apply Eqn. 5.6 and Algorithm 1 to estimate the two emission matrices A_1 and A_2 respectively.

5.5.1 Transition Strategy

Transition matrix measures the probability of a subject moving to next location at each time t, which is defined as $A_{ij} = Pr(a_t = l_j|a_{t-1} = l_i)$. However, based on the common-sense, a subject can only move a step at a time, meaning that it is highly unlikely for the subject to walk from the lower-left corner to the upper-right corner or walk through a bed within a sampling time (0.5s in our case). Therefore, we adopt a *Adjacent Transition* strategy to calculate the probabilities of next candidate locations given current location.

Definition 4 (Adjacent Transition) *The subject can only move to a feasible location that is adjacent (including current location which means still) to current location with equal probabilities, and the probabilities of moving to other locations are very small.*

Based on the proposed strategy, we assume that location l_i denotes the subject appears in zone L_i. Given current location l_i, all the possible locations that the subject can move

to belong to the set ψ_i, and the number of locations contained in the set is $|\psi_i|$. Thus, the transition probability matrix can be expressed as

$$A_{ij} = Pr(l_j|l_i) = \begin{cases} \frac{1}{|\psi_i|} & \text{if } l_j \in \psi_i \\ 0 & \text{if } l_j \notin \psi_i \end{cases} \tag{5.12}$$

where ψ_i is defined according to the proposed strategy.

5.5.2 Viterbi Searching

Having *Emission Matrix* and *Transition Matrix*, we can search the most likely sequence of state transitions in a continuous time stream via the *Viterbi* algorithm defined by $V_j(t)$, the highest probability of a single path of length t which accounts for the first t observations and ends in location L_j:

$$V_j(t) = \arg \max_{l_1, l_2, \ldots, l_t} Pr(l_{1:t-1}, l_t = L_j; \mathbf{o}_{1:t}; C_{1:T}|A, B) \tag{5.13}$$

where A and B can be found in Eqn.5.11. Further, by induction:

$$\begin{aligned} V_j(1) &= A_{1j} = (A_1)_{1j}(A_1)_{1j} \\ V_j(t+1) &= \arg\max_i V_i(t) B_{ij} (A_1)_{t+1,j} (A_2)_{t+1,j} \end{aligned} \tag{5.14}$$

where $(B_1)_{1j} = Pr(\mathbf{o}_1|l_j)$ and $(B_2)_{1j} = Pr(C_1|l_j)$. Finally, we can recovery an optimal path with the maximum likelihood. Next, we need to deal with the latency issue in tracking system.

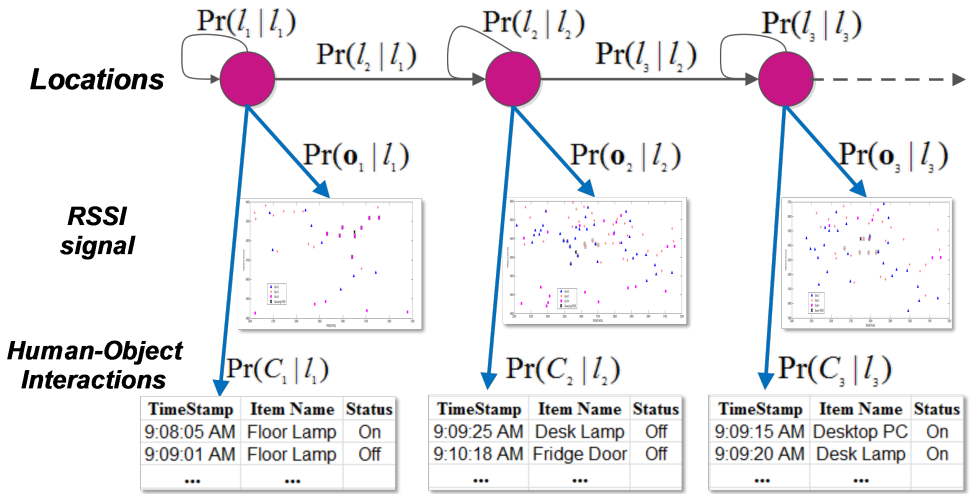


Fig. 5.8 HMM tracking mechanism by fusing RSSI signal and HOI events

5.5.3 Forward Calibration

We find some latency in detecting the subject, which is mainly caused during the RSSI collection process and by the delay of signals sent by passive tags [141]. The RSSI collector is programmed with a timer to poll the RSSI with a predefined order of transmission, taking around $1 \sim 2s$ to complete a new measurement with no workarounds. To cope with the issue, we adopt a forward calibration mechanism to calibrate the estimated location sequences to offset the latency, which uses a moving time averaging window to recalculate the coordinates of location sequence obtained by *Viterbi Searching*. In particular, the estimated coordinates $\hat{c}_i : (\hat{x}_i, \hat{y}_i)$ location l_t at time t can be calculated as:

$$\hat{c}_t = \frac{\sum_{i=t}^{t+|w|-1} \tilde{c}_i}{|w|} \tag{5.15}$$

where $|w|$ is the window length. $\tilde{c}_i$ is uncalibrated coordinate of predicted grids centroid at time t. Based on our pilot experiments, we find that *HOI-Loc* achieves the best performance when $|w| = 7$.

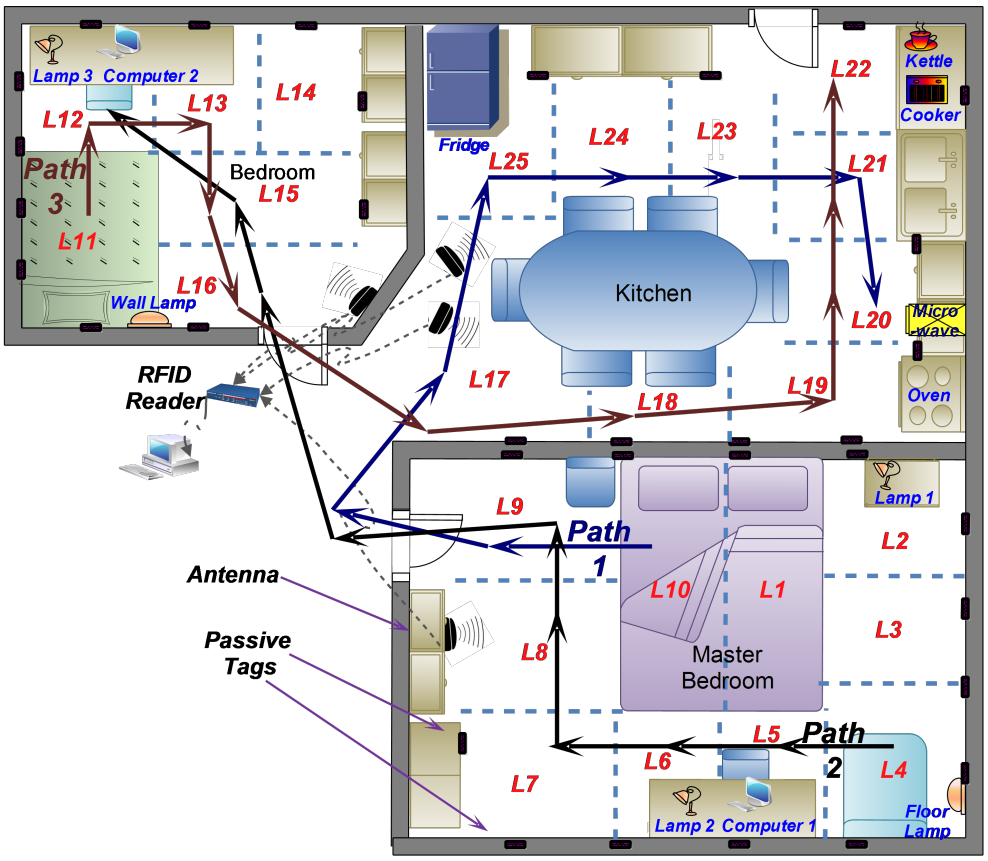


Fig. 5.9 Experiment settings and paths

5.6 Implementation and Evaluation

We set up COTS RFID hardware in a residential house with two bedrooms and a kitchen (see Fig. 5.9), including an Alien ALR-9900+ Enterprise RFID Reader, 4 two-circular antennas, and multiple squiggle Higgs-4 passive tags. The reader operates at 840-960MHz and supports UHF RFID standards such as ETSI EN 302 208-1. We set the sampling rate as $2Hz$ and each tag reading contains time stamp, tag ID, antenna ID and the RSSI value, which are then processed by a computer with an i7-3537U 2.5G processor and 8G RAM, running WINDOWS 7. Based on our preliminary experiments, we place the antenna about $1.7m$ above the ground and facing tags with approximately 45° in order to catch all readings of reference tags in a non-subject environment. We attach passive RFID tags to the wall with an approximate $0.6m$ interval (shown as Fig. 5.10). During the localization and tracking,

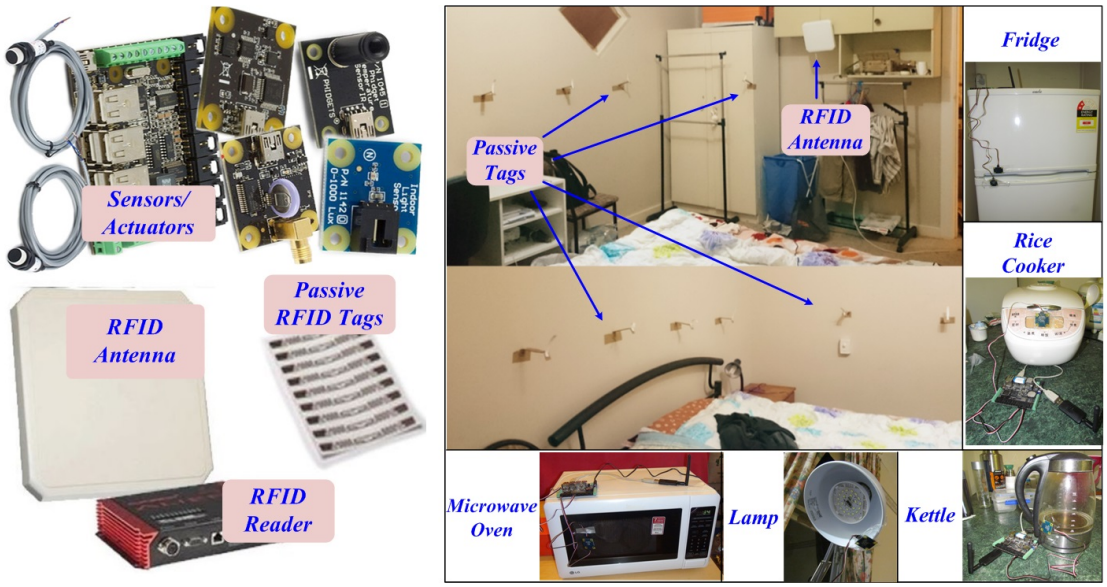


Fig. 5.10 Sensors and RFID hardare deployment
Testing Area: master bedroom: $3.6m \times 4.8m$*, bedroom:* $3m \times 3.2m$*, kitchen:* $3.6m \times 4.6m$

we send an RSSI request to all tags within a sampling period. If we cannot receive RSSI readings of a certain tag, its RSSI value will be set to 0. Thus, for all timestamps, we have the RSSI vectors with the same dimension.

Before evaluating our approach, we need to deal with two practical issues: one is how to decide the zone size, and the other is how to choose the specific HOI events and their corresponding candidate locations. Based on our empirical study, the smaller the grid size, the higher false classification rate will be due to more indistinguishable zones, and more profiling data are needed as well. Thus, smaller zone size can offer high localization resolution but increase the calibration burden. In reality, extreme high resolution in localization is not the main concern. For instance, in an elderly people assistant system, caregivers are generally desirable to know which sub-area the elderly is rather than a very fine-grained location point if achieving the later goal is at expensive cost. Therefore, we defined our virtual zones as shown in Figure 5.9. For HOI events, the priority is given to the objects that the resident frequently interacted or used, and their operation status can be easily monitored based on COTS sensors. We treat the zones that are adjacent to the object as the possible candidate

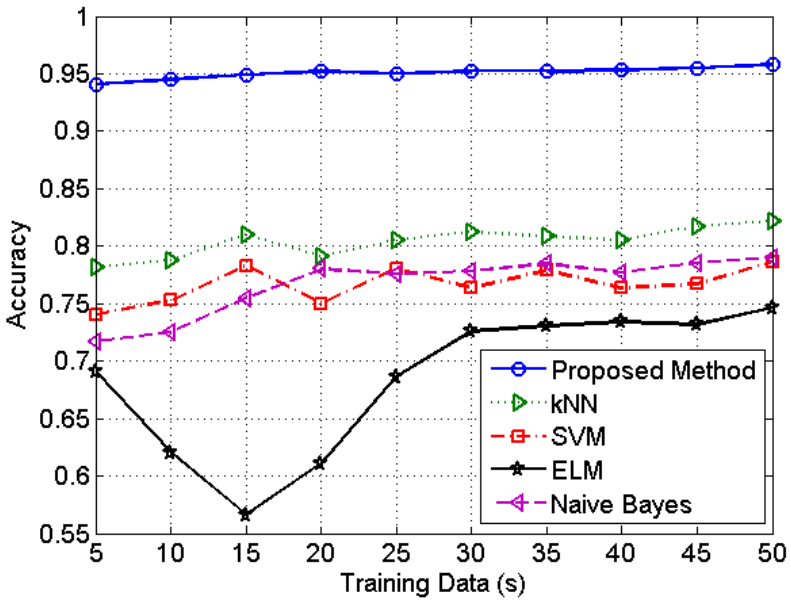


Fig. 5.11 Localization result for Stationary Scenario

locations when interacting events happens, *e.g.,* when the fridge door is open, the potential locations[4] are *L25*, *L24* and *L17*.

5.6.1 Evaluation Metrics

We adopt standard localization accuracy and error distance to measure our proposed approaches in terms of localization and tracking respectively [32]. The *localization accuracy* is defined as

$$Accu. = \frac{\sum_i^N \mathbb{I}(\hat{l}_i, l_i)}{N} \tag{5.16}$$

where $\mathbb{I}(\hat{l}_i, l_i)$ is an indicator, which equals to 1 if estimated zone $\hat{l}_i$ is as same as the ground truth zone l_i, otherwise equals to 0; N is the total number of the testing RSSI measurements.

[4]When we detect the HOI event, the subject may move to a adjacent locations (if not in *L25*) within 0.5*s*.

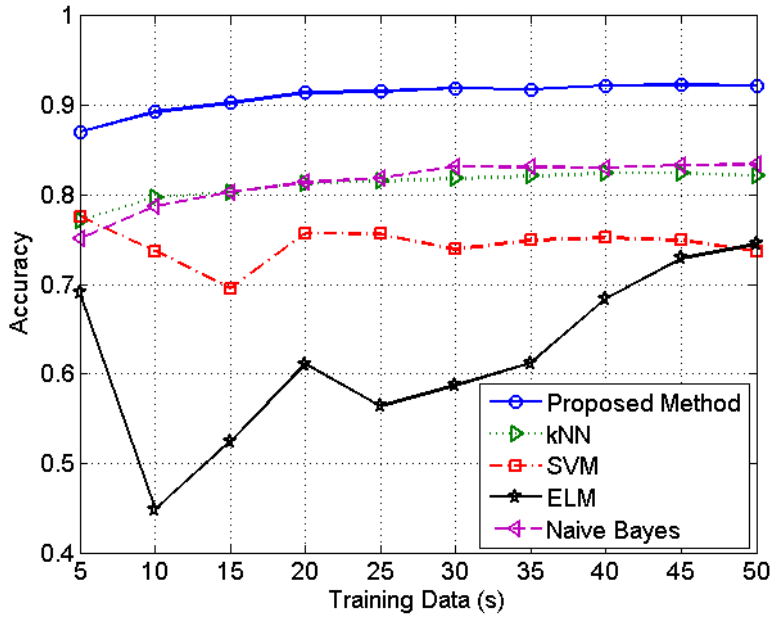


Fig. 5.12 Localization result for Dynamic Scenario

The error distance denotes the averaging accumulated error distance of the testing samples in each continuous trajectory, it is calculated using

$$Dis_{err.} = \frac{\sum_{i}^{|T|} dis(\hat{c}_i, c_i)}{|T|} \tag{5.17}$$

where c_i is the coordinates of the actual location of the subject at time i, and $dis(\hat{c}_i, c_i)$ is the Euclidean distance between predicted coordinates and actual coordinates, $|T|$ is the total number of testing samples generated by a trajectory.

5.6.2 Localization

Shown as Fig. 5.9, we test the performance in a residential home that is divided into 25 virtual grids. To be more practical, we defined the following three scenarios to mimic daily-living activities in our experiments.

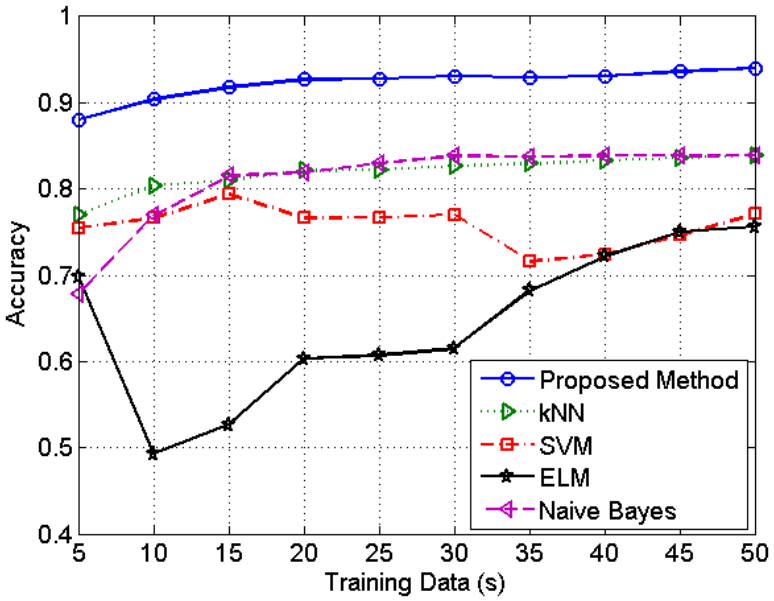


Fig. 5.13 Localization result for Mixed Scenario

- *Scenario 1 (Stationary)*: Assuming a subject is standing/sitting in an unknown place in the monitored area still, such as watching TV or waiting for someone.

- *Scenario 2 (Dynamic)*: Assuming a subject keeps moving around or with some activities in a small unknown area, such as cooking in the kitchen.

- *Scenario 3 (Mixed)*: Assuming a subject presents in an unknown place and performs a combination of Scenario 1 and Scenario 2, such as doing some exercises for a while and then watching TV.

Based on the predefined three scenarios, we collected three types of data to test our method: *i)* a subject is standing in each grid for 120 seconds, *ii)* a subject keeps moving around within each grid for 120 seconds, and *iii)* combining both activities for 120 seconds in each grid[5]. Then we randomly divide it as training data (*i.e.,* 5 seconds∼50 seconds data per grid) and testing data (*i.e.,* 115 seconds ∼70 seconds data per grid). In each case, we do exper-

[5]Three participants with different genders, heights and weights join our experiments. We overall collect 612,000 RSSI readings with one week experiment timespan.

Table 5.1 The percentage improvements for the accuracy of our method over the other approaches

Scenarios	*k*NN	SVM	ELM	Native Bayes
S1	12.63%	17.89%	21.05%	12.63%
S2	10.87%	19.56%	18.47%	9.78%
S3	10.63%	18.08%	19.14%	10.63%

iments 20 times to get the mean accuracy. The testing result is shown as Fig. 5.11 $\sim$ Fig. 5.13. The parameter settings are: k=2 for kNN; SVM (linear kernel, terminate criterion=0.01, C=1, others as default); ELM (*hardlim* activation function, *NumberofHiddenNeurons*=600, others as default, average result of running 20 times); NaiveBayes (normal distribution, uniform prior probabilities, others as default).

For Scenario 1, all classification methods achieve more than 75% localization accuracy with 50 seconds' training data. In particular, the proposed method is able to achieve 95.6% accuracy with only 5 seconds of training data, which exhibits great advantage than other fingerprint-based schemes. In previous work, the shortest time needed for collecting training data to get same localization accuracy is about 60 seconds [45]. Our system only needs to collect 5 seconds of training data to reach a better localization accuracy, improving 12 times. For Scenario 2, the best localization accuracy is 93.7%, achieved by our method. It is worth to mention that, performance in this case is more sensitive to the size of training data. It may lie in fact that more training data can better interpret more informative RSSI patterns for the *dynamic* scenario compared to the *stationary* scenario. For Scenario 3, the localization accuracy can reach 95.2%. Table 5.1 summarizes the percentage improvements for the accuracy results of the proposed approach over the other classification methods such as SVM, ELM, Native Bayes, and kNN. It shows that our method generally improves the accuracy at around 10% $\sim$ 20%. To conclude, our method achieves a better localization performance among all the methods, and it is also more robust to the RSSI uncertainties in case of limited training data.

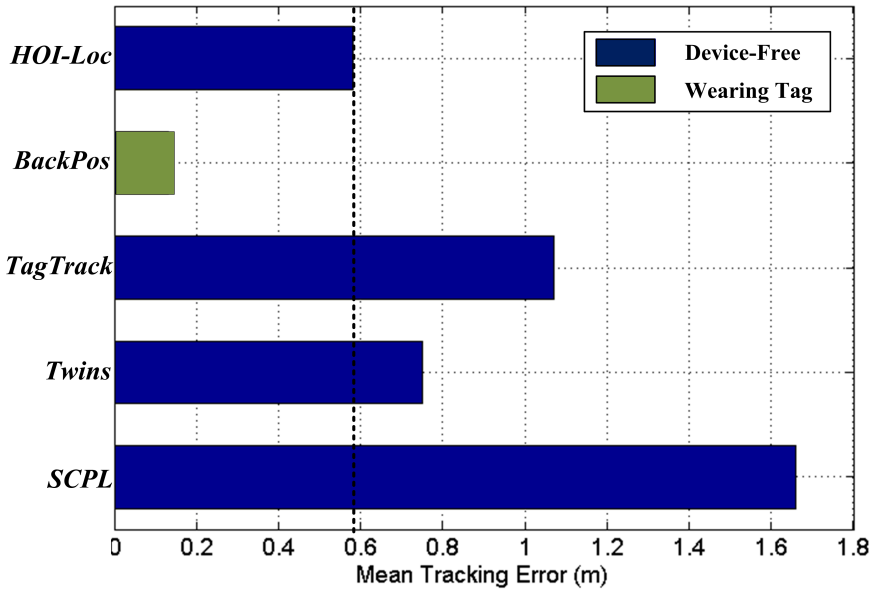


Fig. 5.14 Compare tracking accuracy of *HOI-Loc* with other state-of-the-art systems

5.6.3 Tracking

We evaluate tracking performance on three paths (see Fig. 5.9), which respectively simulate three real-life scenarios: i) the subject gets up from the bed in the master bedroom and opens the *fridge*, takes out some food to do *cooking* in the kitchen; ii) the subject stands up from sofa in the master bedroom and goes to *work on the desk* in the study room, and iii) the subject gets up from the bed in the small bedroom and walks through the kitchen and *boils water* using the *electric kettle*. Three subjects with different heights and weights join the tracking experiments, and each participant walks the three paths 20 times. We also review and compare *HOI-Loc* with the state-of-the-art RFID-based systems, shown as Fig. 5.14 and Fig. 5.15. The parameters used are: $k = 2$ for TagTrack and HOI-Loc; GMM component number=4 for GMM-CRF in SCPL.

- *TagArray:* TagArray [32] is the very first work that utilizes RFID tags to achieve device-free localization. It deploys active tags as an array to localize a subject when RSSI of some anchoring tags variate beyond a threshold. However, it requires high tag density, relatively expensive and needs pre-calibrate the tags' locations.

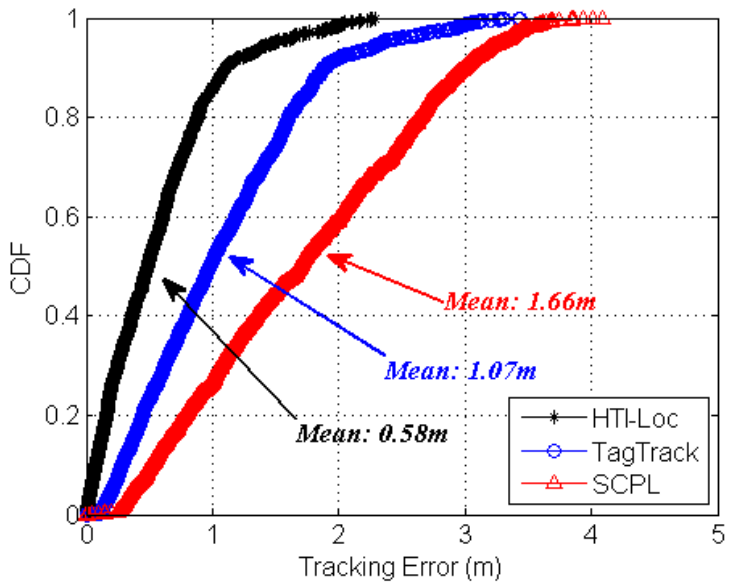


Fig. 5.15 Trackng error CDF (cumulative distribution function) for different device-free methods

- *TASA:* TASA [33] is another tag array-based localization scheme using both active and passive tags, which is more cost-effective comparing to TagArray. However, it tracking error is heavily correlated with the tag density, and it still requires to calibrate all tags' coordinates.

- *SCPL:* It is a device-free localization system based on wireless sensor nodes [37]. SCPL proposes a GMM (gaussian mixer model) based CRF (conditional random field) to track a moving subject. It reports average 1.3m tracking error. We apply its GMM-CRF in our testbed, achieving average 1.66m error.

- *Twins:* Twins [148] is a very recent device-free localization work based on pure passive tag, which leverages observations caused by interference among two passive tags to detect a single moving subject. It reports an average 0.75m tracking error in a relatively spacial warehouse. It requires to know the reference tags' locations in advance.

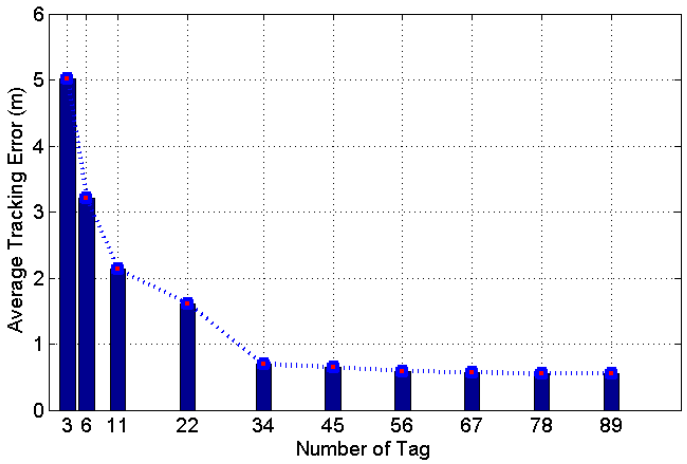


Fig. 5.16 Mean tracking errors using different tag numbers

- *BackPros:* BackPros [135] is the latest RFID-based positioning system that can achieve decimeter-level accuracy (*i.e.,* report 13*cm* mean tracking error). It requires the target to be attached with a tag and exploits the phase differences of backscatter signals to infer the tag's location. It needs carefully calibrate the positions of four antennas beforehand and the tracked subject attached with a tag.

- *TagTrack:* TagTrack [138] is a similar attempt using RFID signals to passively localize the objects. It deploys the passive tags as an array and uses the RSSI changes as the tracking indicator. It improves the result by introducing classification technique, reaching 70*cm* error distance. However, such accuracy is only achievable in a spacial, clear area. We also utilize its method to our test environment, achieving 1.07*m* mean error.

Unlike the above methods, *HOI-Loc* does not require the location contexts of reference tags, achieving 0.58*m* mean error distance in the testbed. As Fig. 5.14 shows, it offers about 1.3×, 1.86× and 2.86× improvement compared with Twins [148], TagTrack [138] and SCPL [37] in a residential house[6]. We also explores the relation of tag density with tacking

[6] In a residential home testbed, we do not compare *HOI-Loc* with TagArray and TASA since these two works need to put the tags in an array which is impractical especially in a full-furnitured house. Firstly, the reader

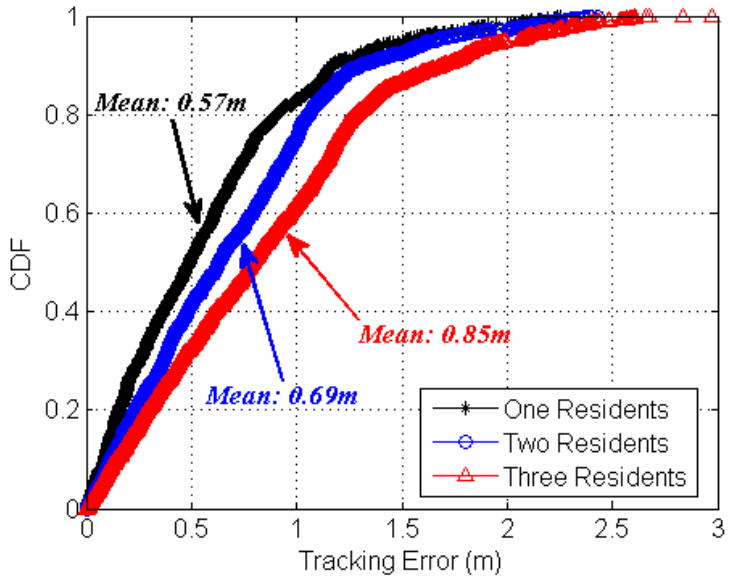


Fig. 5.17 Tracking errors for mutiple residents

error (see Fig. 5.16). We can see that the tracking performance will greatly degenerate when using less tags, e.g. in the case of 6 tags (2 tags per room), the error is more than $3m$. However, adding more tags (e.g. more than 34 tags) cannot enhance the performance significantly since a large number of tags are difficult to be interrogated by an antenna within a sampling time, causing more lost readings. As a result, the overall performance decays in this circumstance. To summarize, *HOI-Loc* can achieve high tracking accuracy using 34 passive tags, which relaxes the requirement of high-density tags deployment in TagArray and TASA

5.6.4 Beyond the Limits

To push the limits of *HOI-Loc*, we also conducted experiments in a multi-resident scenario. Two residents walked randomly among different rooms and interacted with the environment

even cannot catch the readings from passive tags that are deployed in a blanketed ground because signals are strongly blocked by furnitures around and absorbed by the blanket. Secondly, tag-arrays that densely deployed on ground in the residential room strongly obstruct the mobility of the resident, causing uncomfortable and inconvenient. In *HOI-Loc*, the passive tags are attached on the wall which is more practical and considered as less intrusion.

		Sitting	*Standing*	*Lying*	*Walking*	*Recall*
Ground Truth	*Sitting*	880	0	4	20	0.973
	Standing	4	708	56	14	0.905
	Lying	0	8	1160	18	0.978
	Walking	8	16	36	560	0.903
	Precision	0.987	0.967	0.924	0.915	0.947

Fig. 5.18 Confusion matrix of detecting four basic postures

(where the instrumented objects are available), and then for three residents[7]. As shown in Fig. 5.17, our method can track two residents with 0.69m average error and track three residents with 0.85m mean error. Although our experiments exclude the cases that multiple subjects are concurrent in the same room, from a practical view, the test scenarios is still valuable and frequently seen in real-living activities. Mostly, in each room (especially in bedroom) there are usually one resident, and even for a family of two or three, there plenty of time that people stay or work in a room alone. We also attempted to detect different postures of the resident, such standing, sitting, lying down and walking using our system. We observe that, similarly to localization, the RSSI signals embody different patterns when a resident performs different postures, which means the RSSI signal is not only the location indicator but also can be exploited as a human-activity indicator. Thus, we collected RSSI readings to feed into our Probabilistic Polyhedron Isolation method when the resident performing different postures in the bedroom. As Fig. 5.18 shows, we can successful to detect resident's postures with 94.7% accuracy. The results suggest that *HOI-Loc* provides an enabling primitive to recognize postures, besides tracking a moving resident. We can use this capability to better understand a resident's daily-living habits.

[7]Make sure there is only one resident in each room at a certain time, overall we collected 10,800 measurements.

5.7 Conclusion

To summarize, this chapter has shown how human object interaction events can be used to facilitate the COTS RFID-based device-free localization under a rigid probabilistic framework. The real-world experiments demonstrate the feasibility and effectiveness of our system, which marks an important step toward enabling accurate device-free indoor localization in a residential house. In our system, we simply attached passive tags on the walls to enable antennas to capture signals when a resident moving in the room, and installed sensors on the electronic appliances, which is considered not being very practical. However, we can mount readers and antennas on the ceiling, and embed passive tags into wall decorations. Also, with the development of IoT, it is a standard configuration for smart-homes to monitor working conditions of domestic appliances. *HOI-Loc* will be more practical and enable valuable applications with the prevalence of smart-homes in the near future.

Aside from accurately knowing the resident's indoor locations, understanding what the resident is doing is also one of the core functionalities in our living-assistive system. In the next chapter, we will intensively explore how to achieve a high-accuracy device-free human activity recognition based on the same passive RFID hardware.

Chapter 6

Device-free RFID-based Human Activity Recognition

Activity recognition is a fundamental research topic for a wide range of important applications such as remote health monitoring, fall detection, assistive-living system. It is essential for those applications to understand what a user is doing or attempting to do. Many of the existing techniques on activity recognition rely heavily on people's involvement such as wearing battery-powered sensors, which might not be practical in real-world situations (*e.g.,* people may forget to wear sensors). Over the last few years, *device-free* activity recognition has gained a considerable momentum and several solutions have been proposed. In this chapter, we propose a device-free activity monitoring approach using an array of low cost, passive RFID tags. Activity recognition is achieved by learning how the Received Signal Strength Indicator from the pure passive RFID tag array is distributed when a person performs different activities. We systematically examine the impact of tag configurations on performance of activity recognition and propose techniques for determining the optimal subset of RFID tags in the array, which is often missing in the most existing approaches. We further propose to infer activity changes via Dirichlet process Gaussian Mixture Model (DPGMM) based Hidden Markov Model, which effectively captures the nature of the

uncertainty caused by signal strength varieties. We run a pilot study and evaluate the performance with 12 orientation-sensitive activities and a series of activity change sequences. We conduct extensive experiments in two cases: in a lab environment and at a residential home. The experimental results demonstrate that our proposed approach can distinguish a series of orientation sensitive activities with high accuracy in both environments. The experimental results also show that our RFID-based device-free approach offers a good overall performance and has the potential to support the assisted living of elderly people.

6.1 Introduction

Human Activity Recognition (HAR) is one of the core functionalities for a living-assistive system. For instance, by monitoring the activities of a person with dementia, we could track how completely and consistently her daily routines are performed and determine when the person needs assistance. It is also regarded as an important and essential infrastructure for a wide range of applications such as safety surveillance [158, 159], ambient assisted living [160–162], and remote health monitoring and intervention.

To date, many research works on recognize human activities have been emerged. Computer vision based technique is one of main directions, but unfortunately, such solutions demand high computational cost for machine interpretation. In addition, the performance of such vision-based approaches depends strongly on the lighting conditions (*e.g.,* hard to monitor sleep postures at night), and cameras are generally considered to be intrusive to user's privacy. With the development of sensor, RFID, and wireless sensor network technologies, sensor-based HAR has gained growing attentions in the last several years. Inertial sensors are the most frequently used wearable sensors for human activity recognition [163–165, 58, 166, 57, 167]. Although sensor-based HAR systems can better address those issues in computer vision based techniques such as privacy intrusion and poor performance under darkness, most works still requires people to wear the inertial sensors or other electronic

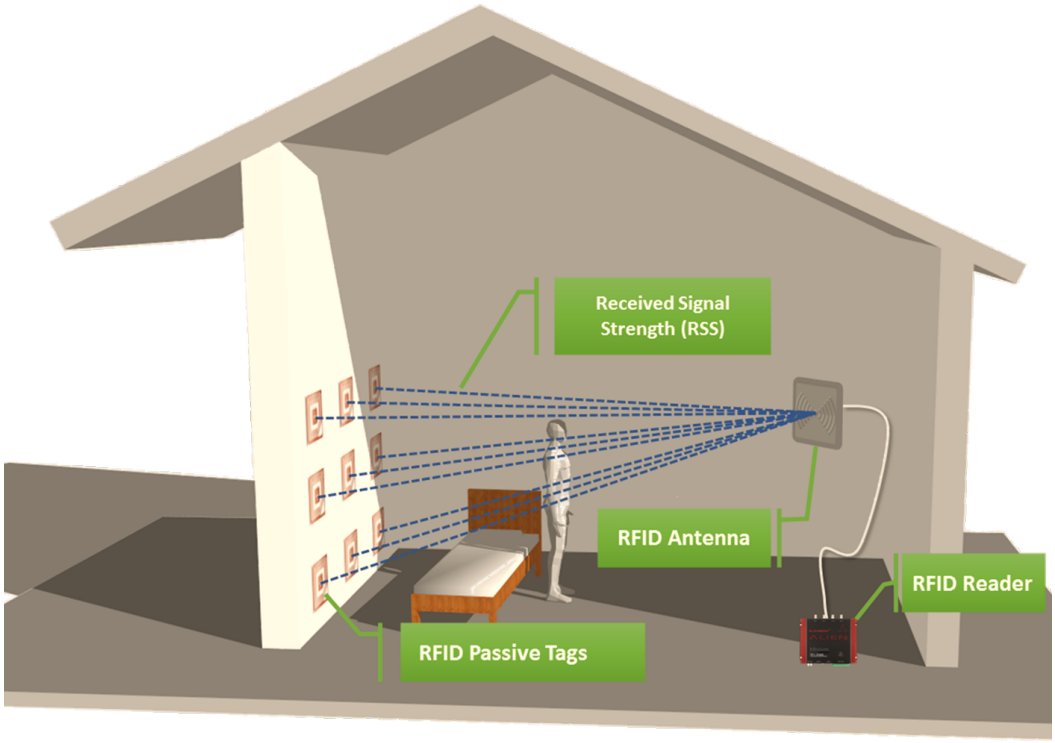


Fig. 6.1 Proposed lightweight setup: a person performs different activities between the wall deployed with an RFID array and an RFID antenna. The activities can be recognized by analyzing the corresponding sensing data collected by the RFID reader.

devices. RFID tags. Consequently, those systems unavoidably *i)* are not always practical, particularly for monitoring elderly persons with cognitive disabilities who usually forget to wear, and *ii)* need users' cooperation and regular maintenance (*e.g.,* battery changes).

To overcome the aforementioned issues, RF (Radio Frequency) based device-free activity recognition has been increasingly popular. Those systems generally deploy RF sensors in environments and analyze the fluctuations of the received signal strength (RSS) induced by the movements of human bodies to recognize the activities [60, 168, 169, 62, 33]. For example, in [61], the authors set up a sensor array to learn informative features from fluctuation of collected signals when people move to different locations or perform different actions. The HAR solution proposed in this chapter also belongs to this technical category, which is built upon a set of lightweight, cost-effective and maintenance-free passive RFID tag-array.

Different to other RF-based HAR approaches, the designed RFID-based system have three unique advantages: *i)* it requires no maintenance such as battery replacement or recharging; *ii)* it is cheap and each passive tag only costs 5 cents; and *iii)* it is convenient for deployment in a residential environment since a passive tag only weights several grams and has a tiny size (*i.e.,* $5cm \times 1cm$). Fig. 6.1 illustrates the basic hardware setup of our HAR system.

Moreover, our HAR system is capable of recognizing fine-grained daily activities such as distinguishing orientations of activities. Activity orientation identification can be valuable by combining with the layout of the place in practice, especially for a living assistive system. For instance, if we know that a table is on the left side of an elderly person, based on the layout, if we can detect the orientation of the person's fall, it is possible to identify how severe the fall would be (*e.g.,* she may hit the table if falling to her left).

In addition to activity monitoring, considering cost and size of deployment in reality, it is desirable to find a minimal set of tags and sensors without loss of performance accuracy. To meet this requirement, we further systematically study the optimal number of RFID tags deployed in the RFID tag array, which is commonly missing in many existing works. A common intuition is that more RFID tags offer better performance in activity recognition. However, research findings show that passive RFID tags can cause some unpredictable effects, *e.g.,* significant signal loss or fading if two tags are put in a certain distance [170]. Some researchers have been exploring the optimal tag configuration. A recent study by Wagner *et al.* [171] investigate the optimal tag placement to alleviate inaccuracy caused by the variability of RSS.

To this end, before HAR, we first search an optimal placement of passive tags, examining and eliminating the redundant correlations to find minimal set of tags so that achieve accurate and discriminative activity recognition performance. Second, to evaluate the performance of our passive RFID system handling highly dynamic variations of RSSI during activity transitions, we propose a Dirichlet Process Gaussian Mixture Model (DPGMM) with Hidden

Markov Model to detect a sequence of different activities (*e.g.,* from sitting to standing to falling). In a nutshell, our main contributions are summarized as follows.

- We address orientation-sensitive activity recognition problem using an array of pure passive RFID tags. Our approach is light-weight, low-cost, and unobtrusive in the sense that only passive RFID tags are deployed. The proposed HAR relaxes the requirement that users need to wear devices (sensors or transceivers) for activity monitoring.

- We examine a series of tag selection techniques including F statistics, relief F, random forest, multinomial logistic regression, to identify and eliminate redundant tags, which not only determines the optimal settings of tag array in terms of performance, but also paves a way to deploy tag arrays in larger scale environments with lower cost and less computational demand. On top of it, we further propose to integrate Dirichlet process Gaussian mixture model with hidden Markov model for an effective activity recognition.

- We conduct extensive experiments to validate our proposed approach. The experimental results demonstrate the feasibility of the proposed approach. In particular, the accuracy of steady activity classification based on our approach achieves over 98% in both lab and a real-world residential environment.

The rest of this chapter is organized as follows. In Sec. 6.2, we discuss the applications that can benefit from our approach, and formulate our problem based on some key observations. We describe our proposed approach in Sec. 6.3. In Sec. 6.4, we report the experimental results. We wrap up the chapter in Sec. 6.5 with conclusion and some future research discussions.

6.2 Background

In this section, we first discuss several representative applications that can benefit from our device-free activity recognition approach. We then formally define the targeted problems.

6.2.1 Application Scenarios

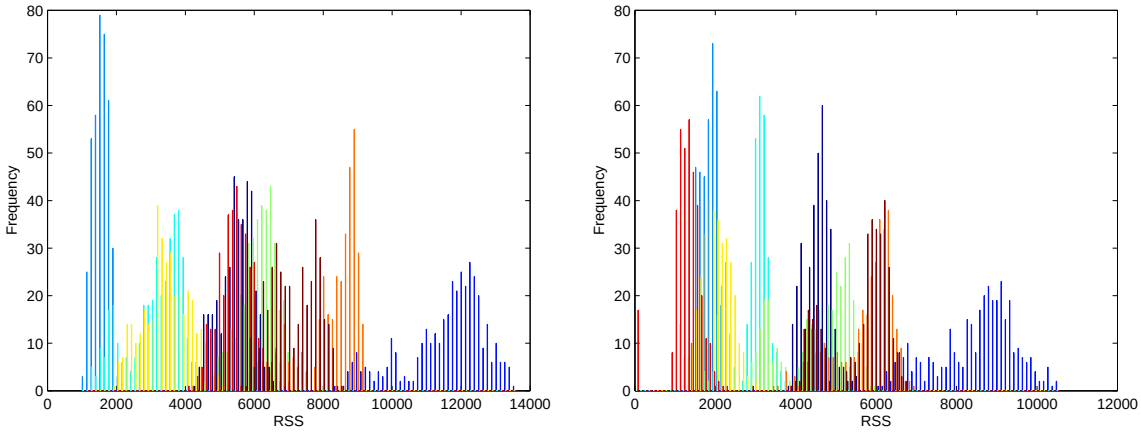
The HAR system we develop in this chapter can potentially be applied to posture monitoring and activity recognition in general, particularly for elderly people or people with cognitive impairment. Here we showcase several examples of its practical applications.

Fall Detection

With the great progress of medical technologies, many developed countries are facing the issue of the *aging society* where there will be a lower proportion of people of working age available to provide the necessary levels of care to elderly people. Meanwhile, the problem of huge nursing cost has a big impact to aged care. The demand for developing home surveillance systems is rising and such systems help old people stay at their own homes longer and safer. Falls are the leading cause of fatal injuries for people aged 65 and above. By monitoring the activities of an elderly, we could detect the likely falls (*e.g.,* getting out of bed) and issue an alert timely. Obviously, it is not practical to require the senior people to carry device all the time.

Ambulatory Monitoring

Activity recognition and monitoring are critical in medical area, *e.g.,* ambulatory monitoring, because physiological responses, such as changes in heart rate or blood pressure, may result from changes in body position and physical activities. Continuous monitoring and automatic detection of subtle behavioral changes can be very valuable for physicians and caregivers to estimate the physical well-being of a person.



(a) (b)

Fig. 6.2 (a) Histogram of RSSI from activity *sit leaning left*; (b) Histogram of RSSI from activity *sit leaning right*

Sleep Monitoring

Sleep activity recognition is crucial for elderly people as sleep disorders can be associated with some particular disease such as restless leg syndrome and diabetes. Device-free activity monitoring is an improvement over camera-based activity monitoring, because the latter suffers from privacy issues and poor performance at low light levels.

6.2.2 Observations and Problem Formulation

Figure 6.1 depicts the hardware setting of our system, where a tag-array containing several passive RFID tags is deployed on the wall of a bedroom and a RFID antenna is placed on the other side, facing these tags with a certain angle for better capturing RSSIs. When a person performs different activities in the room, our HAR intends to decode her activities from the collected RSSIs. However, it is well known that RSSI signal exhibits highly uncertain and complicate fluctuations in an indoor environments due to the signal reflection, diffraction and scattering, *etc.*. And those factors are often affected by the propagation environment, the tagged object properties, and human movements in the signal coverage area. So it is

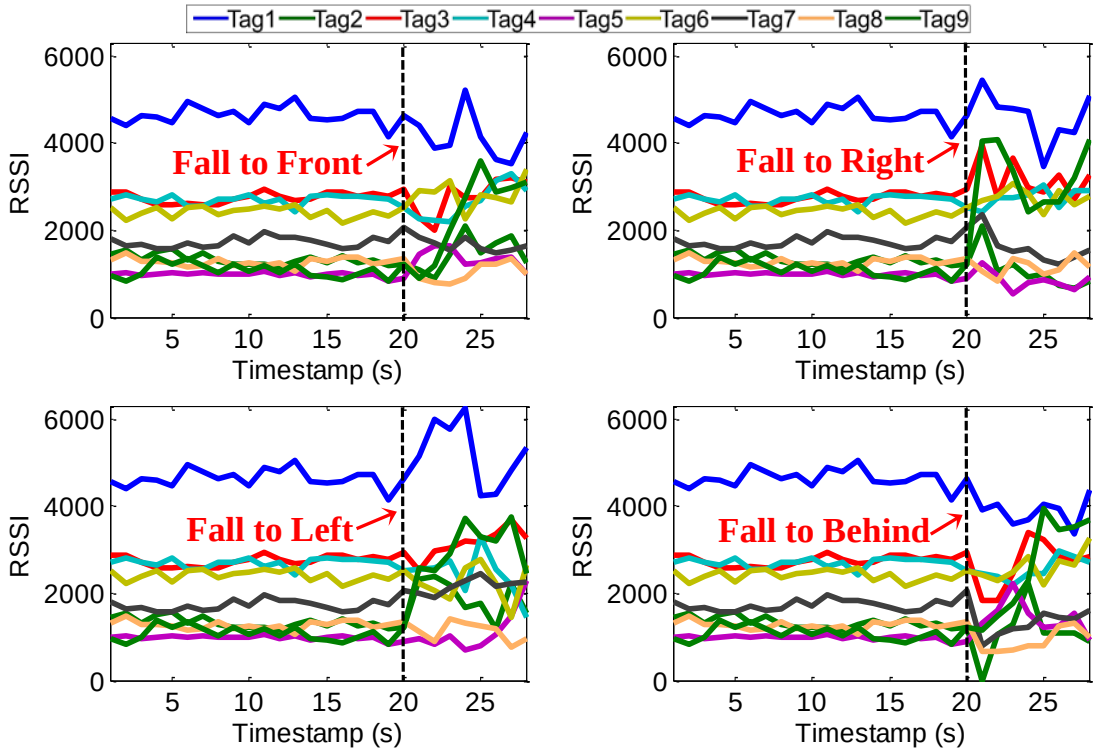


Fig. 6.3 RSSIs from 9-tag array for a fall with different orientations

difficult to quantify the relation of RSSI signals and human activities by the conventional signal propagation model.

However, on the other side, the variations of RSSIs potentially allow us to distinguish different activities if our HAR system can correctly “learn" those patterns. As Figure 6.2 shows, RSSIs for different activities depict distinctive distribution patterns in terms of histogram. Even for a same activity with different orientations, the RSSIs still show distinguishable fluctuation patterns, shown by Figure 6.3. of RSSIs show distinctive changing patterns for a fall activity with different orientations. Based the above observations, we believe that radio frequency signals of passive RFID tags embody certain patterns for different orientation activities and activity transitions, which can be further exploited for our activity recognition task. We therefore formulate our problems in this work as follows.

Let $\mathscr{O} \subset \mathbb{R}^d$ (d is the number of tags) be the domain of observable RSSI $\mathbf{o}$ and $\mathscr{L} \in \{1,...,K\} \subset \mathbb{R}$ be the domain of output activity label l. Suppose we have n RSSI and activity

label pairs $\{(\mathbf{o}_i, l_i)|\mathbf{o}_i \in \mathscr{O}, l_i \in \mathscr{L}, i = 1, ..., n\}$. The training dataset would be represented as:

$$\mathbf{O} = [\mathbf{o}_1, ..., \mathbf{o}_n] \in \mathbb{R}^{d \times n}, \quad \mathbf{l} = [l_1, ..., l_n]^T \in \mathbb{R}^n \tag{6.1}$$

In this chapter, we aim to answer the following two questions.

Problem 6 (Tag Deployment) *Given an RFID tag array, what is the optimal deployment setting of tags to achieve the best performance while using as minimal tags as possible?*

Problem 7 (Activity Recognition) *Given the RSSI observations, how can we accurately recognize a user's activity?*

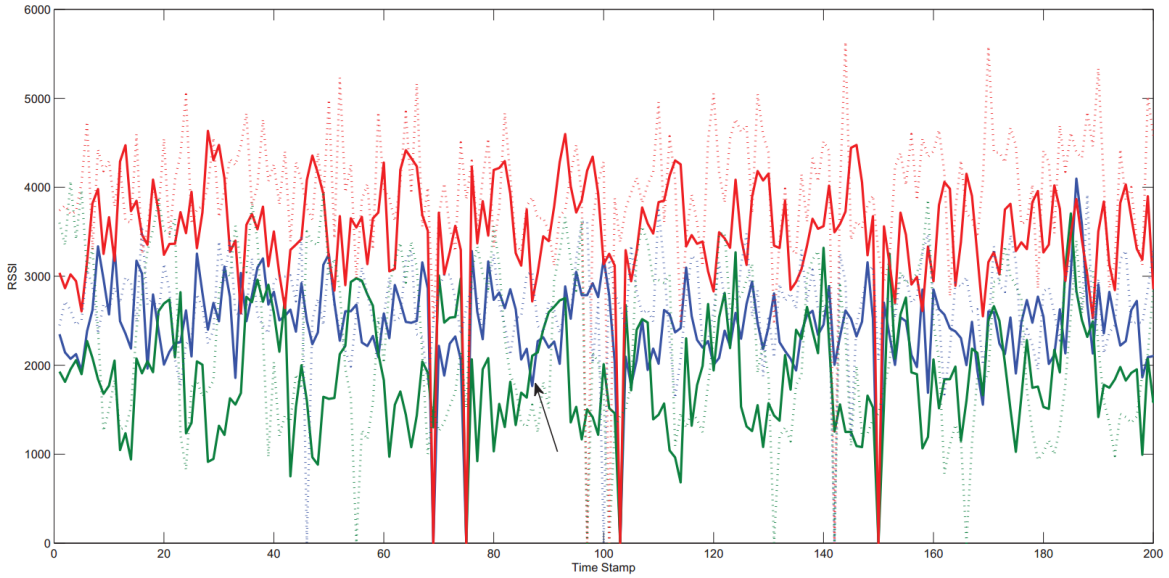
6.3 The Proposed Approach

In this section, we first present tag selection solutions to reduce the unpredictable effects and the number of passive tags. Then we introduce the technical details on human activity recognition.

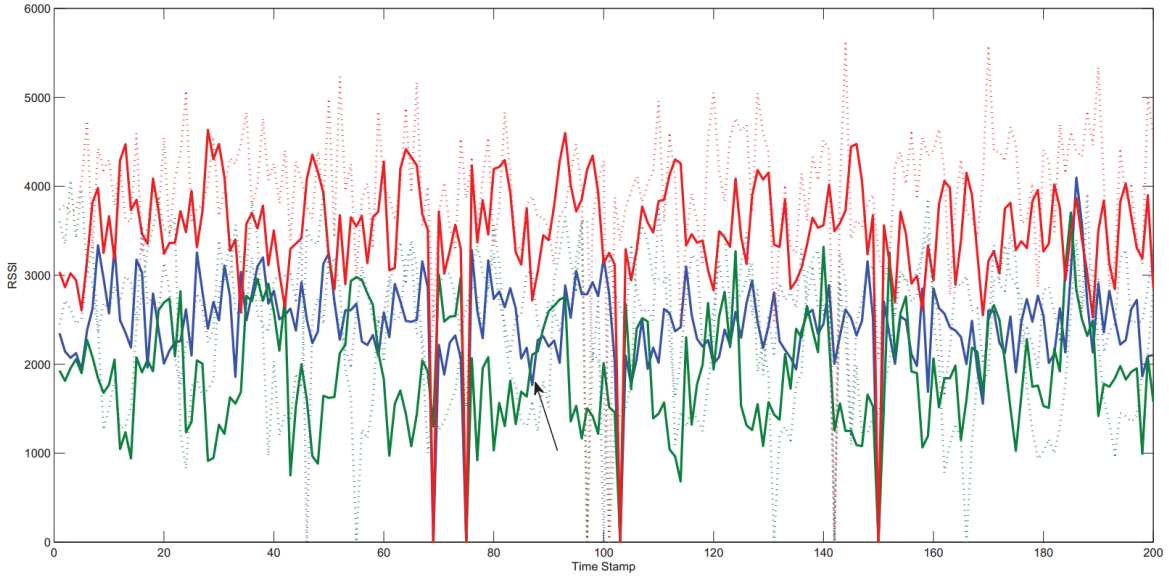
6.3.1 Tag Deployment

To find the optimal tag deployment, the first essential challenge for us is how to set up tags in an indoor setting to obtain the best performance. We first describe some intuitions of tag placement in this work. There are two main reasons why we place tags as an array:

- We conduct empirical studies on different forms of placing tags, such as arranging tags as a single line on the wall. According to our results, single-line tag placement is capable of capturing signal variations, but it may fail to detect fine-grained body movements, such as sitting leaning right or left. Furthermore, it is also not sensitive to capture the signal variations caused by subjects with different heights.



(a)



(b)

Fig. 6.4 Illustration of RSSI fluctuations of falling right and falling left: RSSIs of tag 1, tag 2 and tag 3 (top) and RSSIs of tag 7, tag 8 and tag 9.

- To achieve better accuracy and higher sensitivity, we increase single-line tag placement to multiple lines, eventually forming an *array*. Different lines correspond to different parts of human body. For instance, the upper line of tags would be expected to reflect the variations from upper human body like waving arms or shaking head, and the

middle line of tags would be more sensitive to movements of torso, and the bottom line of tags are supposed to have more response to lower body movements such as falling. In this way, we may perform more robust activity recognition with the collected *full spectrum* of RSSI variations.

As shown in Figure 6.4, the top one shows the RSSI fluctuations of three tags (tag 1, tag 2 and tag 3 are placed as a single line shown in Figure 13) and the lines indicate RSSI variations of falling right, and dash lines indicate RSSI variations of falling left. The bottom one shows the RSSI fluctuations of three tags (tag 7, tag 8 and tag 9 are arranged as a single line) and the solid lines indicate RSSI variations of sitting right, and dash lines indicate RSSI variations of sitting left. We can observe the fluctuations of tag 1, tag 2 and tag 3 are not quite helpful for reflecting different orientation falls as they do not show significant difference, on the other hand, tag 7, tag 8 and tag 9 can distinguish falling right and falling left better. The reason lies in fall action happens on the lower body, the lower location of tag 7, tag 8 and tag 9 can be more sensitive to such actions compared with tag 1, tag 2 and tag 3 in upper location. To capture the RSSI variations in all aspects, we use multiple lined up tags forming a tag array in this work. Existing works such as [61, 172] also show that placing sensors as an array can realize activity recognition with good accuracy.

For Problem 6 *Tag Deployment*, the second challenge is how many tags should be used to form a tag-array in order to obtain the best performance while using a less number of tags. In particular, we intend to answer the following question: how many tags should be actually deployed to reach the optimal trade-off between performance and set up cost of tag array, in other words, more is better or less is more?

We first explore the correlations between tags while activities are performed. As shown in Figure 6.5 (a), the RSSI values of tag 1 and tag 2 are highly correlated with each other, which means tag 1 and tag 2 are redundant in identifying activities. Figure 6.5 (b) illustrates that RSSI values of tag 1 and tag 9 successfully divide the RSSI data space from the series

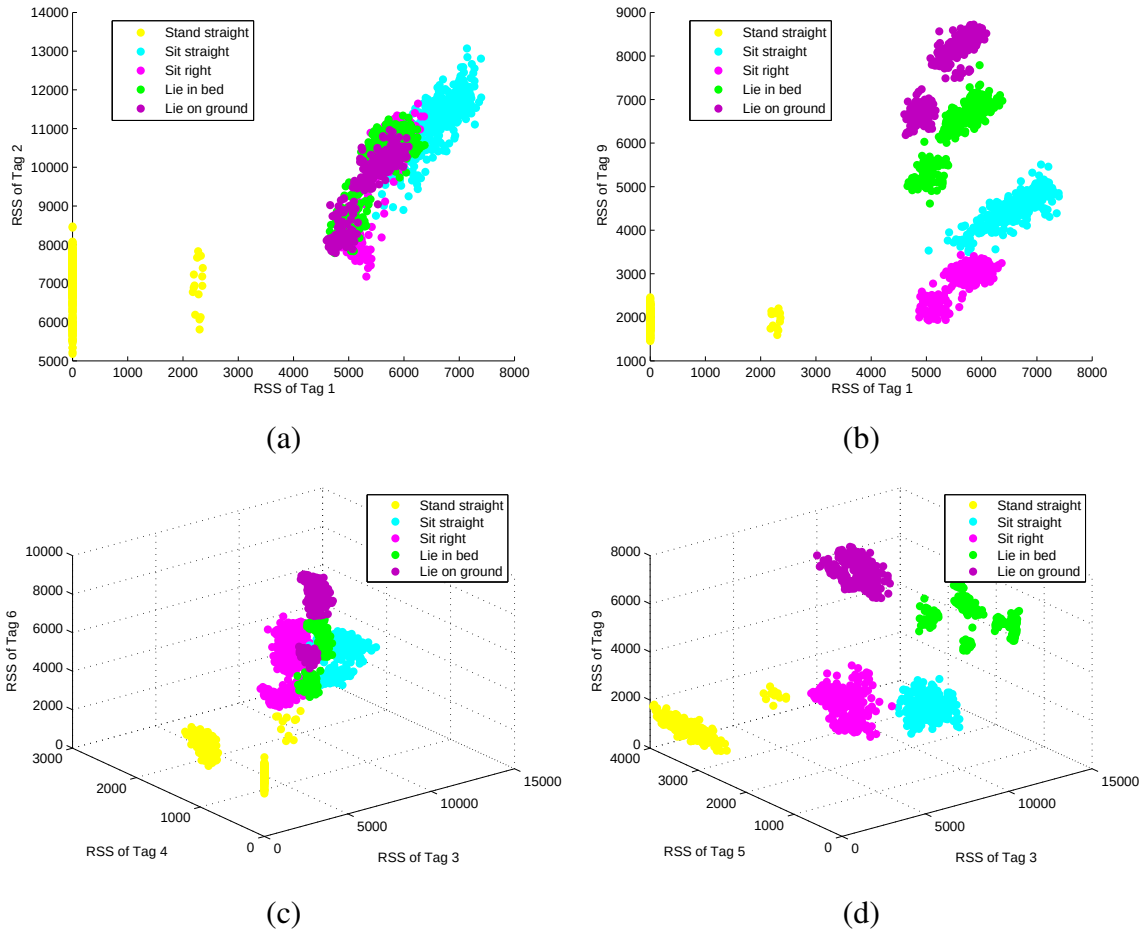


Fig. 6.5 Illustrative examples of tag correlations

of activities like *stand straight*, *sit straight* and *lie in bed*. We also examine the redundant correlations among three tags. For example, Figure 6.5 (d) shows that the RSSI values of tag 3, tag 5 and tag 9 can distinguish the listed activities, while the RSSI values of tag 3, tag 4 and tag 6 are highly correlated, as shown in Figure 6.5 (c). From the above observations, to eliminate the redundancy and select discriminative tags, we introduce a series of techniques to select a salient subset of tags to determine the optimal tag array configuration in our HAR system.

F-Statistics

It is to measure the discrimination of multiple sets of real numbers and is calculated using:

$$F_i = \frac{\sum_{j=1}^{l}(\bar{\mathbf{o}}_i^j - \bar{\mathbf{o}}_i)^2}{\sum_{j=1}^{l}\frac{1}{n_j-1}\sum_{k=1}^{n_j}(\mathbf{o}_{k,i}^j - \bar{\mathbf{o}}_i^j)^2} \tag{6.2}$$

where l is the number of activity classes, n_j is the number of samples in j^{th} activity class. $\bar{\mathbf{o}}_i$ denotes the mean value of tag i in the training dataset. $\bar{\mathbf{o}}_i^j$ is the mean value of i^{th} tag in the j^{th} activity class. The numerator indicates the discrimination between positive and negative sets, and the denominator indicates the one within each of the two sets. The larger the F-score is, the more likely this tag is discriminative in activity recognition.

Relief F

This technique estimates the relevance of features according to how well their values distinguish between the data points of the same and different activity classes that are close each other. It computes a weight for each tag to quantify its merit. This weight is updated for the RSSI samples presented in each activity class, according to the evaluation function:

$$\begin{aligned} w_i = w_i + &\sum_{j\in\mathscr{L}, j\neq l(\mathbf{o}_i)} \frac{P(l_j)}{1-P(l_j)} |\mathbf{o}_i - nearmiss_i^j(\mathbf{o}_i)| \\ &- |\mathbf{o}_i - nearhit_i(\mathbf{o}_i)| \end{aligned} \tag{6.3}$$

where l is the number of activity classes. $nearmiss_j(\mathbf{o}_i)$ and $nearhit_i(\mathbf{o}_i)$ denote the nearest RSSI samples to $\mathbf{o}_i$ from the same and different activity classes respectively.

Random Forest

Random forest (RF) is a classification method [173, 174], which also provides feature importance. Its basic idea is that a forest contains many decision trees, each of which is

constructed by instances with randomly sampled RSSIs. The prediction is made by a majority vote of decision trees. To obtain tag importance, we split the training sets into two parts. By training the first part and predicting the second, we obtain an accuracy value. For the j^{th} tag, we randomly shuffle its values in the second set and obtain another accuracy value. The difference between the two accuracy values can indicate the importance of the j^{th} tag.

Multinomial Logistic Regression with ℓ_1 Regularization

ℓ_1 regularization uses a penalty term that shapes the sum of the absolute values of parameters to be small, which usually leads to a sparse parameter vector. In this work, we integrate the ℓ_1 regularization into linear classifier in the objective term. Given our multi-class activity recognition problem, we combine the ℓ_1 regularization with multinomial logistic regression, which models the conditional probability $P_{\mathbf{w}}(l_j = \mp 1|\mathbf{o})$. The prime problem with ℓ_1 regularization can be calculated by optimizing the log likelihood:

$$\min_{\mathbf{w}} \sum_{k=1}^{K} ||\mathbf{w}_k||_1 - \sum_{i=1}^{n}\sum_{k=1}^{K} l_{ik}\mathbf{w}_k^T\mathbf{o}_i + \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} exp(\mathbf{w}_k^T\mathbf{o}_i)\right) \tag{6.4}$$

RFID tags can then be selected by considering the obtained weight vector $\mathbf{w}$.

Least Square with ℓ_1 Regularization

It can be represented as:

$$\min_{\mathbf{w}\in\mathbb{R}^d} \frac{1}{2}||\mathbf{l} - \mathbf{O}\mathbf{w}||_2^2 + \lambda||\mathbf{w}||_1 \tag{6.5}$$

where $\mathbf{l} = \{l_1, ..., l_n\}$ is the activity labels of training RSSI samples, $\mathbf{O} = \{\mathbf{o}_1, ..., \mathbf{o}_n\}$ is all training RSSI samples, $\mathbf{w} = [w_1, ..., w_d]^T$ denotes the regression coefficients, w_i corresponds to the regression coefficient of the i^{th} tag, λ is the regularization parameter. Same as the multinomial case, $||\mathbf{w}||_1$ regularization tends to produce a sparse solution (*i.e.,* the regression coefficients of irrelevant tags are or close to zero), which indicates the importance of each

tag. We also study the $\ell_{2,1}$ regularization, which is formulated as:

$$\min_{\mathbf{w}\in\mathbb{R}^d} \frac{1}{2}||\mathbf{l}-\mathbf{O}\mathbf{w}||_2^2+\lambda||\mathbf{w}||_{2,1} \tag{6.6}$$

where $||\mathbf{w}||_{2,1}=\sum_{i=1}^{n}\sqrt{\sum_{j=1}^{K} w_{ij}^2}$.

After performing the selection process, all tags are ordered based on their importance and a subset of tags is selected based on a user-defined threshold of top-N ($N<d$) tags.

6.3.2 Steady Activity Recognition

We adopt SVM (support vector machine) with linear kernel to perform steady activity classification. SVM aims at finding the decision boundary via maximizing the distance from the closet sample to the boundary hyperplane. When there are limited training data available, SVM usually outperforms the traditional parameter estimation methods which are based on the Law of Large Numbers. This is mainly due to the fact that SVM benefits from the structural risk minimization principle and the avoidance of overfitting by its soft margin. For activity recognition, SVM classifies activities based on the fact that the smaller the distance between two RSSI samples, the higher probability they belongs to a same activity. SVM method works directly with RSSI using the kernel functions. The topology implicit in sets of RSSI and the activities can be exploited in the construction of possibly non-Euclidean function spaces that are useful for activity estimation. Given the sequence of training RSSI and corresponding activity labels $\mathbf{O}=\{(\mathbf{o}_1,l_i),...,(\mathbf{o}_n,l_n)\}$, where $\mathbf{o}\in\mathbb{R}^d$ and $l\in\{1,...,K\}$, the objective function can be formulated as:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \mathbf{w}^T\mathbf{w}+C\sum_{i=1}^{n}\xi_i \\ &\text{s.t. } l_i(\mathbf{w}^T\phi(\mathbf{o}_i)+b)\geq 1-\xi_i, i=1,2,...,n \\ &\xi_i\geq 0, i=1,2,...,n \end{aligned} \tag{6.7}$$

where ξ_i is a slack variable, C is the penalty of error term, $K(\mathbf{o}_i, \mathbf{o}_j) = \phi(\mathbf{o}_i)^T \phi(\mathbf{o}_j)$ is the kernel function.

The prime problem of optimization in Equation 6.7 can be converted to solve its duality using Lagrange multiplier. Thus, Equation 6.7 can be reformulated as:

$$\begin{aligned} &L(\mathbf{w}, b, \xi, \alpha, \mu) = \\ &\mathbf{w}^T\mathbf{w} + C\sum_{i}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(l_i(\mathbf{w}\mathbf{o}_i + b) - 1 + \xi_i) + \sum_{i=1}^{n}\mu_i\xi_i \end{aligned} \tag{6.8}$$

where $\alpha = (\alpha_1, ..., \alpha_n)^T$ and $\mu = (\mu_1, ..., \mu_n)^T$ is the Lagrange multipliers. To solve Equation 6.8, we can maximize the minimization of duality as:

$$\max_{\alpha,\mu} \min_{\mathbf{w},b,\xi} L(\mathbf{w}, b, \xi, \alpha, \mu) \tag{6.9}$$

After the model is learned, we can recognize the activity class for a given testing RSSI $\mathbf{o}^*$.

6.3.3 Activity Sequence Recognition

To accurately recognize a sequence of activities (*e.g.,* from straight standing to falling to the ground, then to standing up, finally to walking alway) , our goal is to determine the conditional probability $P(l_k|\mathbf{o}_i)$ given a new coming sample $\mathbf{o}$. We propose a HMM based approach, which has shown a powerful performance in modeling activity sequences. In particular, given observation sequences of RSSI $\mathbf{O} = \{\mathbf{o}_1, ..., \mathbf{o}_T\}$, and activity states denoted by activity label sequence $\mathbf{l} = \{l_1, ..., l_T\}$, the HMM models the sequence of observable RSSI $\mathscr{O} = \{\mathbf{o}_1, ..., \mathbf{o}_T\}$ by assuming that there is an underlying sequence of different activities $\mathbf{l} = \{l_1, ..., l_T\}$ drawn from a finite activity set. In our model, each observation $\mathbf{o}_t$ is the RSSI vector, and each state l_t is the activity label (*e.g., sit*, *lie in bed*).

HMM makes two assumptions: *i)* each activity performed at t only depends on its immediate previous activity at time $t-1$, and *ii)* each observable RSSI $\mathbf{o}_t$ only depends on

the current performed activity l_t, which are formulated respectively as:

$$p(l_t|l_{t-1},\mathbf{o}_{t-1},...,l_1,\mathbf{o}_1) = p(l_t|l_{t-1}), t = 1,2,...,T \tag{6.10}$$

$$p(\mathbf{o}_t|l_T,\mathbf{o}_T,...,l_{t+1},\mathbf{o}_{t+1},...,l_1,\mathbf{o}_1) = p(\mathbf{o}_t|l_t) \tag{6.11}$$

With the assumptions, we can model the joint probability of activity sequence $\mathbf{l}$ and observable RSSI sequence $\mathbf{O}$ as:

$$p(\mathbf{l},\mathbf{O}) = \prod_{t=1}^{T} p(l_t|l_{t-1})p(\mathbf{o}_t|l_t) \tag{6.12}$$

where $p(l_t|l_{t-1})$ is the transition probability indicating the likelihood the subject changes from activity l_{t-1} to activity l_t, which is defined by considering the predefined activity transitions applications. For example, people can transit from *sit* to *stand*, but can not transit from *lie in bed* to *fall on ground* directly, whilst they can transit from *lie in bed* to *sit* then to *fall on ground*. We denote the state transition probability distribution as $\mathscr{A} = \{a_{ij}\}$:

$$a_{ij} = p(l_{t+1} = l_j|l_t = l_i) \tag{6.13}$$

On the other hand, $p(\mathbf{o}_t|l_t)$ denotes the observation distribution drawn by different activities. We assume RSSI distribution generated by each activity as a Gaussian mixture model, which is a weighted sum of m component Gaussian densities. It can be defined as $\mathscr{B} = \{b_t(i)\}$:

$$
\begin{aligned}
b_t(i) &= p(\mathbf{o}_t|l_t = l_i) \\
&= \sum_{m=1}^{M_i} \pi_{i,m} N(\mathbf{o}_t, \mu_{i,m}, \Sigma_{i,m}) \\
&= p(\mathbf{o}_t|\mu_m, \Sigma_m) = \frac{1}{(2\pi)^{D/2}\Sigma_m^{1/2}} exp \\
&\left(-\frac{1}{2}(\mathbf{o}_t - \mu_m)^T \Sigma_m^{-1} (\mathbf{o}_t - \mu_m) \right)
\end{aligned}
\tag{6.14}
$$

where $\mathbf{o}$ is d dimensional continuous RSSI observations (d is the number of tags in the deployment) and π is the mixture weights and $p(\mathbf{o}_t|\mu_m, \Sigma_m)$ is the component Gaussian distribution.

The traditional GMM learning process with Expectation-Maximization (EM) limits to determination of how many gaussian components in the GMM. We adopt the Dirichlet process Gaussian mixture model (DPGMM) in observation probability distribution in this work. It uses Dirichlet process as a prior over the distribution of the parameters and there is no need to explicitly declare the number of components. The approximate inference algorithm uses a truncated distribution with a fixed maximum number of components, but almost always the number of components actually used depends on the data [175]. We use two-dimensional RSSI from our dataset to show the advantage of DPGMM over GMM (in Figure 7). GMM with EM learning splits Gaussian components arbitrarily, for example, the two clusters are eventually divided into five clusters in some convergences. Thus it does not reach a good fit even we use AIC (Akaike Information Criterion) [1] as model selection criteria, while the Dirichlet Process GMM model effectively only uses as many as needed for a good fit without defining number of guassian components, it can accurately nail down two clusters and converges to a good fit automatically in this case.

Our goal of detecting activities in the context of HMM is: given a sequence of RSSI observations $\mathbf{o}_1, ..., \mathbf{o}_T$, what is the most likely sequence of activities to produce such ob-

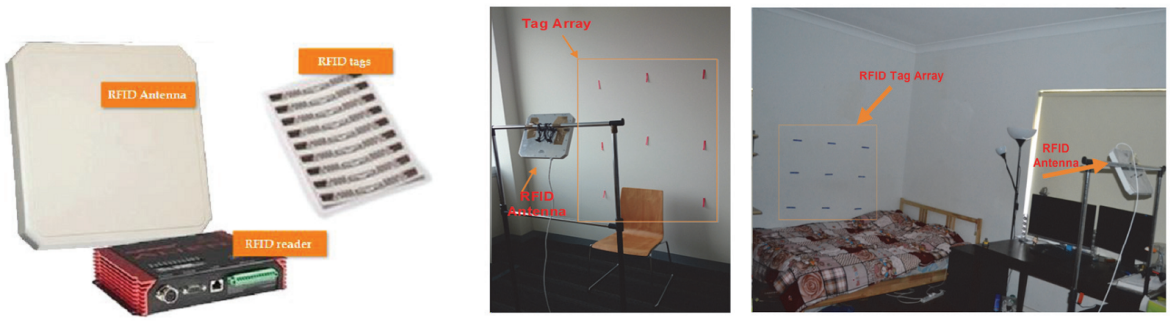


Fig. 6.6 RFID tags/reader/antenna (left); Lab setting (middle) and Bedroom setting (right)

servations? We adopt the Viterbi algorithm to find the most likely state sequence in HMM. Formally, given a continuous sequence of RSSI observations $\mathbf{o}_1 ... \mathbf{o}_T$ and learned HMM (shown in Equation 6.12), we aim to find the most likely activity sequence $l_1 ... l_T$:

$$\delta_t(j) = \max_{l_1 ... l_{t-1}} p(l_t = j, l_{t-1}, ... l_1, \mathbf{o}_t, ..., \mathbf{o}_1 | \mathscr{A}, \mathscr{B}) \tag{6.15}$$

where $\mathscr{A}$ and $\mathscr{B}$ can be calculated from Equation 6.13 and Equation 6.14. By induction, we can have:

$$\begin{aligned} \delta_1(j) &= b_1(\mathbf{o}_j) \\ \delta_{t+1}(j) &= \max_{1 \le j \le N} \delta_{t-1}(j) a_{ij} b_t(\mathbf{o}_{t+1}), j = 1, ..., K \end{aligned} \tag{6.16}$$

6.4 Experiments

This section reports our experimental studies in both lab and realworld residential environments.

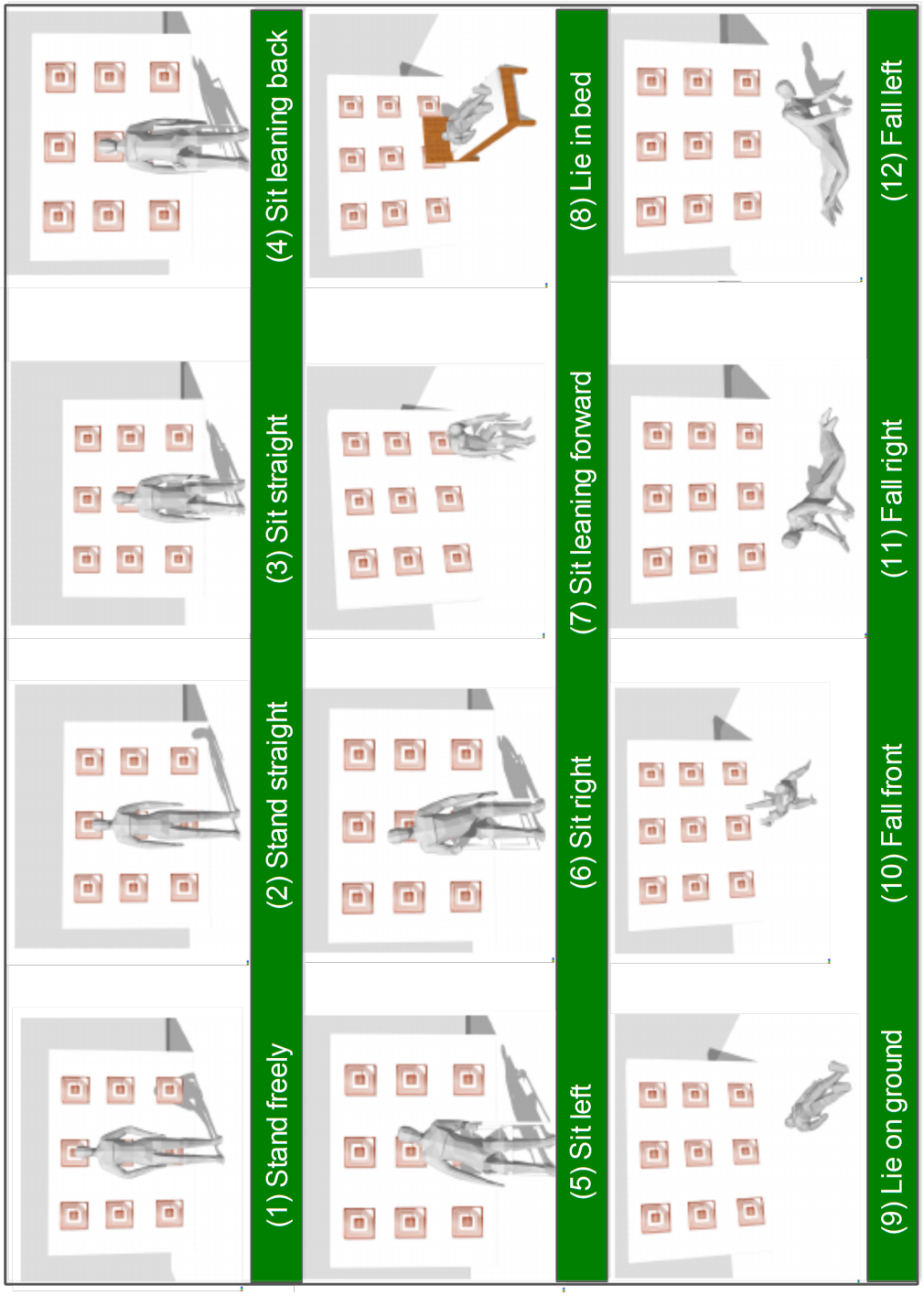


Fig. 6.7 Predefined orientation-sensitive activities

6.4.1 Experimental Settings

Hardware Setup

We used one Alien 9900+ RFID reader, one circular antenna and Squig inlay passive RFID tags in our experiment (see Figure 6.6). The original tag array containing nine tags was placed at a 3×3 grid points on the wall where each grid is roughly $0.58m \times 0.58m$. We call this wall the *active testing area*. The antenna was arranged in $\approx 1.3m$ height facing the active testing area in $\approx 70^{\circ}$ (as shown in Figure 6.6). The subject performed different predefined activities between the wall and the antenna, and the corresponding sequence of RSSI were collected.

Sampling Rate

Passive RFID tags tend to be noisy even in a lab environment. For example, one challenge in existing RFID systems is false negative readings, caused by missed detections (*i.e.,* a tag is in the antenna's reading range, but not detected). In addition, RSSI data is much sensitive to environments, *e.g.,* some disturbance from environment can cause RSSI fluctuations. Appropriate sampling rates can reduce the aforementioned problems. Too small sampling rates make our method more sensitive to the noise of RFID readings, while too big sampling rates blur the inter-class activity boundaries. In our implementation, we collected the continuous RSSI data streams at the sampling rate of 2Hz.

Data Acquisition

We ran a pilot study to evaluate the performance of our HAR system. For collecting the training dataset, we conducted a series of experiments in which a subject entered the active testing area and performed various pre-arranged activities, including *standing*, *sitting*, *lying on ground*, *lying in bed*, and *falling*, *etc.*. Three subjects (two males and one female) participated in the experiment and each performed the set of 12 fine-grained activities

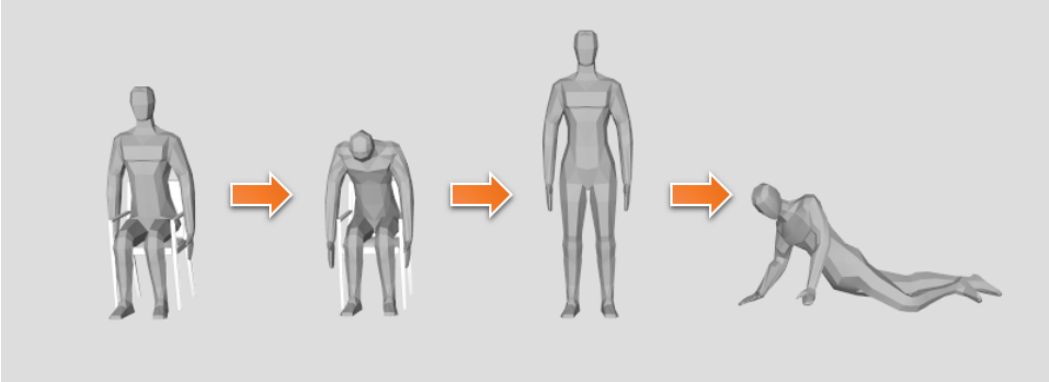


Fig. 6.8 An example of activity changes

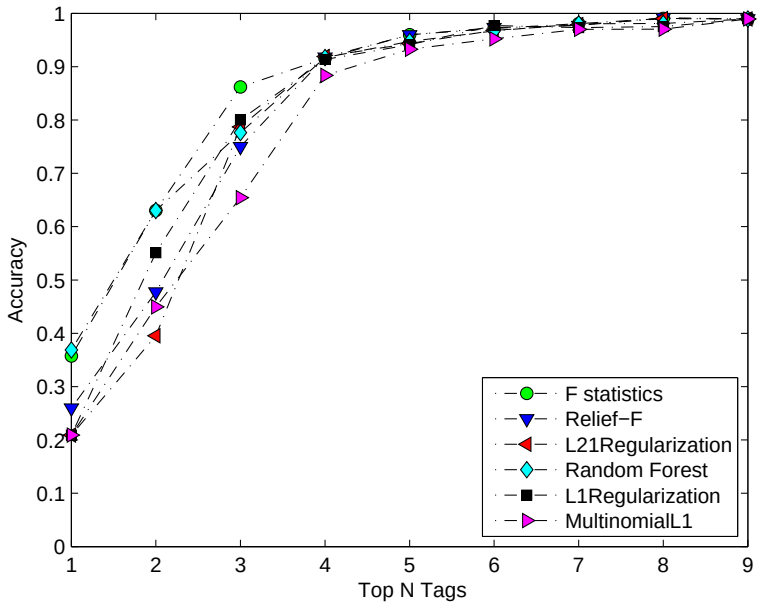
(Shown as Figure 6.7). The subjects also performed different predefined activity sequences for evaluating activity sequence recognition.

The task of the steady activity classification is to model how the signal strengths are distributed when the subject performs different activities. Each subject stands in the active testing area which is between the antenna and the wall deployed with passive RFID tags. We first measured the RSSI values for all tags when the testing area is empty. Then each subject stood in the area and performed the 12 predefined activities.

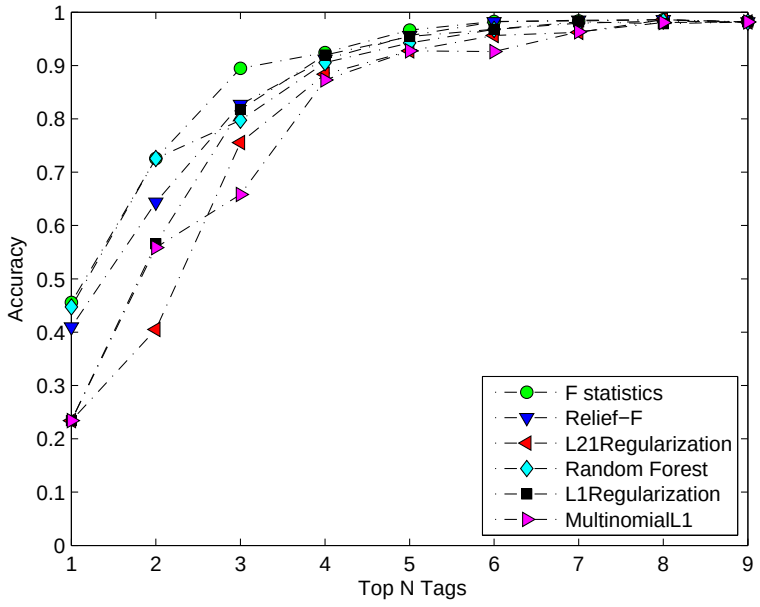
For collecting the activity-sequence dataset, we designed eight different activity sequences to simulate the activity sequences in real world (see Figure 6.8) and collected them using two strategies. In the first strategy, the subject performed and held each activity for 30 seconds and then performed next activity in the order as predefined in the sequence. In the second strategy, the subject performed and held each activity for 60 seconds and then performed the next activity in the order as predefined in the sequence.

6.4.2 Results

To evaluate the effectiveness of the proposed tag selection, we adopted a person-dependent 10-fold cross-validation strategy. For the person-dependent evaluation, we use partial samples of a subject for testing and use the remaining samples of the same participant for training.



(a) Lab



(b) Bedroom

Fig. 6.9 Activity classification comparison with Top N tag selection in (a) lab and (b) bedroom environments

Impact on Tag Selection

To evaluate the impact of tag selection, we sorted the tag importance calculated from six selection approaches (Section 6.3.1) in descend order, and compared the recognition accuracy using SVM with linear kernel by choosing top N tags (N is from 1 to 9 (full set)). Figure 6.9 (a) shows the results comparison over top N tags in the lab environment, and Figure 6.10 (b) shows the results comparison over top N tags in the bedroom environment.

In both cases, the performance are influenced by the selected tags. Activity classification accuracy does not improve much after top 5 selected tags using all selection criteria, and reaches the best point (99.18%) with Relief-F selection when top-7 tags are selected compared with 99.04% without tag selection (full set of tags). In the bedroom case, the impact of tag selection on performance is more obvious. The accuracy is the best when only seven tags are selected, and the performance even slightly drops when more tags are added. From the results, we can see that the tag selection does improve the overall performance in both lab and bedroom environments by distinguishing the salient tags, only subset of intuitively placed tags shows their usefulness and discrimination via implicating the intra-person variability on different activities. The rest of tags degrade the overall performance due to failing to capture the inter-class and intra-class variability. Figure 6.10 shows an example of optimal tag deployments from our experiments.

Steady Activity Classification

To study the feasibility of our approach and sensitivity to size of training data after selecting tags, we further evaluated the activity classification with varying training ratios in terms of tag selection and no tag selection. As shown in Figure 6.11, our approach performs well even only with 10% the training size, the accuracy reaches over 90% in both cases. The accuracy increases with larger training sizes. However, when the training size is around $60\% \sim 70\%$,

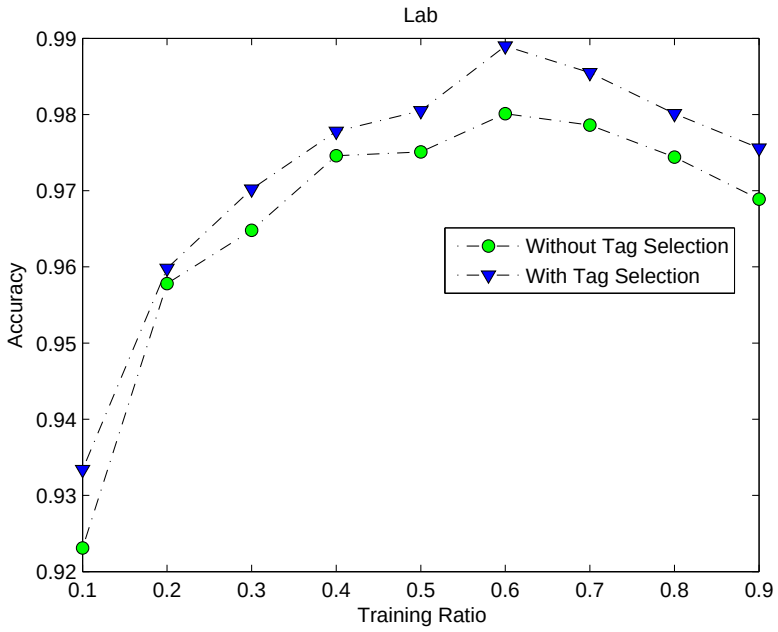
Fig. 6.10 Selected tags

the accuracy begin to decrease. We can see that the overall accuracy is consistently better with the tag selection strategy compared with the case without tag selection.

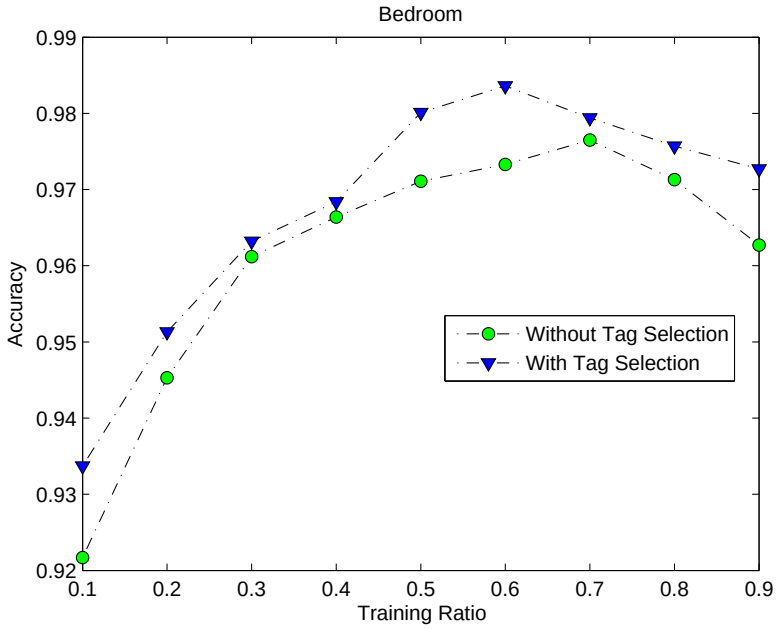
We look closely at the results of the confusion matrices in both the lab and the bedroom cases with the selected subset of tags and 10% training ratio given in Table 6.1 and Table 6.2. Generally, only a few samples of activities, *i.e., stand in free style* (with ID of 1) and *stand straight* (with ID of 2), misclassified in the lab environment. In the in bedroom environment, activities *fall right* (with ID of 11) and *fall left* (with ID of 12) are misclassified, whilst they can be accurately classified in the lab environment. It should be noted that the performance on classifying orientation-sensitive activities still reaches over 98% in the bedroom environment.

Activity Sequence Recognition

To evaluate the accuracy of recognizing sequential activities, we performed activity classification over a series of activity changes and measured how accurately our approach can recognize a activity given new coming RSSI values, as well as how timely our approach can recognize the activity.



(a) Lab



(b) Bedroom

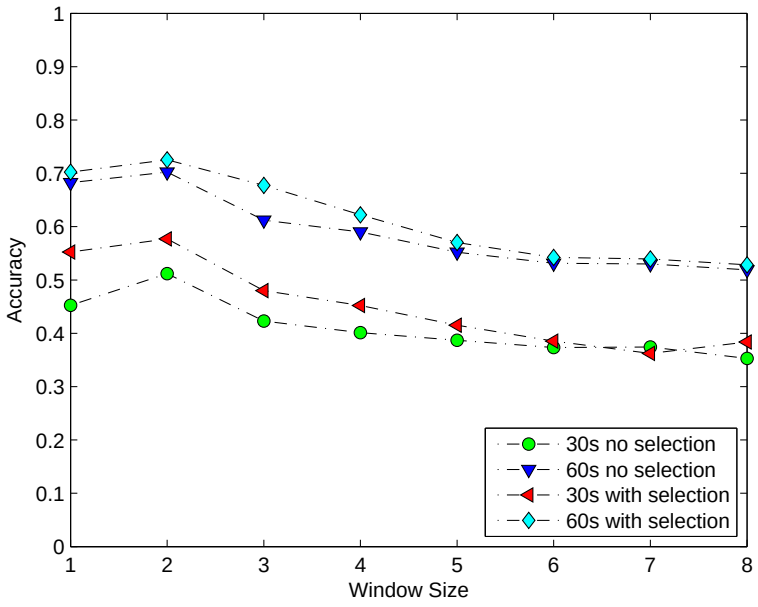
Fig. 6.11 Accuracy comparison with tag selection and without tag selection using different training sizes: (a) lab and (b) bedroom

Table 6.1 Confusion matrix with tag selection in lab

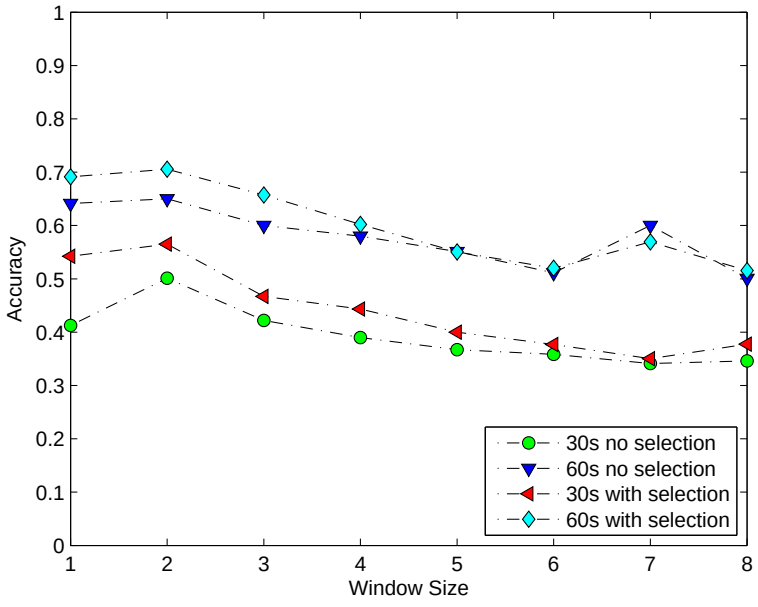
ID	1	2	3	4	5	6	7	8	9	10	11	12
1	35	2										
2	1	34										
3			39									
4				36								
5					36							
6						36						
7							39					
8								36				
9									37			
10										36		
11											36	
12												36

Table 6.2 Confusion matrix with tag selection in bedroom

ID	1	2	3	4	5	6	7	8	9	10	11	12
1	35	1										
2	1	34										
3			36									
4				36								
5					36							
6						36						
7							39					
8								36				
9									37			
10										36		
11											35	2
12											1	35



(a)



(b)

Fig. 6.12 Performance comparison on different window sizes using 30s and 60s strategies without tag selection and with tag selection (a) lab and (b) bedroom

Passive RFID tags are highly sensitive to disturbance, especially when activities continuously change. The RSSI fluctuation result from activity transition exhibits some uncertainty. To cope with the impact of this disturbance, we adopted a *forward calibration* mechanism to calibrate the RSSI streams before detecting activity changes [176]. We used a sliding time averaging window to smooth the RSSIs. The calibrated RSSI stream $\mathbf{o}'_t$ at time t can be calculated as:

$$\hat{\mathbf{o}}'_t = \frac{\sum_{i=t}^{t+|w|-1} \mathbf{o}'_i}{|w|} \tag{6.17}$$

where $|w|$ is the window size.

Intuitively, a larger window size may break the consistency of RSSI samples from one activity, while a smaller window size may not provide the best information for the activity transition process. To determine the best window size in this work, we evaluated the performance of both lab and bedroom settings with and without tag selection strategy by varying the window size. Figure 6.12 shows the results. We can see that the performance does not consistently improve when increasing window size, instead, when the window size is 2, the performance in both settings reached the best result. We further compared the performance in terms of different duration an activity is held. Figure 6.12 shows the results under two durations (30 and 60 seconds) with and without tag selection. From the results, we can see that the longer the activity is held by the subject, the better accuracy can be achieved. The reason is that a longer activity holding time can eliminate both inter-class and intra-class variations, to which RSSI are especially sensitive in recognizing activities. We also can see that the performance using the tag selection strategy significantly outperforms the one without tag selection. The results from steady activity classification and activity transitions detection consistently indicate that an optimal subset of tags can more discriminatingly recognize activities compared with full set of tags. The subset tags have the dominant impact.

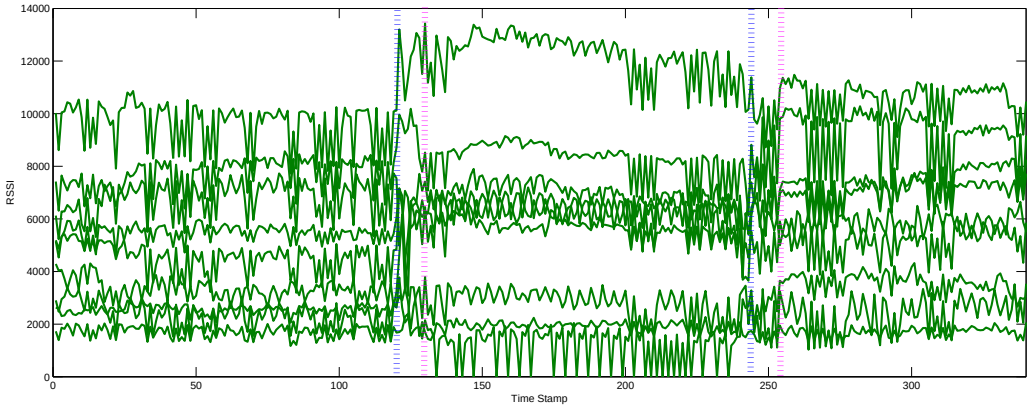


Fig. 6.13 Recognition latency: blue dot vertical line indicates the ground-truth time point of activity change, pink dot vertical line indicates the recognition time point detected by our proposed approach.

Recognition Delay

Fast recognition of sequential activities is critical, particularly for aged-care applications. For example, for fall detection, we can send an alert and notify caregivers as quickly as possible to offer medical assistance for elderly people when a fall occurs. So we conducted the experiments to test how quickly our proposed approach can identify a new activity when a user constantly changes her activities, in other words, the recognition delay.

The results from our experiments show that our system has 3.5 seconds recognition latency, which results from two main reasons. Firstly, our system evaluates subject's activities every 0.5 seconds using the latest 2 seconds of RSSI stream. In other words, if the current system time is at timestamp t, our system will produce the predicted activity in the $[t-2,t-1]$ seconds, and $[t-1,t]$ seconds is used to backtrack check if the predicted label complies with predefined rules. For instance, assume that the label is estimated as: *lying in bed* at $[t-2,t-1]$ interval, if the predicted label in interval $[t-1,t]$ is *nobody*, our system will determine the subject is still *lie in bed*. Secondly, the RSSI collector is programmed with a timer to poll the RSSI with a predefined order of transmission, and needs to take around 1

second to complete a new measurement with no workarounds. From Figure 6.13, we can clearly see our proposed method can promptly detect the activity changes with slight latency.

In summary, in this section, through our extensive experiments, we can see that the overall performance at home environment is a little bit lower than the lab environment (due to furnitures *etc.*). However, it still achieves over 98% accuracy for steady activity classification and 70% for the overall activity transition detection.

6.5 Conclusion

In this chapter, we proposed a device-free activity recognition system for elderly people, by exploiting low-cost passive RFID tags. We focus our study on tag configuration issues, especially tag placement and selection, for achieving the best trade-off between performance and cost. We systematically study these issues by using different configuration settings and applying various tag selection methods. We also propose a Dirichlet Process Gaussian Mixture Model with the Hidden Markov model to recognize different activities. Through our extensive experiments, our HAR system detects 12 orientation-sensitive activities, with an accuracy of 99% and 72% in terms of steady activity and activity sequence recognition in a lab environment respectively, and over 98% and 70% in a real-life home environment. We also demonstrate that deploying more tags does not necessarily improve the recognition performance, which actually decreases the accuracy under some circumstances.

Among all those human activities, falling down is one of the most dangerous actions that deserves our particular attention, especially for the elderly who live alone. In the next chapter, we will design a fine-grained fall detection system that can timely recognize falling events and distinguish the falling orientations.

Chapter 7

Fine-grained Device-free Fall Detection based on Passive RFID Tag Array

Falls are among the leading causes of hospitalization for the elderly and illness individuals. Considering that the elderly often live alone and receive only irregular visits, it is essential to develop such a system that can effectively detect a fall or abnormal activities. However, previous fall detection systems either require to wear sensors or are able to detect a fall but fail to provide fine-grained contextual information (*e.g.,* what is the person doing before falling, falling directions). In this chapter, we propose a device-free, fine-grained fall detection system based on pure passive Ultra-High Frequency (UHF) Radio-Frequency IDentification (RFID) tags, which not only is capable of sensing regular actions and fall events simultaneously, but also provide caregivers the contexts of fall orientations. We first augment the Angle-based Outlier Detection Method (ABOD) to classify normal actions (*e.g.,* standing, sitting, lying and walking) and detect a fall event. Once a fall event is detected, we first segment a fix-length RSSI data stream generated by the fall and then utilize Dynamic Time Warping (DTW) based k Nearest Neighbors (kNN) to distinguish the falling direction. The experimental results demonstrate that our proposed approach can distinguish the living status before fall happening, as well as the fall orientations with a high accuracy. The experiments also show

that our device-free, fine-grained fall detection system offers a good overall performance and has the potential to better support the assisted living of older people.

7.1 Introduction

Falls happen when human body suddenly changes from a normal living status (*e.g.,* sitting, standing, or walking) to the reclining without control [64], which often occur in a very short time without human attentions. Falls may cause moderate to severe injuries including hip fractures, head traumas, even more devastating consequences for the elderly. Based on the Centers for Disease Control and Prevention, one-third population of the elderly who aged 65 and older experience falls each year [177]. Researchers estimate that up to 50% of nursing home residents fall each year and more than 40% of them might fall more than once [177]. Moreover, studies have shown that the medical outcome of a fall is largely dependent on the response and rescue time[68]. The delay of medical treatment after a fall can increase the mortality risk in some clinical conditions, especially for those who live alone [178]. Thus, falls are a major health risk that diminishes the quality of life among the elderly people, strongly motivating the necessity of fall detection systems.

Over the past decades, fall detection (FD) and prevention have been an active research area with several proposed solutions. Both wearable sensor based (*e.g.,* inertial sensors [70], accelerometer [179, 66], specialized cane [68]) and smart-phone based [71, 72] fall detection techniques require the subject to be attached with sensors or phones, which might not be practical (*e.g.,* sensors lost/damaged, or forget to carry by the elderly with dementia). Vision based fall detection systems [74, 78, 77] employ activity classification algorithms on a series of images recorded by a video camera, which is usually regarded as being privacy invasive and causes uncomfortable feeling to the elderly. Vision-based systems also fail to work in dimmed or dark environments, where falls usually happen.

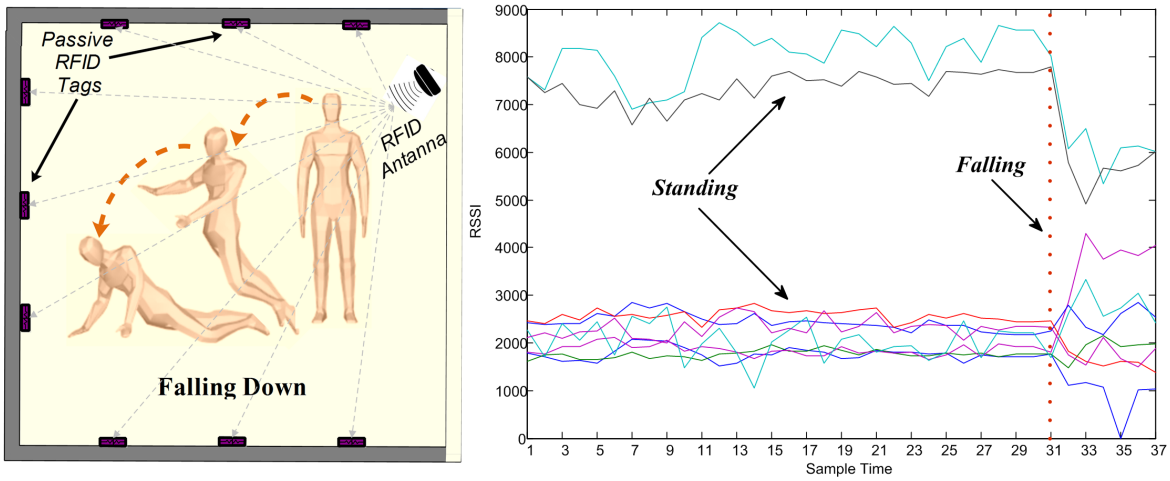


Fig. 7.1 RSSIs variation patterns when falls occur

Recently, some device-free techniques for fall detection have been proposed [81, 80, 83]. However, in most of these systems, some complicated or personalized devices (*e.g.,* pressure sensor, audio sensor, radar) are needed to be implanted in the environment, and then the variations of audio, pressure or microwave signals are used to infer a fall event. As a result, most of them can only sense whether a fall happened, but fail to provide more fine-grained information [63] that is valuable to caregivers. One of the fine-grained contexts is the status (*e.g.,* sitting, standing or walking) before a fall occurs. For example, when people fall down while standing or siting, some serious diseases possibly have happened such as cerebral haemorrhage or cardiopathy [177]. But when people are walking, the falling is possibly caused by knocking some obstacles. Another useful contextual information is the fall orientations (*e.g.,* fall to front, fall to back or fall to the right side), *e.g.,* falling to back may seriously damage the subject's head, while falling to the right side may more likely cause injuries to arms or legs.

Based on those motivations and with recent advances of passive RFID sensing technology, this chapter concentrates on the investigation that whether a device-free, fine-grained fall detection can be achieved without using any wearable device/sensor. Figure 7.1 briefly illustrates the mechanism of our fall detection system built on passive RFID tags. When the

subject falls from standing, the Received Signal Strength Indicators show different fluctuation patterns, indicating the potential for detecting a fall. Compared to other hardware platforms, passive RFID is cost-effective (passive tags cost several cents each) and practical (*e.g.,* no maintenance since no battery needed) [85]. More importantly, as far as we know, until now, there is no research work that explores how to purely utilize passive RFID hardware to achieve fine-grained fall detection.

In this chapter, we propose a device-free, fine-grained fall detection system called *TagFall*, which can not only timely detect a fall event but also accurately recognize regular daily living activities (*e.g.,* standing, sitting, lying and walking) before a fall happens, as well as distinguish the falling directions. To achieve such a fine-grained fall detection, our *TagFall* mainly consists of two detection phases. *i)* Detecting Normal Actions and Falls: we augment Angle-based Outlier Detection (ABOD) [180] method to mine the clustering patterns of RSSIs (generating by normal human actions) and detect an anomaly pattern (caused by falls) simultaneously; and *ii)* Detecting Fall Directions: once we detect a fall happened, we segment a fix-length data stream, which we use to calculate the Dynamic Time Warping (DTW) [181] distance with profiled data streams (known labels). So we can distinguish the fall directions by a majority vote of its k nearest neighbors based on the DTW distances. In summary, the core idea of this chapter is to mine the clustering patterns and change rules of RSSIs when the environment is affected by different human actions (*e.g.,* normal activities and falls with different orientations). The main contributions are listed as follows.

- We exploit the feasibility of using passive RFID tags to achieve unobstructive, fine-grained fall detection. To the best of our knowledge, this is the first work to leverage RSSI signals for device-free fall detection based on pure passive RFID tags.
- We propose a fine-grained fall detection pipeline, which not only can detect a fall event, but also be capable of offering the contextual information of the subject's status before falls occur and the falling directions.

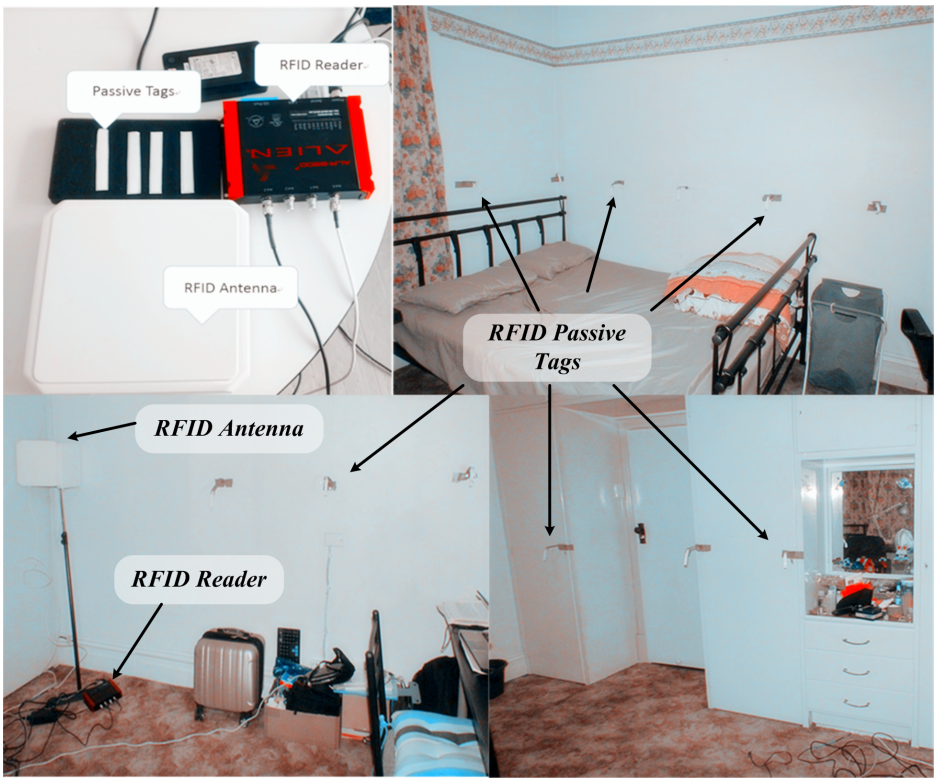


Fig. 7.2 Hardware Deployment

The rest of this chapter is organized as follows. Sec. 7.2 introduces the hardwares and intuitions of our system. We present our system architecture in Sec. 7.3 and propose our solutions in Sec. 7.4. Sec. 7.5 presents experimental results and analysis. Finally, Sec. 7.6 gives the discussion and Sec. 7.7 offers some concluding remarks.

7.2 Hardware and Intuitions

Figure 7.2 shows the system setup, including an Alien ALR-9900+ Enterprise RFID Reader ($20.3cm \times 17.8cm \times 4.1cm$), two-circular antennas ($20cm \times 20cm \times 3cm$), and squiggle Higgs-4 passive tags ($1cm \times 10cm$). The reader operates at 840-960MHz and supports UHF RFID standards such as ETSI EN 302 208-1. We set the sample rate as $0.5s$ and each

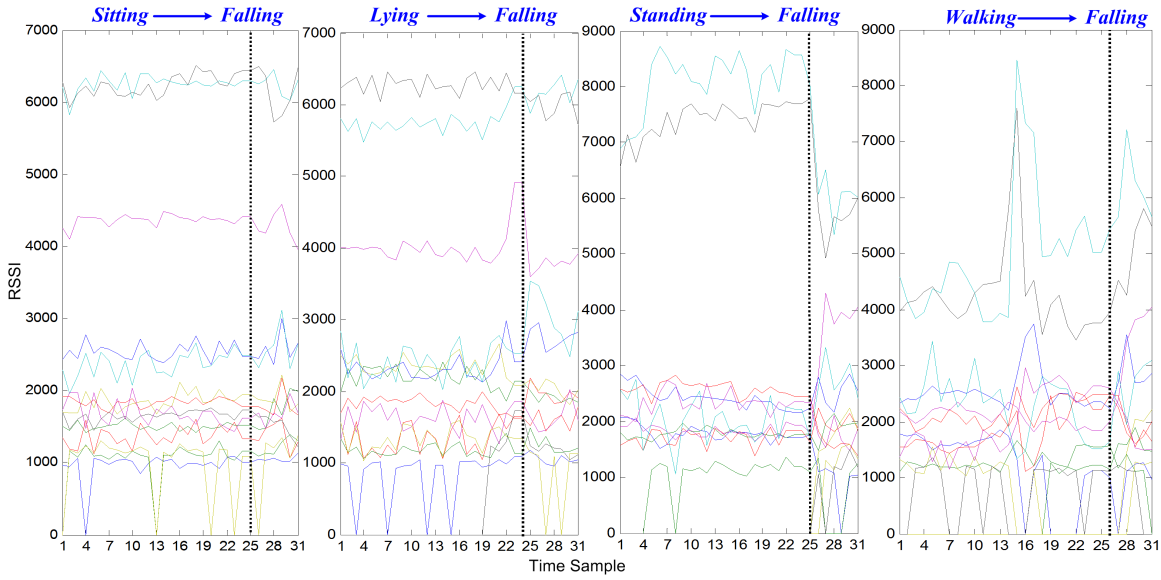


Fig. 7.3 RSSIs variation patterns when a subject falls from different status

tag reading contains a time-stamp, a tag ID, an antenna ID and an RSSI value, which is processed by a WINDOWS 7 PC with an I7-3537U 2.5GHz processor and 8G RAM.

Based on our preliminary experiments, we place the antenna 1.5m above the ground, facing tags with approximately 60°, and attach tags on the wall with an approximate $0.6m$ interval. When people perform activities in the testing area, an antenna can not guarantee to read all tags (interfered by human body), particularly for passive tags. To avoid this, we send an RSSI request to all tags within a sampling time. If we cannot receive RSSI readings of a certain tag, the RSSI value is manually set to 0. Thus, mathematically, for all time stamps, we have the RSSI vectors with the same dimensions. In our settings, the tag detection range can be up to 6 meters. It is worth mentioning that, the electromagnetic fields generated by the readers, in any case, remain lower than the limitation of thresholding value for humans based on the report [182], which means the hardware used have no health risk to subjects.

Based on the hardware, we first run a series of pilot experiments to validate our intuitions. Figure 7.3 demonstrates that when the subject in different living status, the RSSIs display different invariant patterns. When the subject falls down from normal status (*e.g.,* sitting,

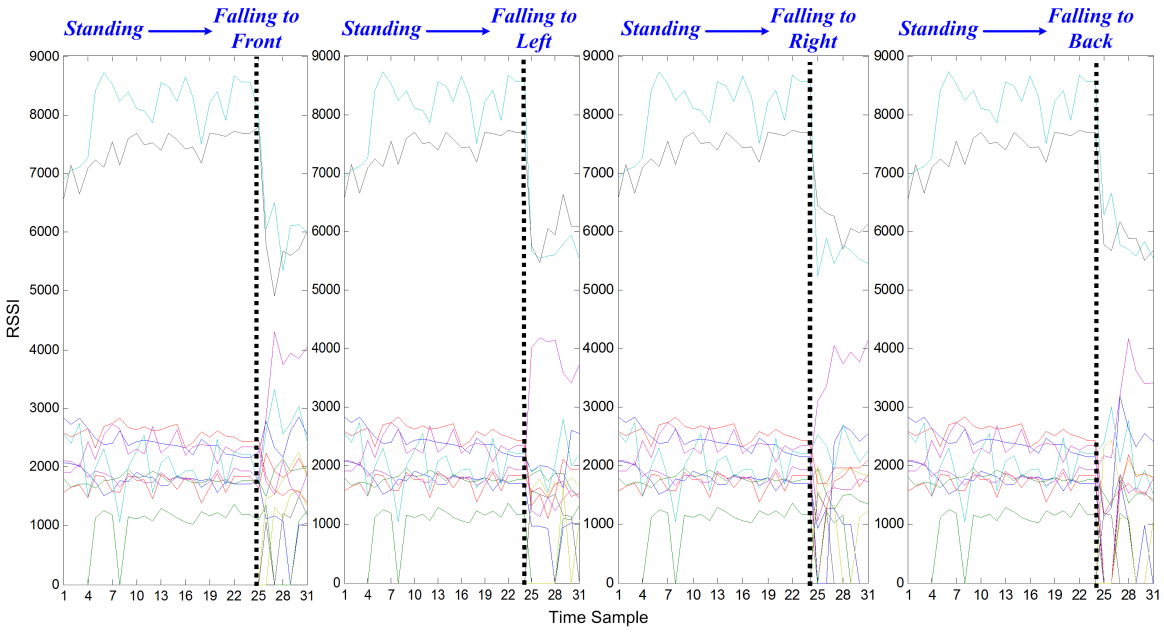


Fig. 7.4 RSSIs variation patterns when a subject falls to different directions from standing

lying, standing or walking), the RSSIs reflect some unique variations that are different from previous stable patterns. Underpinned by these observations, it is possible for us to utilize some supervised classification algorithms to distinguish resident's regular living actions, as well as adopt anomaly detection method to detect an abnormal event (*e.g.,* falls). In the next section, we will introduce how to tackle both problems by extending the traditional ABOD method.

Figure 7.4 shows that the measured RSSIs can reveal varied fluctuation patterns due to the subject's falling down to different directions (*e.g.,* front, left, right or back side) from standing, which allow us to utilize such underlying trends to recognize the falling orientations. Motivated by the observation, we could adopt stream data classification methods to classify resident's falling directions. In this chapter, we solve this problem by using DTW distance based *k*NN. Overall, our preliminary studies have shown the feasibility and potential of our TagFall system to achieve a fine-grained fall detection.

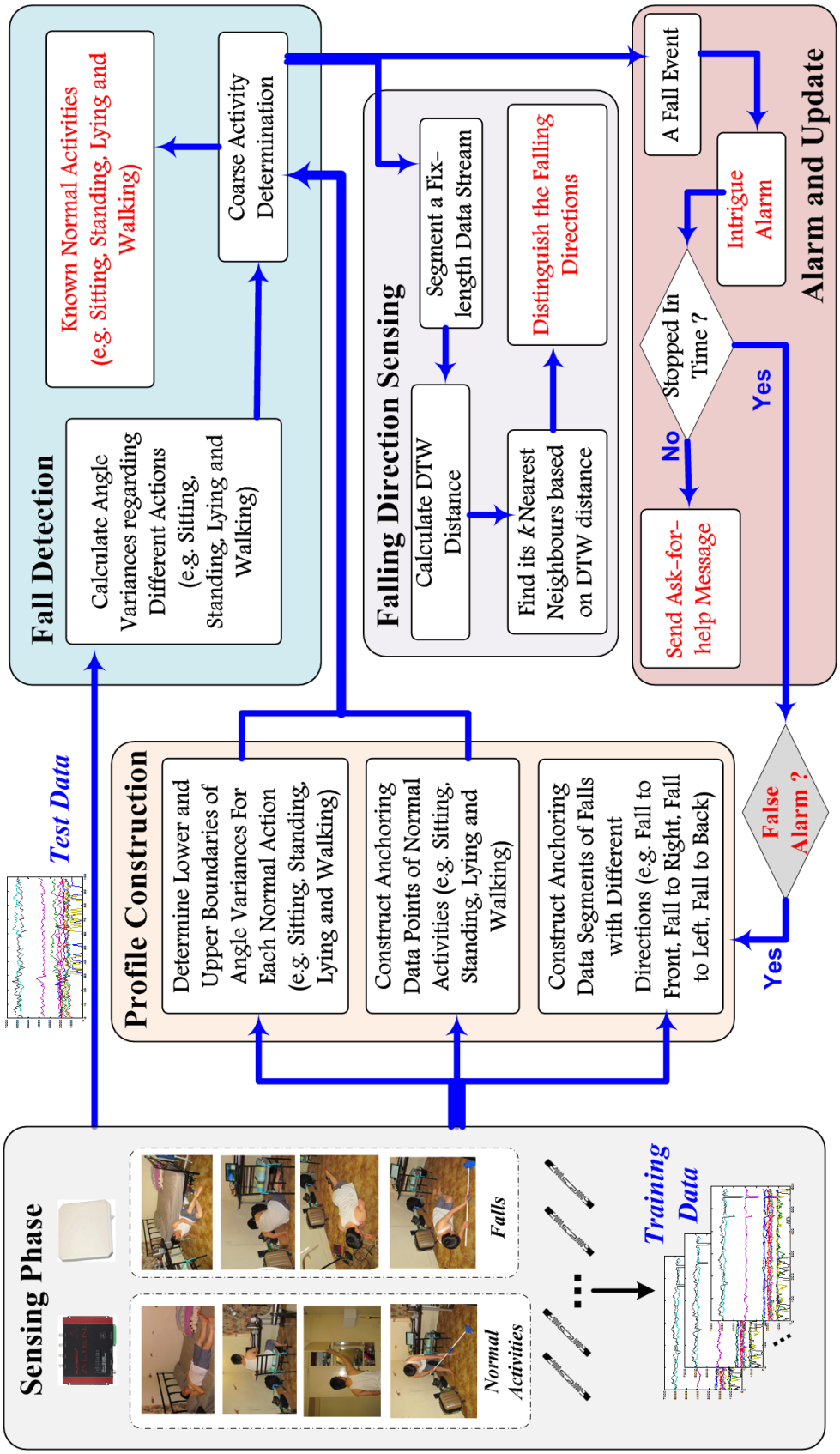


Fig. 7.5 System Architecture

7.3 System Architecture

Figure 7.5 shows an overview of proposed system. It consists of five main phases: the sensing phase, the profile construction phase, the fall detection phase, the falling direction sensing phase, and the altering phase.

7.3.1 Activity Sensing Phase

The antenna in the test area collects RSSI readings propagated by passive tags and then sends them to the reader, which delivers the data package including RSSI values, time stamp, antenna ID and tag ID to a desktop computer for further processing.

7.3.2 Profile Construction Phase

We first utilize slide average smoothing to filter the noises caused by temperature, humidity changes [85] and categorize daily activities into four categories (i.e., sitting, standing, lying and walking, shown by Figure 7.12). Then, we calculate the angle variances of vector pairs formed by same action category and decide the upper and lower boundary of variances, which contain most likely variances. In the meantime, we sample the most representative data point for each regular action category to speed up the later online angle variance calculation. We also collect segmented data streams generated by falls with various falling directions to build the anchoring data streams for the later DTW distance calculations.

7.3.3 Fall Detection Phase

We perform the same smoothing as Profile Constriction phase and then calculate the angle variances of vector pairs formed by an observed RSSI and profiling data points for each normal action category. Based on the calculated variances (i.e., 4 variances in our case) and learned variance bounds, we identify the target's current actions by judging whether the

variance lie in corresponding boundaries. If the variances are within the bounds of multiple action categories, we assign the activity label that has the most likely variance. When all four variances are beyond the bounds of known regular actions, the observed RSSI is regarded as an anomaly, which means the subject currently is experiencing a fall.

7.3.4 Falling Direction Sensing Phase

Once we detect a fall event, we first segment a data stream with the same length as the anchoring data streams. Then, we calculate the DTW distances between segmented data stream and all anchoring data streams. At last, we can distinguish the falling direction by a majority vote of its k nearest neighbors regarding the DTW distance.

7.3.5 Altering and Update Phase

In the meantime, we issue an alarm (*e.g.,* ring an alarm bell) when a fall event is detected. If the user does not timely stop the alarm, we send an ask-for-help SMS or call. Also, if the alarm is timely stopped but is a false alarm, we update the profiling data by adding the error-detected samples into right action category to enhance the detection performance.

7.4 Device-free Fine-grained Fall Detection

The key phases of our *TagFall* are how to efficiently distinguish the normal daily living actions and a fall event, and how to accurately classify the falling directions. In this section, we will introduce technical details on how these two problems are solved.

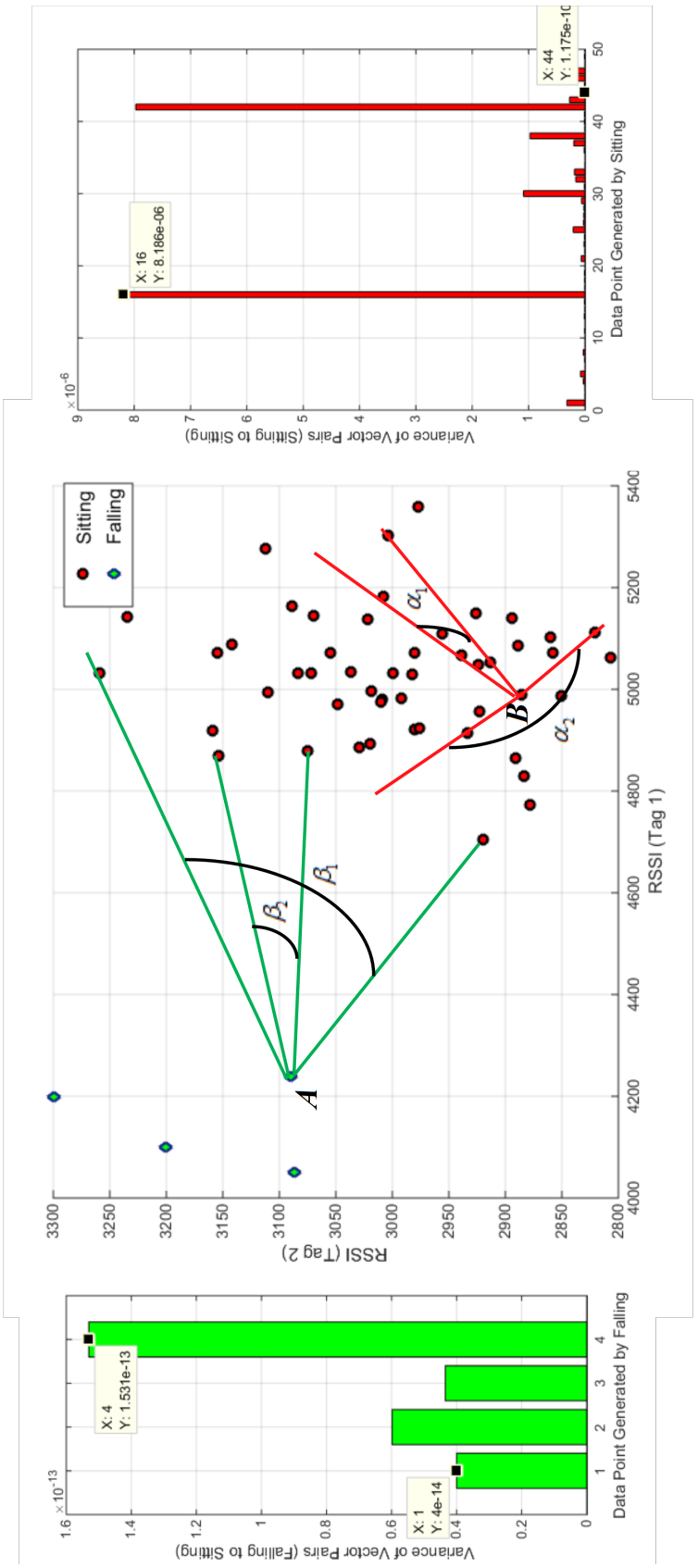


Fig. 7.6 Intuition of angle-based outlier detection

7.4.1 Fall Detection

One of the challenges in this chapter is to detect the anomalous patterns in RSSI signals. A fall involves a series intensive posture changes (*e.g.,* human postures sudden alter from standing, sitting or lying to the ground), which result in sudden, wide range fluctuation of RSSI patterns (see Figure 7.3 and 7.4). To tackle the challenge, we propose a p-partial Angle-based Outlier Detection that can identify p categories of human regular actions and isolate the anomaly patterns. Angle-based Outlier Detection is first proposed by Hans-Peter Kriegel et al. [180] for finding anomalous data points caused by a different responsible mechanism. Unlike purely distance-based approaches (*e.g.,* Local Outlier Factor [183]), ABOD does not rely on any parameter selection influencing the quality of achieved results. Here, we extend ABOD to do both classification and anomaly detection by mining the different patterns of angle variances paired by intra-action and inter-actions.

Figure 7.6 illustrates the basic intuition of our approach. For points (can be multi-dimension) generated by a same human activity, the angles between different vector pairs differ widely, which means a large angle variance (*e.g.,* angle α_1, α_2, the variance of angle paired by data points of sitting is ranged from $1.17 \times 10^{-10} \sim 8.18 \times 10^{-6}$). The angle variance of vector pairs generated by different human activities is smaller since most points are clustered in some directions (*e.g.,* angle β_1, β_2, the variance of angle paired from data points of falling to data points of sitting is ranged from $4 \times 10^{-14} \sim 1.53 \times 10^{-13}$). Therefore, we can classify different regular actions (easily collected, *e.g.,* standing, sitting and walking), and detect abnormal actions (difficultly obtained, *e.g.,* falling) by measuring the angle variances between a testing data point and the constructed profiling dataset. We first give the definition of Angle-based Outlier Factor (ABOF) [180] which measures the angle variance of a data point paired with other data points.

Definition 5 (ABOF) *Given a database $\mathscr{D} \in \mathbb{R}^d$, a point $A \in \mathscr{D}$, and a norm $||\,||$. The scalar product is denoted by $< .,. >$. For two points $B, C \in \mathscr{D}$, $\overline{BC}$ denotes the difference vector*

$C-B$. The angle-based outlier factor $ABOF(A)$ is the variance over the angles between the difference vectors of A to all pairs in D weighted by the distance of the points:

$$ABOF(A) = VAR_{B,C\in\mathscr{D}}\left(\frac{<\overline{AB},\overline{AC}>}{||\overline{AB}||^2\cdot||\overline{AC}||^2}\right) = \frac{\sum_{B\in\mathscr{D}}\sum_{C\in\mathscr{D}}\left(\frac{1}{||\overline{AB}||\cdot||\overline{AC}||}\cdot\frac{<\overline{AB},\overline{AC}>}{||\overline{AB}||^2\cdot||\overline{AC}||^2}\right)^2}{\sum_{B\in\mathscr{D}}\sum_{C\in\mathscr{D}}\frac{1}{||\overline{AB}||\cdot||\overline{AC}||}} - \left(\frac{\sum_{B\in\mathscr{D}}\sum_{C\in\mathscr{D}}\frac{1}{||\overline{AB}||\cdot||\overline{AC}||}\cdot\frac{<\overline{AB},\overline{AC}>}{||\overline{AB}||^2\cdot||\overline{AC}||^2}}{\sum_{B\in\mathscr{D}}\sum_{C\in\mathscr{D}}\frac{1}{||\overline{AB}||\cdot||\overline{AC}||}}\right)^2 \tag{7.1}$$

Based on ABOF, our proposed p-partial ABOD works as follows. Given that we already have profiling dataset regarding the resident's different regular living activities, we first compute off-line the angle variances of vector pairs in the dataset generated by the same activity for all p categories ($p = 4$ in our case, i.e., sitting, standing, lying and walking, see Figure 7.7 (a)$\sim$(d)). Then, based on the variances, we decide the lower and upper bounds for the p categories by a box and whisker diagram, which is a standardized way of displaying the distribution of data (see Figure 7.12). We then can online calculate the angle variances of an observed RSSI vector paired with the p-category profiling datasets (i.e., falling to sitting, falling to standing, falling to lying and falling to walking, see Figure 7.7 (e)). We assign labels to the test sample whose angle variances are within the corresponding boundaries. If multiple labels are assigned, we choose the category that the corresponding variance of testing data point lies in the middle of box as most (*e.g.,* assume that the angle variances of a test data point paired with walking data and standing data are both 10^{-12}, but the variance for walking data is in the middle of the box more, see Figure 7.12, we assign the test data as walking). If no labels are assigned, we treat it as a potential outlier.

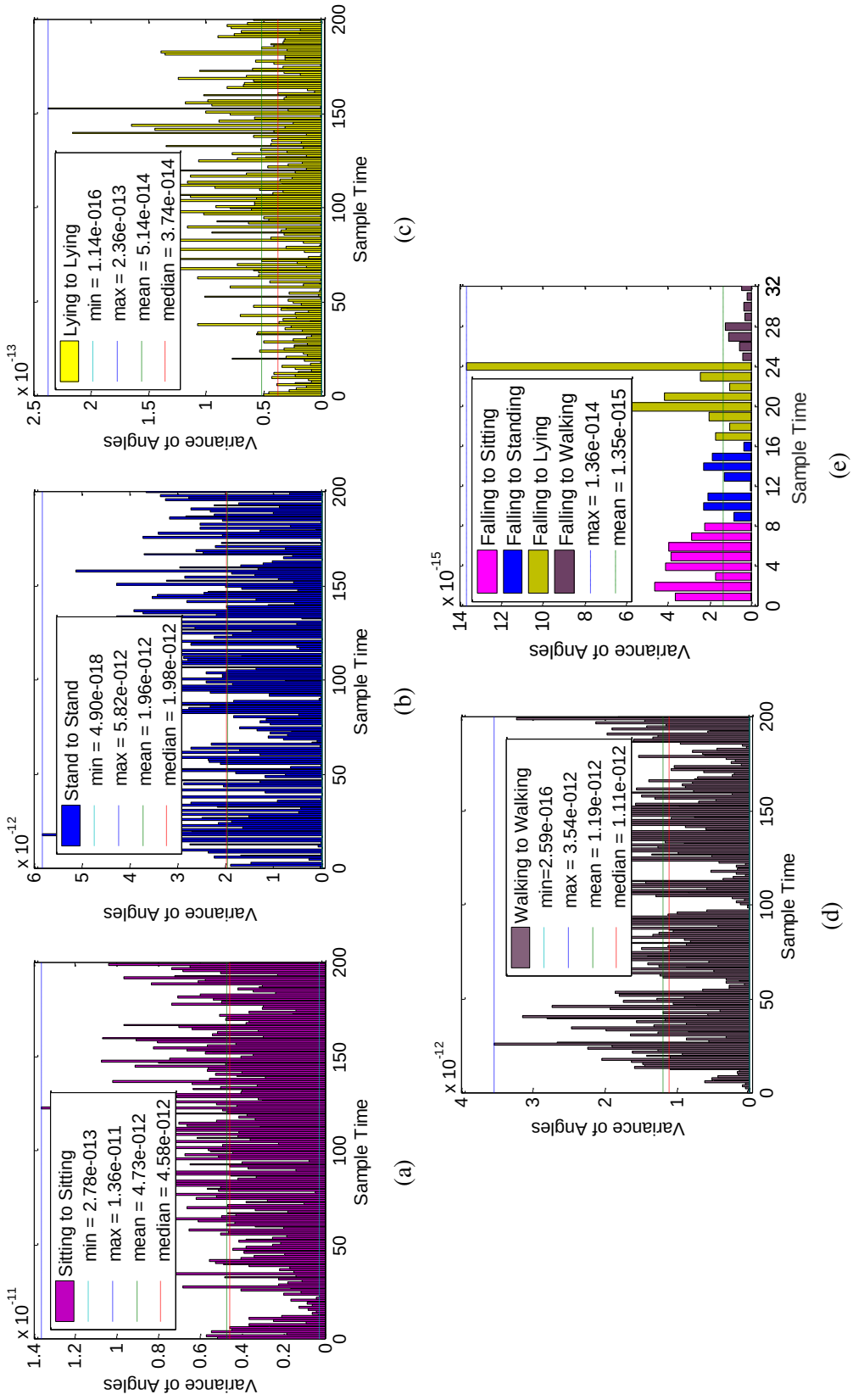


Fig. 7.7 Intuition of pABOD

Algorithm 3: Accumulated Cost Matrix

Input: Two multi-dimensional time-series: $\mathbf{X}, \mathbf{Y}$, Local Cost Matrix $\mathbf{C}$
Output: Accumulated Cost Matrix $\mathbf{M}$

$N \leftarrow$ length of $\mathbf{X}$
$M \leftarrow$ length of $\mathbf{Y}$
$\mathbf{M} \leftarrow new[N \times M]$
$\mathbf{M}(0,0) \equiv 0$
for $i = 1; i \leq N; i++$ **do**
  $\mathbf{M}(i,1) \leftarrow \mathbf{M}(i-1,1) + \mathbf{C}(i,1)$;
end
for $j = 1; j \leq M; j++$ **do**
  $\mathbf{M}(1,j) \leftarrow \mathbf{M}(1,j-1) + \mathbf{C}(1,j)$;
end
for $i = 1; i \leq N; i++$ **do**
  for $j = 1; j \leq M; j++$ **do**
    $\mathbf{M}(i,j) \leftarrow \mathbf{C}(i,j) + min\{\mathbf{M}(i-1,j); \mathbf{M}(i,j-1); \mathbf{M}(i-1,j-1)\}$;
  end
end
return $\mathbf{M}$

Unlike previous fall detection systems which usually either utilize some learning-based classification to distinguish a falling event from other activities [63], or first adopt anomaly detection method to detect an outlier point and then perform one-vs-all classification [82], our method focuses on mining the clustering patterns of RSSI data based on the intra angle-variance and inter angle-variance of multiple-group dataset to avoid parameter tuning, as well as performs the detection and classification simultaneously. In practical, the proposed method needs to proceed an off-line Angle Factor learning before achieving the real-time fall detection, which is less satisfactory and will be improved in our future work.

7.4.2 Falling Direction Sensing

Sensing the falling direction in fact is a *time-series classification* problem. We need to classify the segmented RSSI data stream with a label (i.e., falling directions). To tackle the problem, we introduce a DTW based kNN method.

Algorithm 4: Optimal Warping Path

Input: Accumulated Cost Matrix **M**
Output: Optimal Warping Path *Path* and DTW distance *Dis*

```
Path[] ← new array
i = rows of M
j = columns of M
Dis = 0
while (i > 1)&(j > 1) do
    if i == 1 then
        j = j − 1
    end
    if j == 1 then
        i = i − 1
    end
    else
        if Path(i−1, j) == min{M(i−1, j); M(i, j−1); M(i−1, j−1)} then
            i = i − 1
        end
        if Path(i, j−1) == min{M(i−1, j); M(i, j−1); M(i−1, j−1)} then
            j = j − 1
        end
        else
            i = i − 1; j = j − 1
        end
        Path.add((i, j)); Dis = Dis + M(i, j)
    end
end
return Path, Dis
```

DTW is an efficient algorithm for measuring similarity between two temporal sequences which may vary in time or speed (*e.g.,* walking pattern, speech recognition) [181]. Given two multi-dimensional time-series, $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ and $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_M\}$, where $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^D$, algorithm starts by building the local cost matrix $\mathbf{C}$ representing all pairwise distances between $\mathbf{X}$ and $\mathbf{Y}$:

$$\mathbf{C} \in \mathbb{R}^{N \times M} : c_{i,j} = \left\| \mathbf{x}_i - \mathbf{y}_j \right\|, i \in [i : N], j \in [1 : M] \tag{7.2}$$

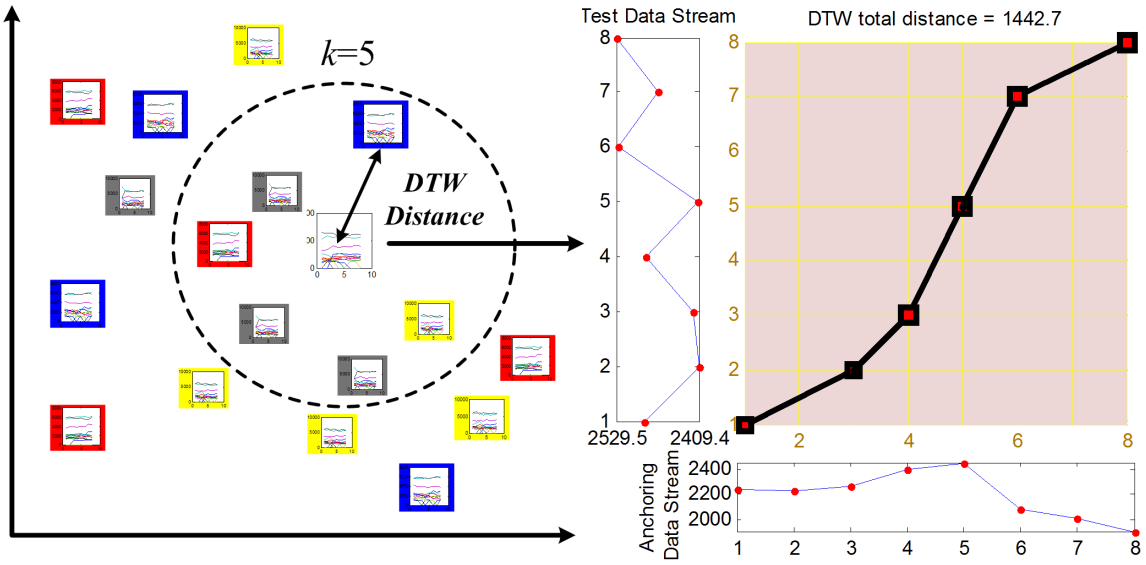


Fig. 7.8 Outline of DTW based kNN

Based on the Local Cost Matrix $\mathbf{C}$, we can construct the Accumulated Cost Matrix $\mathbf{M}$, which contains all possible warping paths (see Algorithm 3). Then Dynamic Programming is used to find the optimal warping path and DTW distance, starting from the point $p_{end} = (M,N)$ to the $p_{start} = (1,1)$ (Algorithm 4).

To our case, when an anomalous RSSI pattern is detected, we first segment a data stream with m continuous time sample (we choose $m = 8$), starting from where we detect as an abnormal point. Then, we calculate all the DTW distances between the segmented data stream and the profiling data streams using the multi-dimensional DTW by optimal matching between two given RSSI sequences. Finally, we can classify the falling directions based on a majority voting by its top k smallest DTW distances. Figure 7.8 illustrates the general idea of our DTW based kNN.

7.5 Evaluation

We evaluate our system in a real-world living bedroom (size: $3.9m \times 3.6m$). Fig 7.9 shows the experimental setup and furniture deployment. Two subjects participate in the experiments, one male ($Age: 28, Height = 172cm, Weight = 68kg$) and one female ($Age: 27, Height = 163cm, Weight = 49kg$).

7.5.1 Evaluation Metrics

For regular actions and falling direction classification, we use standard *precision*, *recall* and *accuracy* to measure our proposed approaches [63]. For fall detection, we evaluate our result in terms of *Detect Rate* and *False Detect Rate* [82].

$$DetectRate = \frac{\text{True Positive}}{\text{\# of Fall Events}} \tag{7.3}$$

$$FalseRate = \frac{\text{False Positive}}{\text{\# of Non-Fall Events}} \tag{7.4}$$

7.5.2 Sensing Normal Activities and Falls

We first collect our profiling data, which involves normal daily living activities (see Figure 7.10, time span is one day). Then we mimic overall 20 different fall events, including various falling directions and locations, shown by 7.11. All fall events are conducted by both participants repeating 3 times each (i.e., 120 fall events).

Based on the collected profiling data, we first calculate the angle variance of RSSI vectors paired by same action category (i.e., sitting, standing, lying and walking, illustrated by Figure 7.7 (a)$\sim$(d)). The mean value of variances for regular actions is ranged from 5.14×10^{-14} to 4.73×10^{-12}, but the maximum value of angle variance paired by falling to regular actions (shown by Figure 7.7 (e)) is 1.36×10^{-14}. Thus, we can easily separate

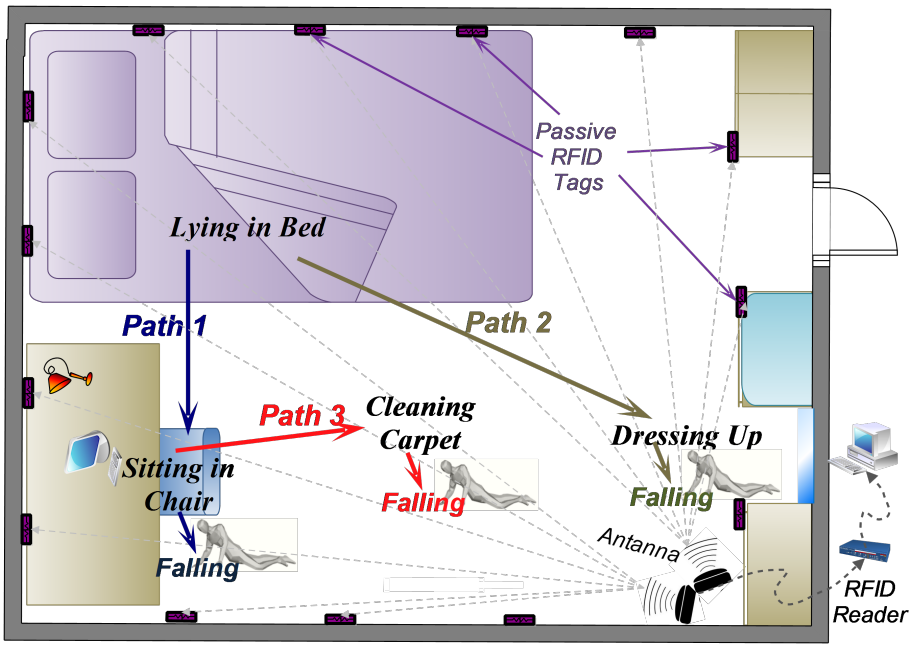


Fig. 7.9 Room layout and three representative action paths

Locations	Activities
Lying in Bed	play mobilephone, read books, listen music
Siting in Bed	play mobilephone, play laptop, read books
Siting in Chair	paly computer, read books, drink water, writting
Standing in Front of Mirror	dress up
Standing beside Bed	read books, stretch body
Standing beside Desk	play computer, read books
Standing in Front of Wardrobe	take out clothes, put in clothes
Standing in Center of Bedroom	do morning exercise
Walking Clean Carpet	clean carpet
Walking Enter 1	walk from door to bed
Walking Enter 2	walk from door to chair
Walking Enter 3	walk from door to the front of mirror
Walking Out 1	walk from bed to door
Walking Out 2	walk from chair to door
Walking Out 3	walk from the front of mirror to door
Walking Random in Bedroom	

Fig. 7.10 Types of normal activities

Fall Location	Fall Direction
Falling from Bed 1	N/A
Falling from Bed 2	N/A
Falling from Bed 3	N/A
Falling from Chair	Right
Falling from Chair	Left
Falling from Standing in Center of Room	Front
Falling from Standing in Center of Room	Back
Falling from Standing in Center of Room	Left
Falling from Standing in Center of Room	Right
Falling from Standing in Front of Mirror	Left
Falling from Standing in Front of Mirror	Right
Falling from Standing in Front of Mirror	Back
Falling from Standing in Front of Wardrobe	Left
Falling from Standing in Front of Wardrobe	Right
Falling from Standing in Front of Wardrobe	Back
Falling while Cleaning Carpet	Front
Falling while Cleaning Carpet	Back
Falling while Cleaning Carpet	Right
Falling while Cleaning Carpet	Left
Falling while Enter the Room	N/A

Fig. 7.11 Different falls in the experiments

the space of regular actions with falls. Figure 7.12 shows our predefined regular activity categories and the learned variance boundaries. We set the lower and upper bound of box diagram as 15% and 85%, so the interquartile range includes 70% of most possible variances. From the box and whisker diagrams, we can easily determine the variance range for each action. The boundaries of regular actions are ranged from 7.42×10^{-14} to 1.23×10^{-12}, in which the lowest value of all four lower boundaries (i.e., 7.42×10^{-14}) is bigger than the maximum value of angle variance calculated by falling to regular actions (i.e., 1.36×10^{-14}). This further verifies the feasibility of our method.

After the boundaries for each normal action are learned, we collect 3,492 non-fall events (varies in time length) generated by regular activities to test our method (*e.g.,* reading book in bed, cleaning carpet, see Figure 7.10), which can achieve overall 94.7% accuracy.

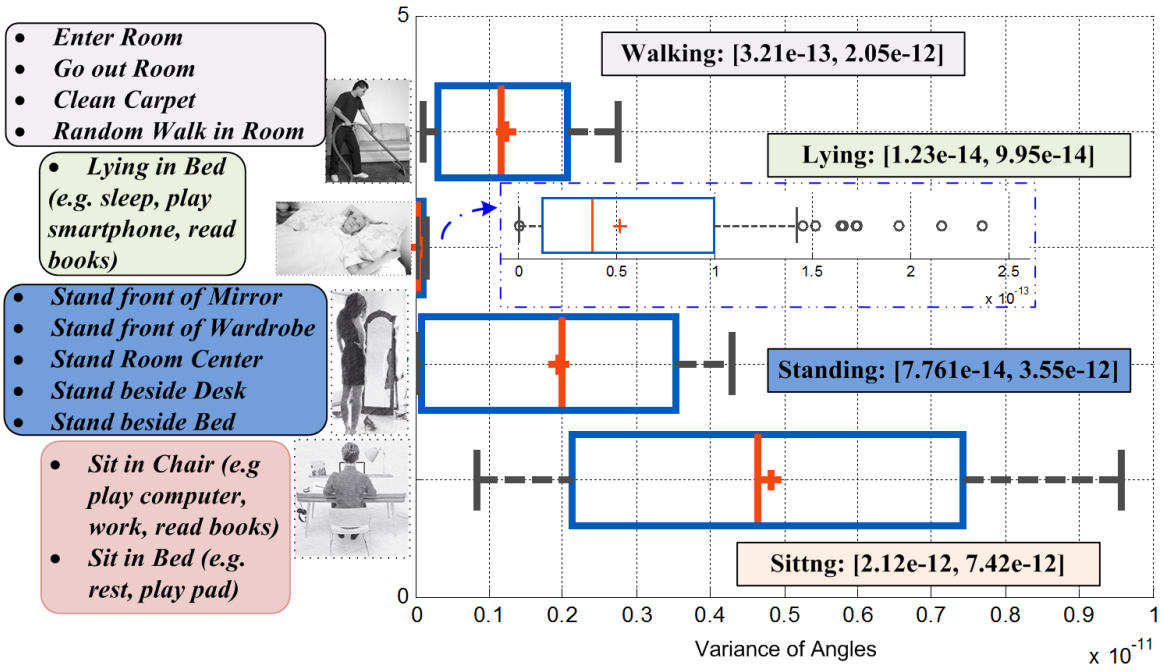


Fig. 7.12 Regular activity categories and boundaries

Figure 7.13 (a) (b) illustrate the performance of sensing regular activities. kNN and SVM are two classification methods that are frequently used by other fall detection systems [63]. Thus, we compare our method to kNN [69, 80] ($k = 5$) and SVM [82, 72] (linear kernel, termination criterion=0.015, C=100, others as default [184]). Our method performs well in distinguishing sitting (98.7% accuracy) and standing (96.7% accuracy) actions but slightly worse in lying (92.4% accuracy) and walking (91.5% accuracy). kNN method only achieves 81.9% in classifying walking action. Our method does not require tuning any parameters and achieves comparable good accuracy, although SVM performs slightly better in distinguishing walking (93.1% accuracy) and Lying (93.5% accuracy) action.

Ground Truth	Sitting	Standing	Lying	Walking	Recall
Sitting	880	0	4	20	0.973
Standing	4	708	56	14	0.905
Lying	0	8	1160	18	0.978
Walking	8	16	36	560	0.903
Precision	0.987	0.967	0.924	0.915	0.947

(a)

Algorithms	Sitting	Standing	Lying	Walking
Our Method	0.987	0.967	0.924	0.915
kNN[1][20]	0.962	0.913	0.921	0.819
SVM[12][14]	0.981	0.963	0.935	0.931

(b)

Algorithms	# of Falls	Detect Rate	# of non-falls	False Detect Rate
Our Method	120	0.908	3492	0.121
LOF[12]	120	0.817	3492	0.163

(c)

Fig. 7.13 Confusion Matrix and Detection Performance

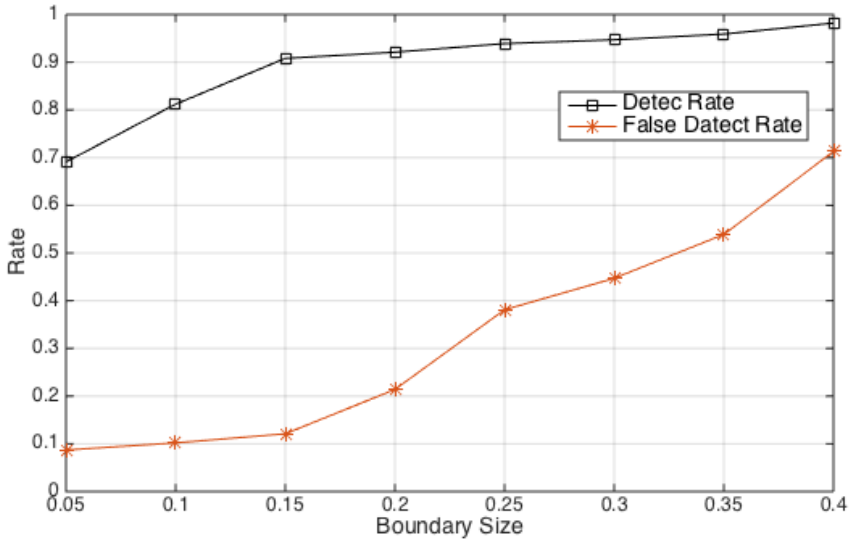


Fig. 7.14 Detection rate and false detection rate varies with the boundaries size (X-axis only shows the lower boundary, so upper boundary should be $100\% - LowerBoundary$, the boundary range should be $UpperBoundary - LowerBoundary$)

Figure 7.13 shows the capability of our method in detecting falls. We test overall 120 fall events, including falls from working at desk, dressing up, cleaning the carpet, and falling to different orientations (*e.g.,* falling to front, to back, to right and to left, shown by Figure 7.11) and 3,492 non-fall events. The result shows that our method can achieve 90.8% detection rate and 12.1% false detection rate. As a comparison, we also utilize LOF (adopted by WiFall [82]) to our dataset, which receives a 81.7% detection rate and 16.3% false detection rate. In this setting, we set the boundaries of box diagram (Figure 7.12) as from 15% to 85%. We can choose different boundaries of the box diagram (*e.g.,* 5%∼95%, 10%∼90%, 20%∼80%, see Figure 7.14). It illustrates that both the detection rate and false detection rate increase when the boundary size becomes smaller (spans from 90% to 20%). However, the false detection rate experiences dramatic growth but the true detection rate in fact does not significantly increase (from 90.8% at 15% to 98.2% at 40%). Thus, we choose 15% and 85% as our lower and upper boundaries in term of the box and whisper diagram.

Ground Truth		Fall to Front	Fall to Left	Fall to Right	Fall to Back	Recall
	Fall to Front	9	2	1	0	0.750
	Fall to Left	2	27	0	1	0.900
	Fall to Right	2	1	26	1	0.867
	Fall to Back	0	1	1	22	0.917
	Precision	0.692	0.871	0.929	0.917	0.875

Fig. 7.15 Confusion Matrix of DTW based kNN ($k = 3$)

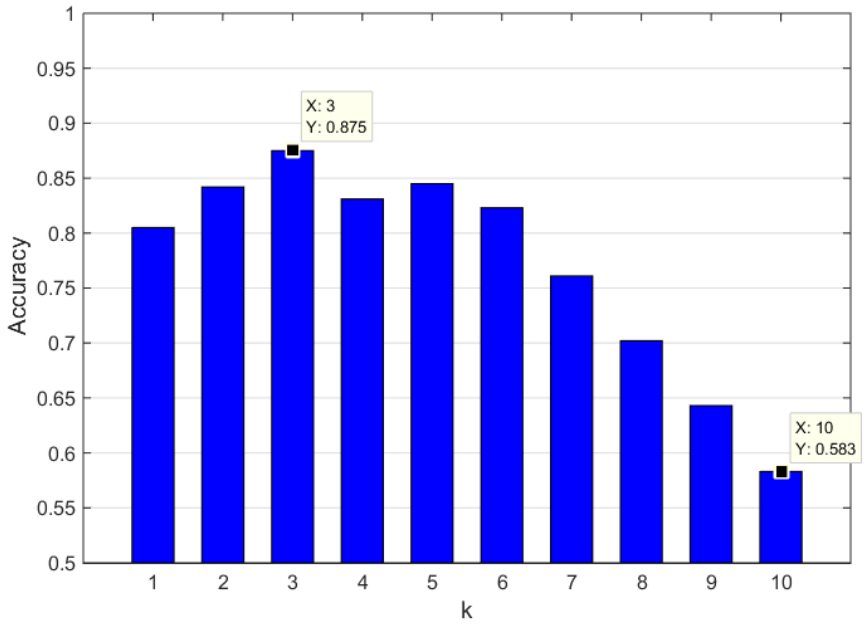


Fig. 7.16 Accuracy of classifying falling direction varies with parameter k

For falling direction classification, we choose 16 fall events which contain falling directions (see Figure 7.11, some falls have no direction context such as falling from bed). Each fall is conducted by both participants and repeated 3 times each (overall 96 fall events). Figure 7.15 shows the confusion matrix of our DTW based kNN method choosing $k = 3$.

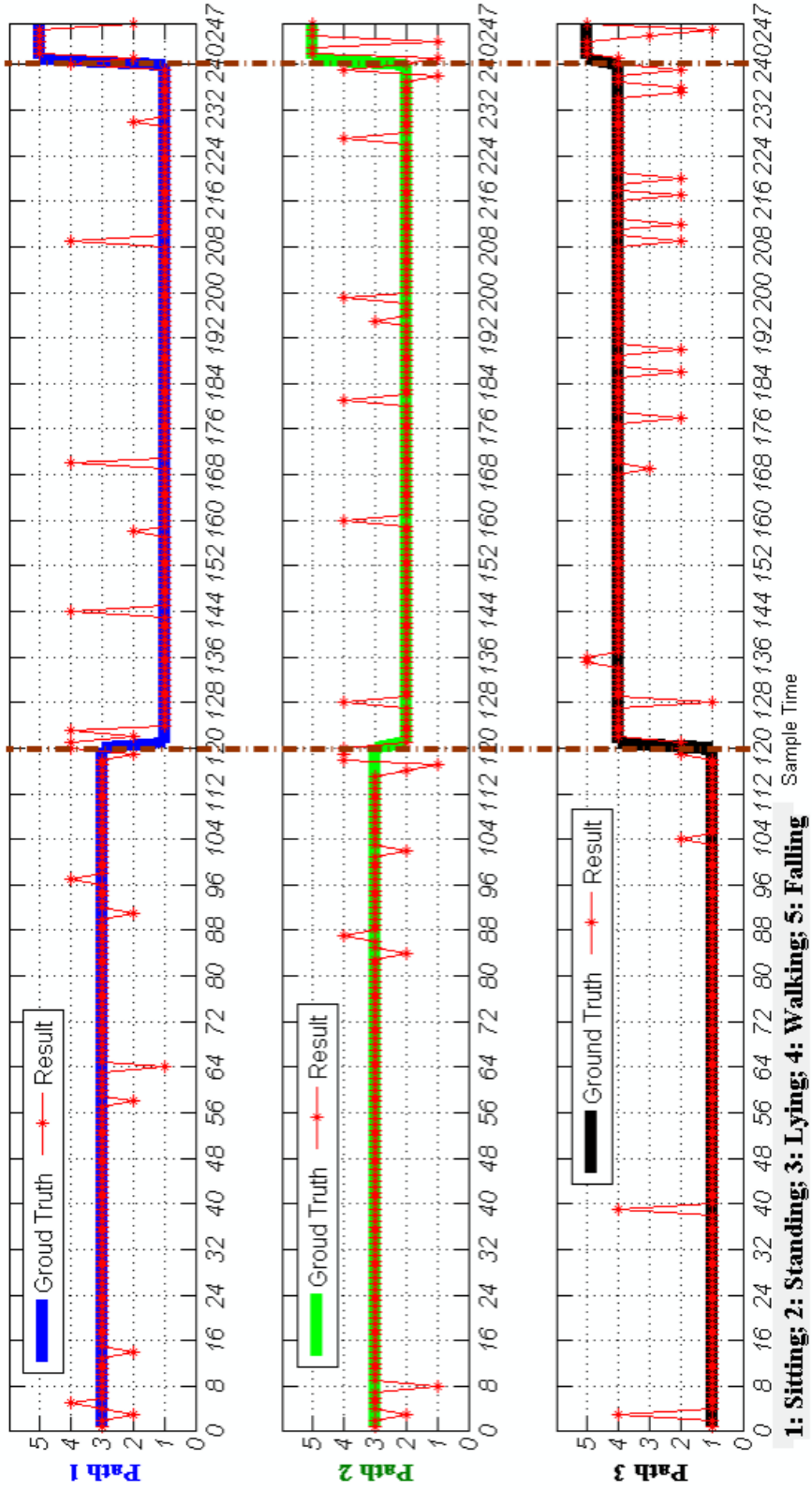


Fig. 7.17 Detect fall events in action paths

We can observe that the overall accuracy is 87.5%, but the precision and recall in classifying *falling to front* are only 69.2% and 75%. The reason may lie in fact that *falling to front* and *falling to right/left* are quite similar in some cases since these two falling directions are adjacent. Our method performs good at distinguishing the *falling to back* (precision and recall are both 91.7%) which possible cause severe damage to the head. The key parameter in DTW based kNN is the k value that heavily affects the classification accuracy. Figure 7.16 illustrates the relation of classifying accuracy with parameter k. As it shows, the accuracy at first increases with the growth of k value, climbing the peak at $k = 3$, then gradually decreases along with the increase of k. Thus, we choose $k = 3$ in our experiments.

Figure 7.17 shows performance of our system in three representative action paths. The bold lines are the ground truth, the Y-axis from 1 to 5 represent action categories (i.e., sitting, standing, lying, walking and falling). The three action paths are shown by Figure 7.9:

- Lying in Bed (60 seconds) $\Longrightarrow$ Sitting in Chair (60 seconds) $\Longrightarrow$ Falling Down,
- Lying in Bed (60 seconds) $\Longrightarrow$ Dressing Up in Front of Mirror (60 seconds) $\Longrightarrow$ Falling Down,
- Sitting in Chair (60 seconds) $\Longrightarrow$ Cleaning the Carpet (60 seconds) $\Longrightarrow$ Falling Down.

We can see that in the first action path, our method can timely distinguish the actions and detect the fall event, although generating some unstable predictions when the resident transfers from getting up from bed to sitting in chair. Our system on the second action path displays the same classifying capability, but it outputs some bad predictions after the resident falls from dressing up although it successfully detects a fall event in the first few points. From the third action path, we observe that the classification result is not as good as previous two paths when the subject is cleaning the carpet, for the reason that cleaning involves plenty of activities that may generate some similar RSSI patterns as sitting and standing actions. In summary, when an activity shift occurs (*e.g.,* from lying to sitting in the

chair, from sitting to cleaning the carpet), the sensing results are usually decayed, which is normal due to unpredictable movements of human body. After detecting a fall, the continuing sensing result is unstable since people usually lie or sit on ground after a fall, which is similar to our predefined regular activities.

7.6 Discussion

7.6.1 Computation Cost

In the Profile Construction phase (off-line part), for a daily recorded activities, based our configuration, the calculation time for angle variances is around 70 minutes. With the constructed profile data, we can online process each given data sample (i.e., the fall detection phase) within 0.4 seconds. For the falling direction sensing phase, the calculation complexity of DTW is $O(NM)$ [181] (both equal to 8 in our case), so calculation itself is fast. However, we need to segment a fix-length data stream beforehand, which results in a latency (about 4 seconds in our case). However, in the direction sensing phase, we aim to provide fine-grained contexts regarding the happened falls, which does not affect timely detecting a fall and sending an alarm (done in the fall detection phase).

7.6.2 Hardware

We use standard, commercial RFID system with passive tags in our work. The passive tags are more cost-effective and, due to their simple structure and protective encapsulation, more robust than the sensor nodes. Passive tags operate without batteries. Once deployed, no further maintenance is required. The devices that require power in our sensing system is the RFID reader and antenna. But recent technical trends show that low-cost, low-power RFID readers are becoming commonly available by integrating into the smart phones, making our work potentially beneficial to the more users in the future.

7.6.3 Detection Methods

As for fall detection techniques, current fall detection systems mainly adopt supervised classification-based method to detect a fall event, such as Support Vector Machine (SVM) [70, 66, 82], Neural Network [74] or Extreme Learning Machine [76], which have to tune many parameters to achieve satisfied accuracy. But in our fall detection phase, we aim to *mine* the clustering patterns of RSSIs based on the variances of angle paired by data point of different actions when the environment is affected by diverse human activities. Thus, different to the traditional classification or distance-based anomaly detection methods, our proposed method relaxes the requirement of tuning parameters that is time-consuming and sensitive to different test scenarios [185].

7.6.4 Limitations

One of the limitations is that the current system is designed for and tested with only a single resident. We believe that this is an important use case, particularly in an aging-in-place setting, which aims to ensure that a single person can live in his/her home and community safely and independently regardless of age and ability level. However, the number of profiles needed with multiple persons would increase exponentially. A more promising approach therefore would be to find techniques that can isolate concurrent activities in separate space from each other and match them against profiles separately, which we will consider in our future work. Another limitation is that labeling profiling data is time-consuming and labor-intensive, which is also an issue shared by other fall detection systems. In the Profile Construction phase, we have to use a camera to record the daily living activities, and then synchronize the camera and RSSI reading based on the time stamp, finally label and segment data streams into different action categories to build a labeled profile dataset based on the video records.

7.7 Conclusion

To detect a fall event in our daily living environments, we present an unobstructive, fine-grained fall detection system based on pure passive RFID tags. By proposing a p-partially Angle-based Outlier Detection method, our system can simultaneously identify regular activities and detect a fall event. By adopting DTW-based kNN, the proposed system can distinguish different falling orientations. Our approach relaxes the requirement of tuning parameters and provides more fine-grained contexts regarding fall events comparing to the current fall detection systems.

From Chapter 4 to Chapter 7, we design a series of functionalities so that our living-assistive system can accurately localize the resident's indoor locations, recognize her activities and detect the falling events. Another important facet in an intelligent living-assistive system is that how the user can conveniently and unobstructively interact with those upper-layer systems and applications. In the next chapter, we will design a novel human-machine interaction approach using in-air hand gestures.

Chapter 8

Realizing Human-Machine Interactions Using Touch-free Hand Gestures

One important research issue for an intelligent residential home is how to accurately and conveniently control the domestic electronic appliances (*e.g.,* automated window curtain, brightness-adjustable lamp, TV and air conditioner). For example, we enter a smart house and turn on the TV by simply waving a hand in the air, then we can use another hand gesture to turn on the Air Conditioner as well, furthermore, by several continuous up-and-down hand-waves, we can adjust the Air Conditioner into a comfortable temperature. To realize such an envisioned functionality, in Chapter 8, we present a touch-free human-machine interaction approach via a novel, device-free, multi-module Hand Gesture Recognition (HGR) system, called *AudioGest* .

Hand gesture nowadays becomes one of most popular means of interacting with consumer electronic devices, such as mobile phones, tablets and laptops. Our designed device-free HGR system in this chapter can accurately sense the hand in-air movement around user's devices. Compared to the state-of-the-art, *AudioGest* is superior in using only one pair of built-in speaker and microphone, without any extra hardware or infrastructure support and with no training, to achieve a multi-modal hand detection. In particular, our system is not

only able to accurately recognize various hand gestures, but also reliably estimate the hand in-air duration, average moving speed and waving range. We achieve this by transforming the device into an active sonar system that transmits inaudible audio signal and decodes the echoes of hand at its microphone. We address various challenges including cleaning the noisy reflected sound signal, interpreting the echo spectrogram into hand gestures, decoding the Doppler frequency shifts into the hand waving speed and range, as well as being robust to the environmental motion and signal drifting. We extensively evaluate our system on three electronic deivces under four real-world scenarios using overall 3,900 hand gestures collected by five users for more than two weeks. Our results show that *AudioGest* detects six hand gestures with an accuracy up to 96%. By distinguishing the gesture attributions, it can provide more fine-grained control commands for various applications.

8.1 Introduction

The booming of consumer electronic devices has greatly stimulated the research on novel human-computer interactions. Hand gestures are a natural form of human communication with devices that have aroused enormous attentions from both industry and academia [103, 186, 187]. Researchers and companies try to integrate the hand-gesture recognition into our daily devices, including laptops [111], tablets [109], smartphones [188], and gaming consoles [189]. However, a crucial prerequisite of these applications is that the device can accurately and robustly detect gestures anytime (*e.g.,* poor light condition at night), anywhere (*e.g.,* in rural area without wireless connection) in a device-free manner (*e.g.,* no need to wear extra devices/sensors) [109, 190, 191].

Over the last decade, many state-of-the-art hand gesture recognition (HGR) systems have been developed using various hardware platforms, such as computer vision [192], inertial sensors [193], ultrasonic sensors [111], infrared sensors (*e.g.,* Leap Motion), and depth sensors [189, 194]. While promising, most of these systems, however, can only partially

meet those requirements [103]. For example, vision-based techniques are sensitive to the light conditions (*i.e.,* performance greatly decreases in poor lighting conditions), and are usually regarded as privacy-intrusive. Although some commercialized HGR systems (such as Kinect, Leap Motion) achieve enormous success, their applications are still limited in computers, also need relatively high installation and instrumentation overhead (around 50$\sim$250 USD). The wearable sensor based approaches (*e.g.,* attaching 3-axis accelerometers or gyroscopes on hand) unavoidably require the user to wear additional devices. Although those systems can achieve fine-grained and multi-level hand motion detection in high precision, they may not be practical in real-world applications (*e.g.,* user may feel uncomfortable or forget to wear the devices).

Many WiFi-based solutions have recently been proposed to overcome the above limitations. For example, WiGest [103] exploits the influence of in-air hand movement on the wireless signal strength of the device from an access point to recognize the performed gestures. Melgarejo *et al.* [108] leverage a directional antenna and WARP board to access various wireless features such as Received Signal Strength (RSS), signal phase differences and CSI (channel state information), then through matching the features from users' gestures with a standard set of pre-trained templates to recognize user's hand gestures. WiSee [107] exploits the doppler shift in narrow bands extracted from wide-band OFDM (orthogonal frequency-division multiplexing) transmissions to recognize nine different human gestures. Although WiFi-based systems can work under any lighting conditions and do not require dedicated hardware modification, those systems, however, require the mobile device to be always connected to a wireless transmitter/receiver, which is impractical for some circumstances such as on a train/bus or traveling in a rural area.

To tackle these challenges, we develop AudioGest, a device-free system that can transform consumer device into an active sonar system by utilizing the embedded microphone and speaker of the mobile device. Compared to other HGR systems, AudioGest exploits only

one pair of built-in speaker and microphone without adding any extra cost on hardware. AudioGest does not require the model-training to achieve a multi-modal hand gesture detection. The system not only can recognize hand gestures but also is able to accurately estimate the hand in-air time, average waving speed, and the hand moving range. We call such capability as *multi-modal* hand motion detection.

Implementing such a practical system, however, requires addressing a number of non-trivial challenges. First, the ambient noise (*e.g.,* human conversation, electronic noise) dominates the recorded audio signals (see the experiments in Sec. 8.3.1). It is hence difficult to perceive the weak Doppler frequency shifts, let alone decoding the hand waving directions, speed, and range. Another challenge is the signal drifting brought by the device diversity and time elapse (see the experiments in Sec. 8.3.2). Since we emit a high-frequency audio signal ($> 18kHz$, making it inaudible to human), the Operational Amplifier (OA) in microphone and speaker both experience attenuation, making the magnitude of recorded echoes unstable. Moreover, different microphones/speakers have various OA attenuations, also resulting in signal drifting.

In AudioGest, we propose three main techniques to tackle the aforementioned challenges. First, we introduce a FFT-based normalization that substantially adjusts the magnitude of FFT frequency bin in different timestamps to the same level, removing the influence of OA attenuation in high-frequency part (see details in Sec. 8.5.1). We then perform *Squared Continuous Frame Subtraction*, in which we first subtract the spectrum of current audio frame by previous frame and square the magnitudes of frequency bins, further eliminating the nearby human motion influence (see details in Sec. 8.5.2). Furthermore, we apply a Gaussian smoothing filter [195] to transfer the discrete shifted frequency bins into a contouring area. We decode it into the real-time hand moving velocity curve based on the Doppler frequency shift (see details in Sec. 8.5.4). Finally, according to the velocity curve, we estimate hand

gesture, moving speed, and waving range (see details in Sec. 8.5.5). In a nutshell, our main contributions are summarized as follows:

- We introduce an approach that utilizes one pair of COTS microphone and speaker to accurately detect the hand movement and to estimate fine-grained hand waving attributes. Our in-situ experiments with five users over a period of two weeks demonstrate the feasibility and accuracy of AudioGest in various living environments.

- We propose a denoising pipeline that not only abstracts the Doppler frequency shifts from weak echo signals, but also deals with the signal drifting issue caused by hardware diversity and time elapse.

- AudioGest is a training-free system that accurately recognizes 6 hand gestures with an accuracy of 95.1% on average, precisely distinguish the magnitude differences of various hand speed and moving range, providing up to 54 control commands by randomly choosing two attributes.

The rest of the chapter is organized as follows. We introduce the preliminaries in Sec. 8.2. Sec. 8.3 depicts the empirical studies and challenges. We present the system conceptual overview in Sec. 8.4 and details the *AudioGest* system in Sec. 8.5. The evaluation results are reported in Sec. 8.6. Finally, we discuss the limitations and conclude our work in Sec. 8.8.

8.2 Preliminaries

This section briefly introduces the Doppler effect and the COTS hardware utilized in our research (*i.e.,* speakers and microphones).

8.2.1 Doppler Effect

Most of current HGR systems utilize labeled sensor readings (including images) to train a classification model, and then distinguish hand gestures, which is lack of physical interpretation. It is also hard for those systems to detect some context information regarding the hand gestures, such as hand's moving speed and in-air waving duration. AudioGest system in this chapter, conversely, is inspired by a prevalent law in the physical world namely the *Doppler Effect*.

Doppler effect illustrates and quantifies the wavelength changes when wave energy like sound or radio waves travel between two objects if one or both of them move. The Doppler effect causes the received frequency of a source to differ from the sent frequency if there is motion that is increasing or decreasing the distance between the source and the receiver. The general equation of measuring frequency shift is as follows:

$$\Delta f = \frac{\Delta v}{v_{wave}} f_{source} \tag{8.1}$$

where $\Delta f = f_{receiver} - f_{source}$, called *Doppler Frequency Shift*; $\Delta v = v_{receiver} - v_{source}$, is the velocity of the receiver relative to the source: it is positive when the source and the receiver are moving towards each other.

In our case, the wave source (*i.e.,* speaker) and the receiver (*i.e.,* microphone) are both motionless but the reflector (*i.e.,* human hand) are moving. Hence, though most of sound waves stay unchanged, a part of acoustic waves that is reflected by a moving hand experiences a Doppler frequency shift measured by Eqn. 8.2:

$$f_{received} = \frac{1 + v_{rad}/v_{sound}}{1 - v_{rad}/v_{sound}} f_{sound} \tag{8.2}$$

where v_{rad} means the radial speed of hand to microphone. Such Doppler effect caused by the motion of a reflector is widely adopted in modern radar systems or underwater sonar systems.

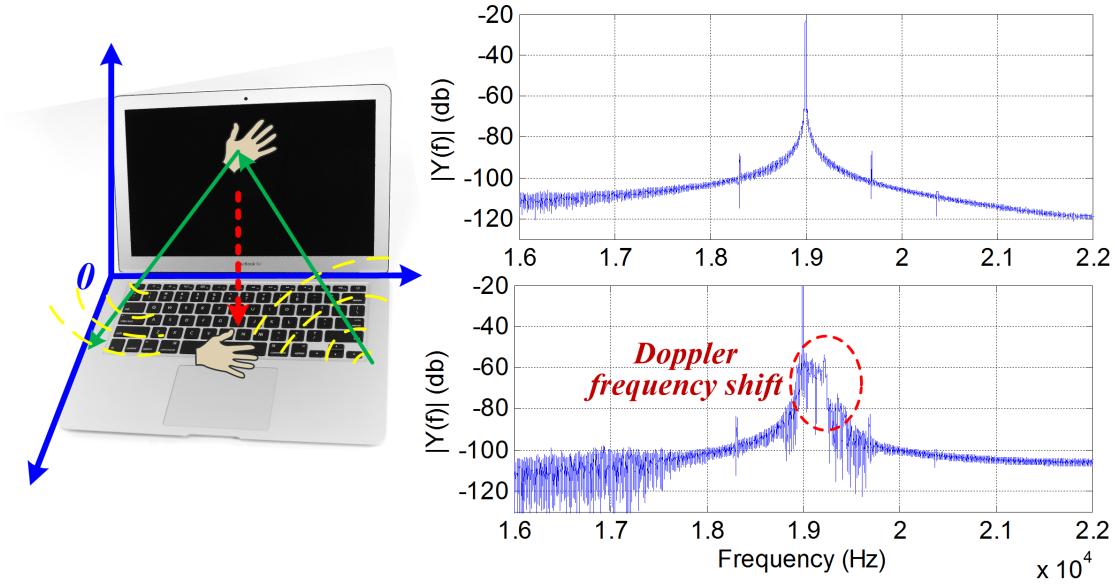


Fig. 8.1 Illustration of Doppler Frequency Shift

Motivated by this intuition, *AudioGest* aims to sense such doppler frequency shift of weak reflected acoustic waves by a moving hand. As shown in Fig. 8.1, when a hand moves in different directions or at different speeds, it will cause different Doppler frequency shifts (*e.g.,* different shapes, different intensities and durations). Our AudioGest targets to decode such Doppler frequency shifts, to recognize the gestures, and to estimate the moving speed and duration of a hand in air.

8.2.2 COTS Speakers & Microphones

In this chapter, we aim to turn the COTS speakers and microphones into an active sonar system to detect fine-grained hand gestures without annoying normal human audition. Such a system, however, needs the support of high-definition audio capabilities.

Normally, human audible signal lies between 20Hz∼18kHz. Assuming that maximum hand waving speed is less than 4m/s, it requires 0.47kHz extra bandwidth (under a sampling rate of 44.1kHz, see Sec. 8.5 for details on how to calculate the frequency bandwidths). As a result, the speakers and microphones needed should be at least with a capability of up to

Fig. 8.2 Speakers and microphones in COTS mobile devices

18.47kHz frequency-response. According to the Nyquist–Shannon sampling theorem, to accurately recover a 20kHz signal, the microphones at least support a 40kHz sampling rate. Fortunately, mobile devices are increasingly supporting high-definition audio capabilities targeted at audiophiles. In particular, such advancement includes high-frequency response range, microphone arrays for stereo recording and noise cancellation, and 4× improvement in audio sampling rates. Fig. 8.2 shows COTS microphones and speakers of three typical mobile devices. They all can support up to 22kHz response frequency and typical 44.1kHz or 48kHz sampling rate, making it possible to achieve fine-grained hand detection.

8.3 Empirical Studies and Challenges

In this section, we will conduct some empirical studies and identify the challenges that we need to deal with.

8.3.1 Weak Echo Signal

As Fig. 8.1 shows, we transmit a 19kHz sine acoustic wave (for 3s) from the right channel of the speaker in a laptop (*i.e.,* MacBook Air). Simultaneously, we record the ambient sound signal using a microphone. At the same time, a participant waves his hand in different

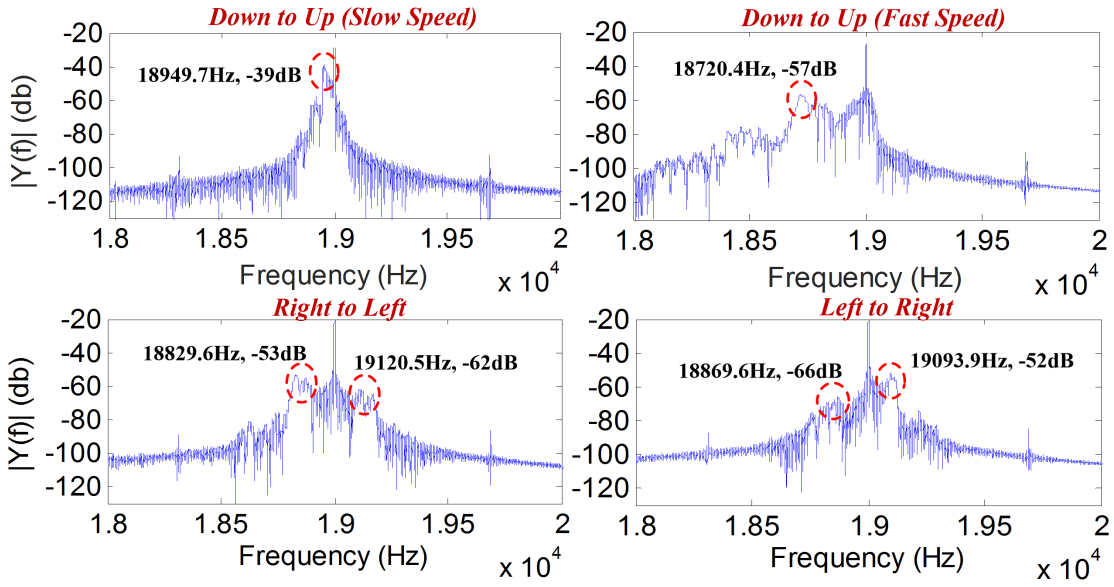


Fig. 8.3 The Doppler frequency shifts caused by different hand gestures and waving speeds

directions and speeds. Then we conduct an FFT to see the frequency shift of audio signal caused by hand motion.

From Fig. 8.3, we can observe that the waving hand from down to up results in an observable magnitude increase in the lower frequency bins, but moving hand from left-to-right/right-to-left is less obvious and the echo signal is weak (*i.e.,* the bins marked by the red circles, the left sides of 19kHz bin). In particular, we find that the motion speed of the hand is highly related with the location of such increased frequency bins, *i.e.,* moving hand in a slow speed causes a risen magnitude in 18,949.7Hz bin, but with a fast speed, it leads to an increase in 18,720.4Hz bin. Also, moving hand from right to left and left to right will arouse a frequency shift in both sides but with opposite intensities (*e.g.,* -53dB and -62dB for right-to-left, -66 dB and -52dB for left-to-right).

In summary, such observable frequency shifts highly motivate our AudioGest system but also bring us a challenging task - how to abstract such weak, vulnerable frequency-bin changes from wideband[1] audio signals. Moreover, we intend to decode the fine-grained hand

[1]Normally, a microphone can resolve 0∼22.05kHz sound signal for a 44.1kHz sampling rate.

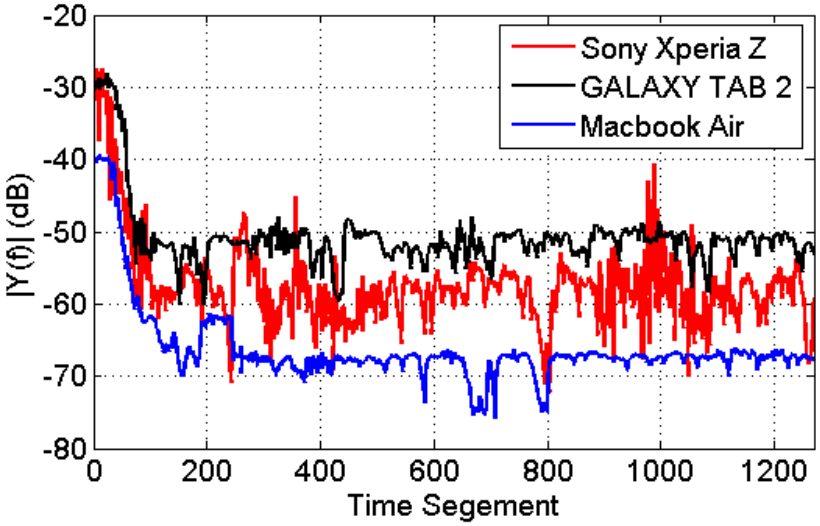


Fig. 8.4 The sound signal drifts for different mobile devices at different time slots

moving speed, in-air duration and motion range beside the hand-gesture recognition. With ambient noise (such as human conversation, electronic noise and environmental sound), it is even harder for us to perceive these Doppler frequency shifts. We will illustrate our solution in Sec. 8.5.2.

8.3.2 Audio Signal Drift

Another challenge is about the audio signal drifts, which can be categorized into two types: *i)* temporal signal drift: audio signals received in different time slots depict various magnitudes for a same frequency bin; and *ii)* diverse-device signal drift: audio signals record by different microphones reveal various magnitudes for the same frequency bin.

Fig. 8.4 illustrates the experiment we conducted under a *static* environment[2], where microphones from various types of mobile devices record 1-hour reflected audio signals while speakers of the same device continuously emit 19kHz inaudible sinuous sound-waves. We divide the 1-hour soundwave into 1,270 signal frames, and further apply 2,048-point

[2] *Static* means no hand moving, same meaning applied in the rest of the chapter.

FFT. We plot the strengths of frequency bin at 19Khz over the time for three different mobile devices in Fig. 8.4. We find that for different mobile devices, the frequency magnitudes are diverse. Even for a same electronic device, the signal strengths fluctuate over the time, and the mobile phone exhibits a stronger signal drift. We also observe that the recorded audio signals drop significantly during first 10 minutes, which lies on two reasons. One reason is that the OA is the main component of the speaker and microphone, and emitting high-frequency sound-waves (*i.e.,* 19kHZ audio signal) will let the OA work on the upper-boundary of its capability, thus is unstable. The other reason is that with the time evolving, continuous ringing of the speaker generates fair amount of heat that increases the working temperature of the electronic components, especially in the first 10 minutes when speakers just start to work[3]. It is well known that the electronic device is very sensitive to temperature, which inevitably influence the performance of the speaker. To summarize, such signal drifting will greatly hinder the system's scalability, which means an HGR approach that works well in one device may be incapable for other devices or in different time-slots. We will deal with this challenge in Sec. 8.5.1.

8.4 System Conceptual Overview

This section will introduce the system architecture of AudioGest, mainly including three conceptual layers - the *gesture detection* layer, the *gesture categorization* layer, and the *application* layer, as shown in Fig. 8.5.

The gesture detection layer is the key part of the whole system (the details shown in the right part of Fig. 8.5). This layer outputs four kinds of gesture contexts - *waving direction, hand's average speed and in-air duration*, as well as *waving range*. Specially, to detect such fine-grained gesture features, we first eliminate the noise of received raw acoustic signal

[3]The working temperature will gradually reach the thermal equilibrium, that is why the signal fluctuation is less significant than the first 10 minutes.

which contains two steps - FFT normalization and background noise subtraction (*i.e.,* dealing with the *Audio Signal Drift* challenge). Then, we need to accurately identify the audio signal segments caused by hand's motion, consisting of two parts - Gaussian smoothing and segmenting the frequency shift area (*i.e.,* tackling the *Weak Audio Signal* challenge). Next, based on the magnitude changes and temporal locations of segmented frequency bins, we interpret such Doppler frequency shifts, thus estimate the hand waving directions. Finally, we put things together, further quantify the hand in-air durations, waving ranges and average speeds.

The gesture categorization layer categorizes different basic gesture characteristics from previous layer into different semantics. As Fig. 8.5 shows, we define overall six gesture directions and three intensity levels for the moving speed, in-air duration and waving range. Unlike previous systems that only detect one or two hand gesture contexts [103, 109], AudioGest provides three types of hand motion attributes except the basic hand gestures. By randomly choosing two motion attributes, AudioGest can theoretically provide up to $6 \times 3 \times 3 = 54$ control commands, which we thus call *multi-modal* hand gesture recognition. It is noted that AudioGest can support a more fine-grained categorization (*e.g.,* classify the in-air duration into four or five levels) which leads to more control commands but degrade the detection accuracy possibly. Vice-versa, we can use a course-grained categorization to increase the estimation accuracy. For example, for an e-book application (only needs 4 commands, *next page*, *previous page*, *full screen*, *normal screen*), we can choose four types of hand waving directions (regardless of waving speed, in-air duration and range) to control these command buttons. This layer provides flexible controlling choices to the application layer.

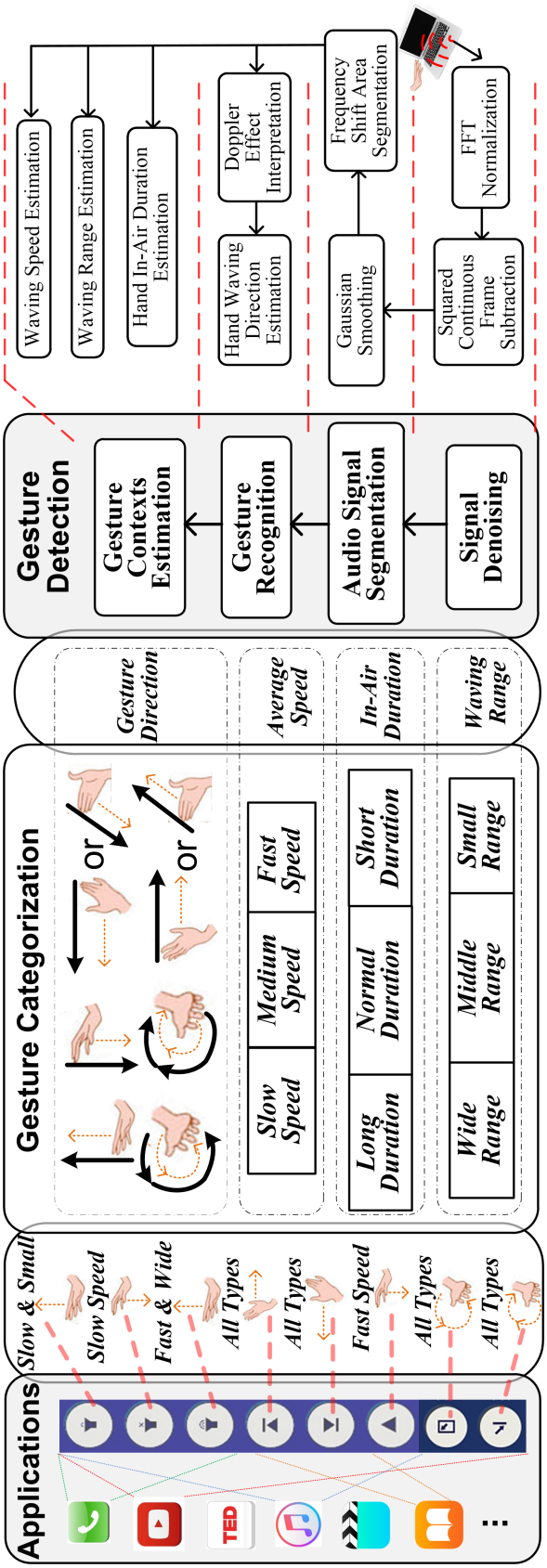


Fig. 8.5 Overview of the system for hand gesture detection

The application layer maps different gestures to control commands for various applications. Typically, one action is mapped to one gesture type and the developer can pick one or more hand gestures to represent an action. For example, for a media player application, a *play* action can be performed with a *Up-Down* hand gesture while a *volume up* action can be mapped to moving the hand up. The volume changing rate can be controlled by the speed or range of the hand waving.

8.5 Realizing the *AudioGest* System

In this section, we will illustrate how to achieve gesture detection and address the associated challenges. Before that, we first introduce how to design the transmitted audio signal. Human normal audible scope is 20Hz∼18kHz. To avoid annoying human audibility, under no circumstance, should AudioGest produce the sound signal below 18kHz (to be more safe, we make it 18.5kHz). Assuming that the fastest hand moving speed is $4m/s$ [111], then the largest Doppler frequency shift[4] $\Delta f_{doppler} = (2v_{hand}/v_{sound})f_{transmit} = 470.6Hz$. Hence, if the mobile device transmits a 19kHz sound, then the received audio signal is 18,529.4Hz∼19,470.6Hz, satisfying the requirement. Also, we save a bandwidth ($2\Delta f_{doppler} = 941.2Hz$) for another possible audio channel[5]. Although microphones in some devices can support a 48kHz or even 192kHz sampling rate, we adopt a more general 44.1kHz sampling rate.

[4]Since we do not know the transmitted sound frequency beforehand, we use a larger possible transmitted frequency 20kHz, $v_{sound} = 340m/s$ under 15 $^{\circ}C$.

[5]It means we can use another speaker channel to transmit a 20kHz sound, and the received signal is 19,529.4Hz∼20,470.6Hz, which lies in the recording capability of a microphone but without inference with another speaker channel.

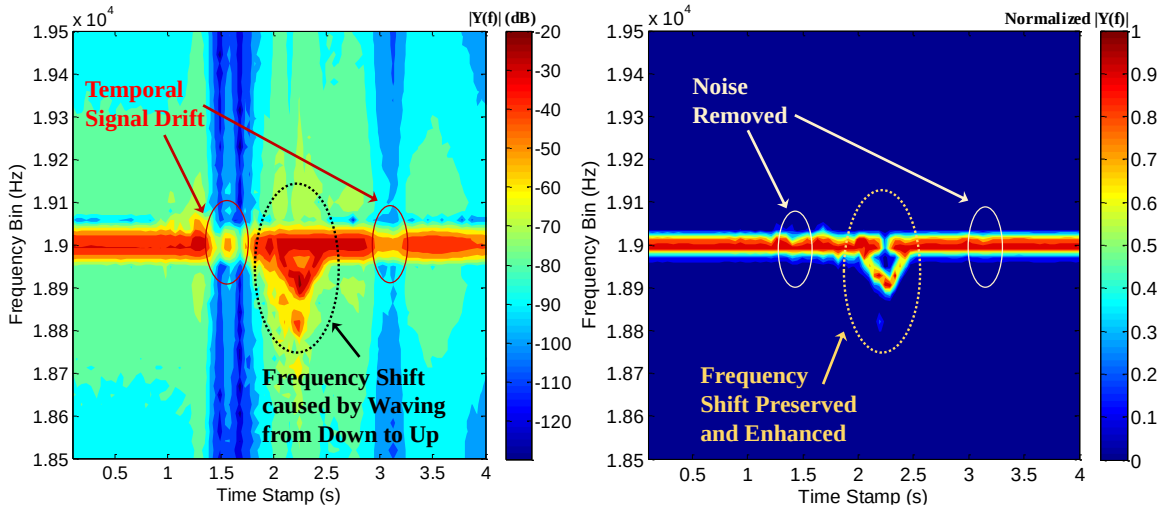


Fig. 8.6 Left Figure: raw audio spectrogram; Right Figure: audio spectrogram after FFT normalization

8.5.1 FFT Normalization

As aforementioned, the raw data recorded by microphones not only contain audible noise but also introduce the signal drifts due to temporal changes and diverse hardwares. This section introduces an FFT-based normalization technique to deal with such issues. Since our targeted sound frequency band is 18.5kHz∼19.5kHz, intuitively, we may need a band-pass filter or high-pass filter. However, the introduced FFT normalization is based upon the frequency domain of the recorded audio signal. We only perform analysis to the FFT bins within the targeted narrow bandwidth. Such processing will naturally filter out the influence of audible noise without adding an extra signal filter.

In order to observe how the Doppler frequency shifts along the time, we first adopt a 2,048-point hamming window to segment the filtered signal into audio frames[6], then apply a 2,048-point FFT[7] to each frame to get the sound spectrogram, shown as the left graph in

[6]Each frame represents $2,048/44,100 = 0.0464s$ audio signal.

[7]With a 44.1kHz sampling rate, the velocity detection resolution $v_{res} = (f_s/FFT_{points})(v_{sound}/f_{source} = 0.39m/s$.

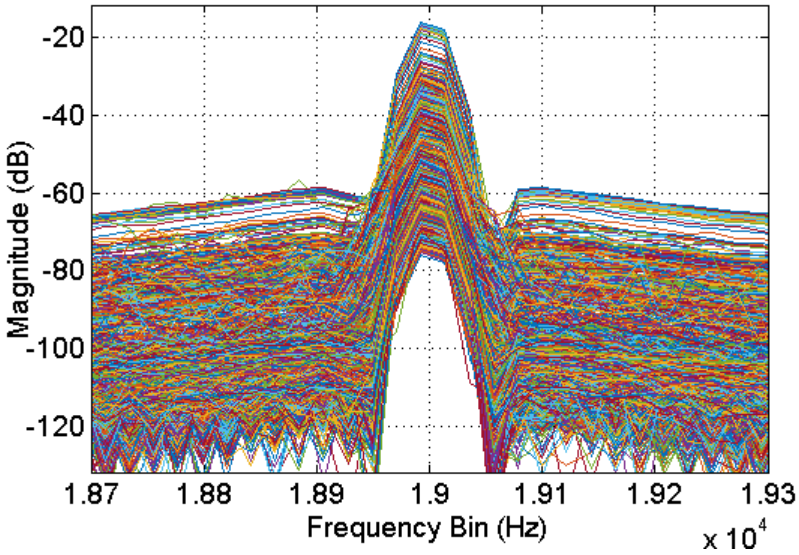


Fig. 8.7 All spectrums of audio signal frames: each line represents a spectrum of each frame

Fig. 8.6. We can see the signal drift severely interferes the audio spectrogram, displaying an unstable magnitude (*e.g.,* the part marked by the red ellipses).

To deal with this challenge, we collect 3,600 seconds 19kHz sound signal using three different mobile devices and then segment the signal into frames of 2,048-point length. As Fig. 8.7 shows, we plot the spectrum of 78,260 audio frames in the same graph. We can observe that, although the magnitude of the frequency bins for different frames show unpredictable signal excursions (*e.g.,* the magnitude in 19kHz bin spans from -83dB$\sim$-24dB), the relative magnitudes for every single sound frame are stable and robust to the time-elapse and device diversity (*i.e.,* each spectrum shows a similar shape). Because we intend to perceive the Doppler frequency shifts to infer hand gestures, we are more concerned about how the peak frequency bin changes over the time instead of absolute magnitude of each frequency bin. Based on this intuition, we normalize the magnitudes of frequency bins for each audio frame. Shown in the right graph of Fig. 8.6, after a simple FFT-based normalization, the audio spectrograms produced by waving hand from *Down to Up* show a stable and interpretable Doppler frequency shift and the signal drift is removed.

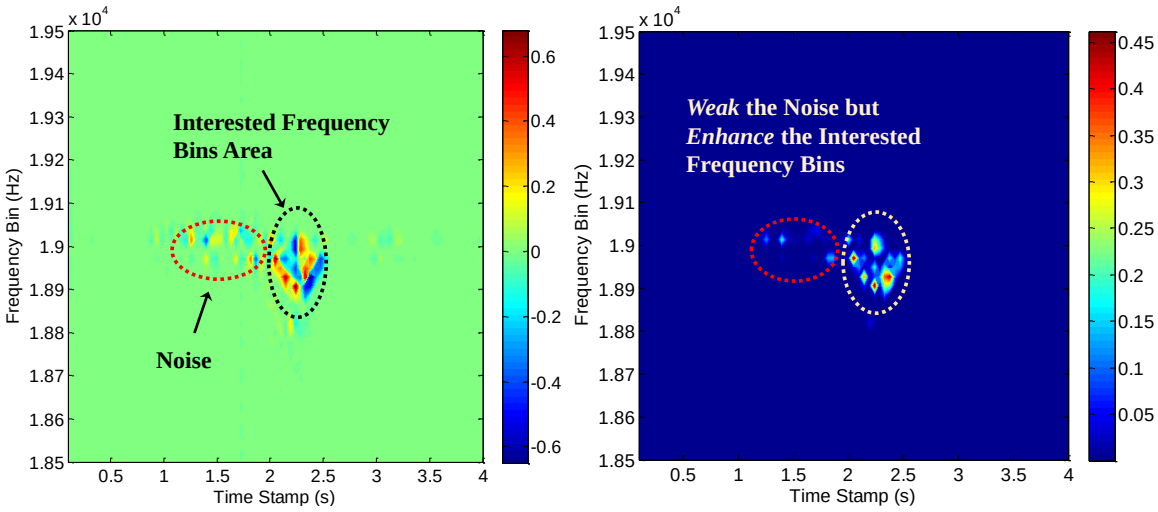


Fig. 8.8 Left Figure: the spectrogram after continuous frame subtraction; Right Figure: the spectrogram after the square calculation

8.5.2 Audio Signal Segmentation

Thus far, we have a denoised audio spectrogram which is robust to the temporal signal drift and device diversity. But we still need to figure out how to precisely segment the area where Doppler frequency shift happens.

Squared Continuous Frame Subtraction

To perceive the magnitude changes of frequency bins, we further conduct a *Squared Continuous Frame Subtraction*, in which we first subtract the normalized spectrum of current audio frame by previous frames and then square the magnitudes of frequency bins. The continuous subtraction essentially eliminates the static frequency bins and save the changed bins, shown as the left graph in Fig. 8.8 (*i.e.,* remove the unchanged 19kHz bin in Fig. 8.6 and highlight the changed frequency bins). The square calculation will further enhance the frequency-bin changes caused by hand's movement but weaken the bins due to the noise (see the right graph in Fig. 8.8, the noise marked by the red dot oval is further eliminated). Next, we need to accurately segment the frequency shift area based on those discrete frequency bins.

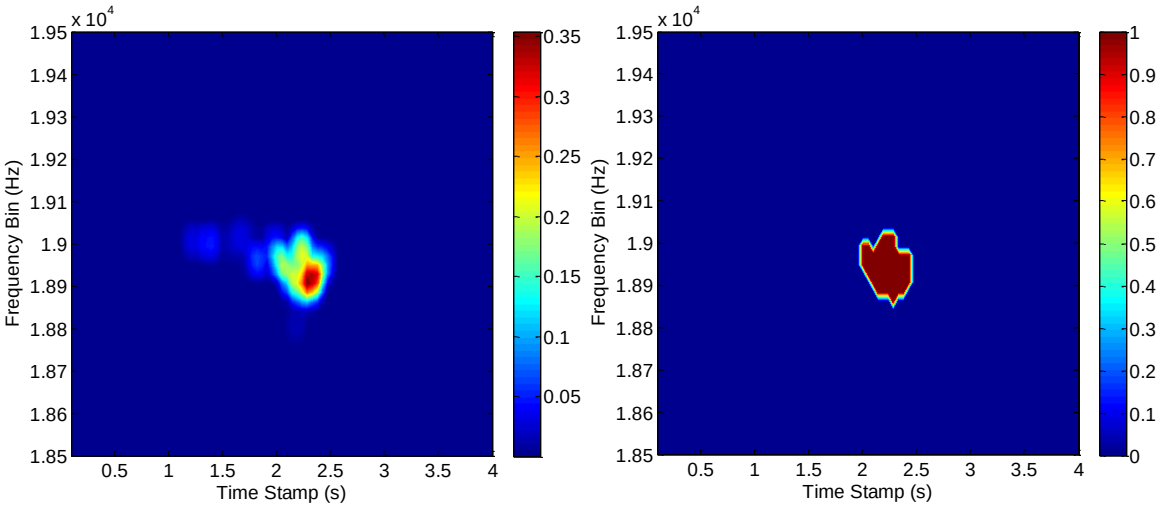


Fig. 8.9 Left Figure: the spectrogram after Gaussian Smooth Filter; Right Figure: the segmented area where Doppler Frequency shift happens

Gaussian Smoothing

Revisit the right graph of Fig. 8.8, the x-axis represents the time-stamps in a 0.046 second resolution, the y-axis indicates the frequency bins in Hz, the colors ranging from blue to red quantify the changing magnitude of frequency bins. Intuitively, we thereby can view such spectrogram graph as an image, then what we are interested is to connect those pixels and augment it into a zone. To do so, we introduce a Gaussian Smoothing method to blur the whole image. The Gaussian smoothing is a type of image-blurring filter that uses a Gaussian function for calculating the transformation to apply to each pixel in an image. In particular, each pixel's new value is set to a weighted average of that pixel's neighborhood. The original pixel's value receives the heaviest weight (having the highest Gaussian value) and neighboring pixels receive smaller weights as their distance to the original pixel increases. For our two-dimensional image, the following function is used for smoothing:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{8.3}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution. Intuitively, this formula produces a surface whose contours are concentric circles with a Gaussian distribution from the center point, which preserves boundaries and edges well. As the left graph in Fig. 8.9 shows, after Gaussian smoothing, those peak pixels are well augmented into a zone. Furthermore, we set a threshold ω to conduct the image binarization, *i.e.,* set the pixel value to zero if its value is less than ω, set the pixel value to one otherwise. As shown in the right graph of Fig. 8.9, we can successfully segment the frequency zone that Doppler shift happens. More de-noising and segmentation examples can be found in APPENDIX B.

8.5.3 Doppler Effect Interpretation

In this section, by using two typical hand-waving examples, we will interpret how a hand movement generates the shifted audio spectrogram based on the *motion law* of the hand movement.

From Eqn. 8.2, since $v_{sound} \gg v_{rad}$, we have

$$\Delta f = \frac{2 f_{sound} v_{rad}}{v_{sound}} \tag{8.4}$$

where $\Delta f = f_{received} - f_{sound}$. As Fig. 8.10 shows, assuming that hand moving path has θ_{hand} with the microphone and the hand moving speed is v_{hand}, we have

$$v_{rad} = v_{hand} \cos \theta_{hand} \tag{8.5}$$

Furthermore, we can derive the relation based on Eqn. 8.4 and Eqn. 8.5 as follows:

$$\Delta f = \frac{2 f_{sound} v_{hand} \cos \theta_{hand}}{v_{sound}} \propto v_{hand} \cos \theta_{hand} \tag{8.6}$$

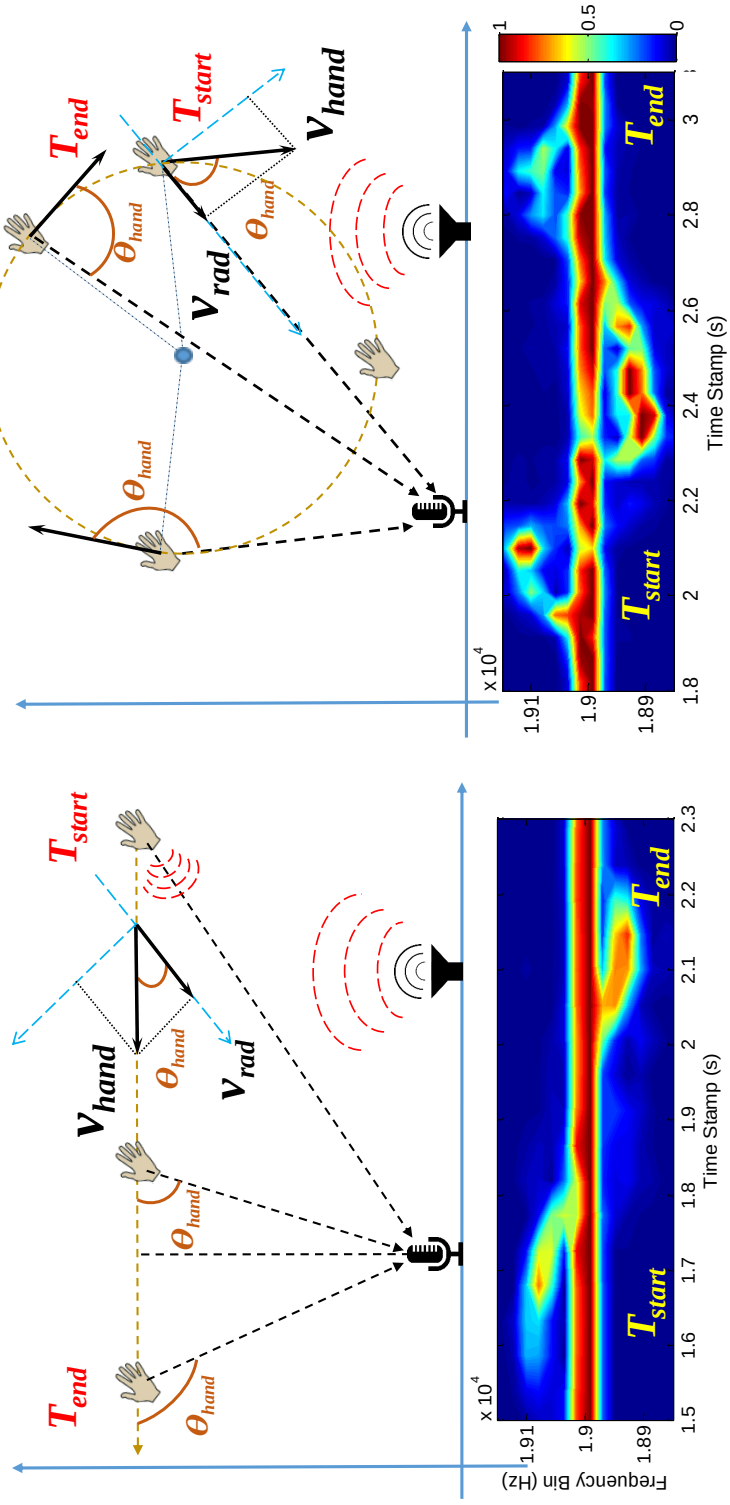


Fig. 8.10 The hand moving path with its generated audio spectrogram. Left Figure: hand moving from Right to Left; Right Figure: hand moving along Clockwise Circle

We take two examples[8] to interpret Eqn. 8.6, showing how we link real-time hand moving gesture with the audio spectrogram. As Fig. 8.10 depicts, when the hand moves from *Right to Left*, θ_{hand} gradually increases (*e.g.,* from $\pi/6$ to $\pi/2$ then to $2\pi/3$), hence the $\cos\theta_{hand}$ decreases[9] to 0, then to a negative value (*e.g.,* from $\sqrt{3}/2$ to 0, then to -1/2). As a result, the frequency shifts from high-frequency (*i.e.,* higher than 19kHz) to zero, then to low-frequency (*i.e.,* lower than 19kHz). For the most complicated case *clockwise circle*, the θ_{hand} first decreases from a certain angle to zero, then gradually increases from zero to π, and then decreases from π to the previous angle (*e.g.,* θ_{hand} experiences $\pi/3 \rightarrow 0 \rightarrow \pi/2 \rightarrow \pi \rightarrow \pi/3$ the right graph of Fig. 8.10). Thus, the audio frequency shifts towards high-frequency at first, then goes back to 19kHz, further moves to the low-frequency, then it goes back to zero, continuously moves to high-frequency[10].

8.5.4 Transforming Frequency Shift Area into Hand Velocity

This section will introduce how to estimate the real-time hand radial velocity based on the segmented frequency shift area. It should be noted that the peak bin locates in 19kHz under a no hand-waving environment (using $v_0 = 0$ represents such case). Based on Eqn. 8.6, we can model the frequency shift with real-time hand radial velocity as

$$\begin{aligned} f_{received}(t) - f_{sound} &= \frac{2f_{sound}}{v_{sound}} v_{hand}(t)\cos\theta_{hand}(t) \\ &= \frac{2}{\lambda_{sound}} v_{rad}(t) \end{aligned} \tag{8.7}$$

Furthermore, we can derive hand radial velocity $v_{rad}(t) = 0.5\lambda_{sound}(f_{shift}(t) - f_{sound})$.

[8] We choose two typical but complicated gestures for the interpretation.

[9] $\cos\theta$ is a monotony decrease function in $[0, \pi]$.

[10] Based on Eqn. 8.6, Δf actually is determined by both v_{hand} and $\cos\theta_{hand}$. And v_{hand} represents the hand speed (a nonnegative scalar), being zero at starting and ending point of hand moving, hence $\cos\theta_{hand}$ (ranging between -1 to 1, and traversing 0 multiple times) dominates the frequency shift.

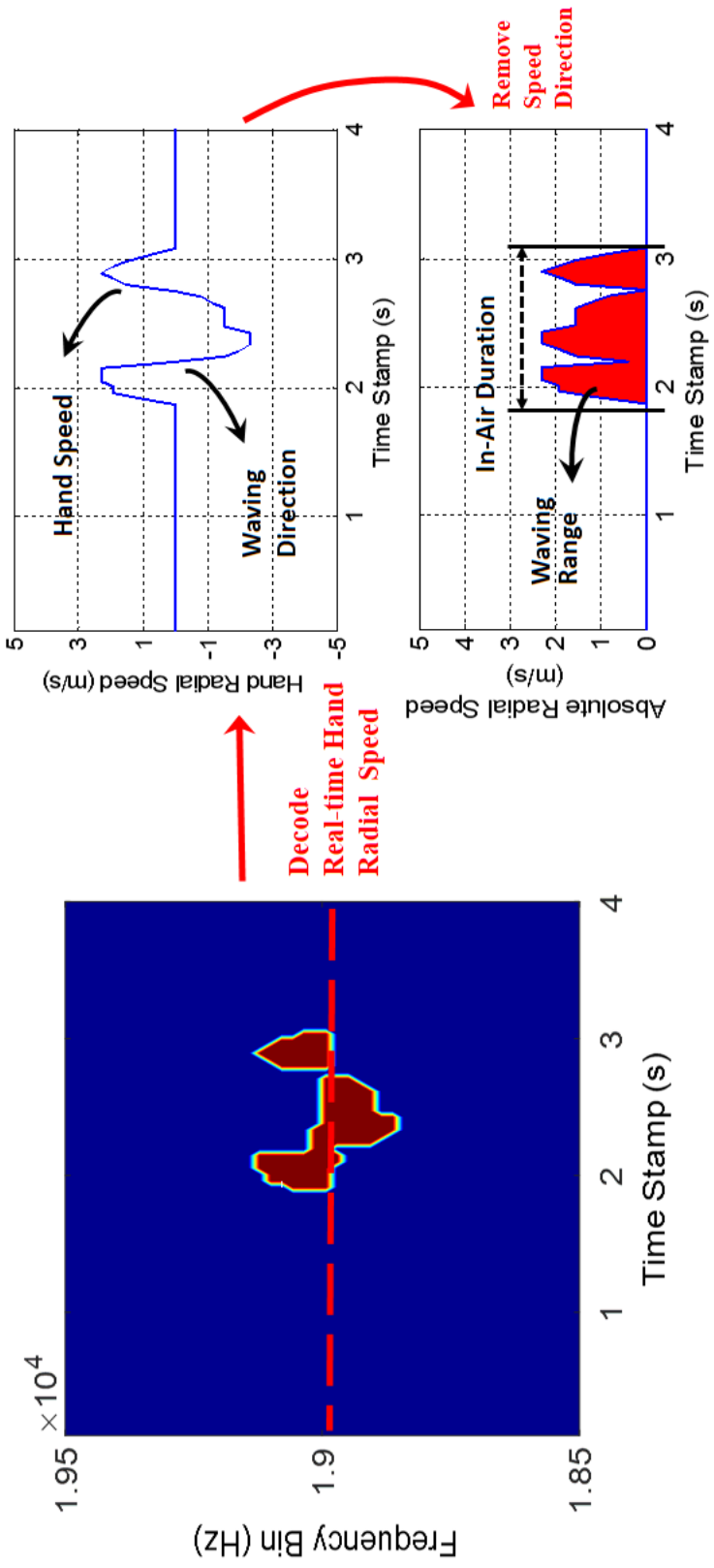


Fig. 8.11 The illustration of transforming frequency shifts into hand velocity, in-air duration and waving range

As the left graph of Fig. 8.11 shows, at each time-stamp, the length of frequency interval marked by red color represents $(f_{shift} - f_{sound})$. Therefore, we can estimate the real-time radial velocity of hand as shown in the right top graph in Fig. 8.11. Essentially, the sign of hand radial velocity indicates the hand moving direction (*i.e.,* hand gesture type), and the time interval of non-zero velocity represents the hand in-air duration. Also, we can measure the hand waving range based on the area covered by the velocity curve.

8.5.5 Gesture Recognition

In this section, we introduce in detail how we estimate the hand waving direction, speed and in-air duration as well as moving range given the hand radial velocity curve.

Recognizing the Waving Direction

Last section illustrates that how we link the hand moving directions with the audio spectrogram. Similarly, based on the direction changes of radial velocity (*i.e.,* whether its value is negative or positive, determined by $\cos\theta_{hand}$), we hence can estimate the angle ranges of the hand movement (*i.e.,* in angle categories: $[0,\pi/2]$ or $[\pi/2,\pi]$), as well as its corresponding time duration in each angle category. Based on a sequence of angle categories and its durations, we can further detect different gesture types. AudioGest adopts a rule-based method to infer the types of hand gestures. These rules are originated from the interpretation of *Doppler Effect*, which first exploit the frequency shifting direction to decode $cos\theta_{hand}$, then to further estimate θ_{hand}, *i.e.,* the hand waving direction towards the microphone. Finally, based on the hand waving direction sequence $\theta_{hand}(t)$, we estimate the hand waving directions.

We summarize the gesture recognition rules as follows: *up to down:* the angle of hand motion is in range $[0,\pi/2]$; *down to up:* the angle of hand motion is in range $[\pi/2,\pi]$; *right to left:* the angle of hand waving is $[0,\pi/2] \rightarrow [\pi/2,\pi]$ and the time duration in $[0,\pi/2]$ is longer than in $[\pi/2,\pi]$; *left to right:* the angle of hand motion is $[0,\pi/2] \rightarrow [\pi/2,\pi]$ but the

time duration in $[0, \pi/2]$ is shorter than in $[\pi/2, \pi]$; *anticlockwise circle:* the angle of hand motion is from $[\pi/2, \pi] \rightarrow [0, \pi/2] \rightarrow [\pi/2, \pi]$; *clockwise circle:* the angle of hand motion is $[0, \pi/2] \rightarrow [\pi/2, \pi] \rightarrow [0, \pi/2]$. More hand gesture recognition examples can be found in Appendix C. It is noted that many hand-gestures recognition systems highly depend on semi-supervised/supervised machine learning methods [190]. Our AudioGest system does not need to collect labeled training data to train a classifier.

Estimating Waving Duration and Speed

For estimating the hand in-air duration, we can directly measure the time interval that hand radial velocity is not equal to zero (*e.g.,* the time length marked by dot-line in Fig. 8.11). Then the remaining problem is how we measure the average hand moving speed. Please note that the velocity curve we estimate is the hand radial speed (towards the microphone) instead of the real hand moving speed that we are interested in[11]. In this chapter, as aforementioned, we aim to first recognize different hand gestures, then to be able to distinguish different hand speed, in-air duration and moving range to provide more control commands for serving various applications. Hence, for a same gesture type, we want to evaluate if the hand is in slow, medium or fast speed (see Fig. 8.5).

In particular, we first transfer the hand velocity (with moving direction) into a speed (ignore the direction), the transformation shows as the right-top graph to the right-bottom graph in Fig. 8.11. We observe that, for the same gesture with different speeds, θ_{hand} actually experiences a same angle range (*e.g.,* $\pi/6 \rightarrow ... \rightarrow \pi/2 \rightarrow ... \rightarrow 2\pi/3$: moving from right to left as in the left graph of Fig. 8.10) but in different timestamps. As a result, according to Eqn. 8.5, we can infer that $E(V^1_{hand}) > E(V^2_{hand}) \iff E(V^1_{rad}) > E(V^2_{rad})$, where $V^1_{rad} = \{v^1_{rad}(t_1), v^1_{rad}(t_2), ...\}$ represents the first sequence of hand radial speed we estimated,

[11] Theoretically, with a single microphone, we cannot estimate the moving velocity of hand since we cannot accurately measure the angle between hand and microphone. To do so, we at least need two microphones which will leave to our future work.

V^2_{rad} indicates the second sequence of hand radial speed[12]. Hence we define a *speed-ratio* to evaluate the relative magnitude for different hand speeds. Assuming that the time interval between two adjacent timestamps is T (*e.g.,* 0.0464s using a 2048-point frame), the hand waving duration is $t_{waving} = nT$, then we calculate the *speed-ratio* as

$$S_{ratio} = \frac{E(v_{rad}(t))}{E(v^0_{rad}(t))} = \frac{\frac{1}{n}\sum_{i=1}^{n} v_{rad}(iT)}{E(v^0_{rad}(t))} \tag{8.8}$$

where $E(*)$ means expectation or mean value; $v^0_{rad}(t)$ represents a baseline of the hand moving speed set as $E(v^0_{rad}(t)) = 1$ for simplicity[13]. Hence, we have $S_{ratio} = \frac{1}{n}\sum_{i=1}^{n} v_{rad}(iT)$, namely the mean value of our estimated radial-speed. Intuitively, a bigger S_{ratio} represents a faster hand movement.

Estimating Waving Range

Similar to the waving speed, we cannot estimate exactly how much distance the hand moves using one microphone. By inheriting the idea in evaluating the waving speed, we also define *range-ratio* to measure the relative magnitude of hand waving range:

$$R_{ratio} = \frac{R_{rad}}{R^0_{rad}} = \frac{\sum_{i=1}^{n} T v_{rad}(iT)}{R^0_{rad}} = \frac{nTS_{ratio}}{R^0_{rad}} \tag{8.9}$$

where R^0_{rad} represents the baseline of hand waving range that we assume equals to 1. Hence we can compare the hand waving ranges using $R_{ratio} = nTS_{ratio}$ (*i.e.,* the area of the zone covered by red color in Fig. 8.11), where n and S_{ratio} is the estimated hand in-air duration and speed-ratio. In APPENDIX C, we illustrate the examples of estimating hand waving speeds, in-air time and waving ranges.

[12] Essentially, V^1_{rad} and V^2_{rad} represent two different moving speeds for a same certain hand-gesture type.
[13] We can definitely find a certain hand waving meets such requirement.

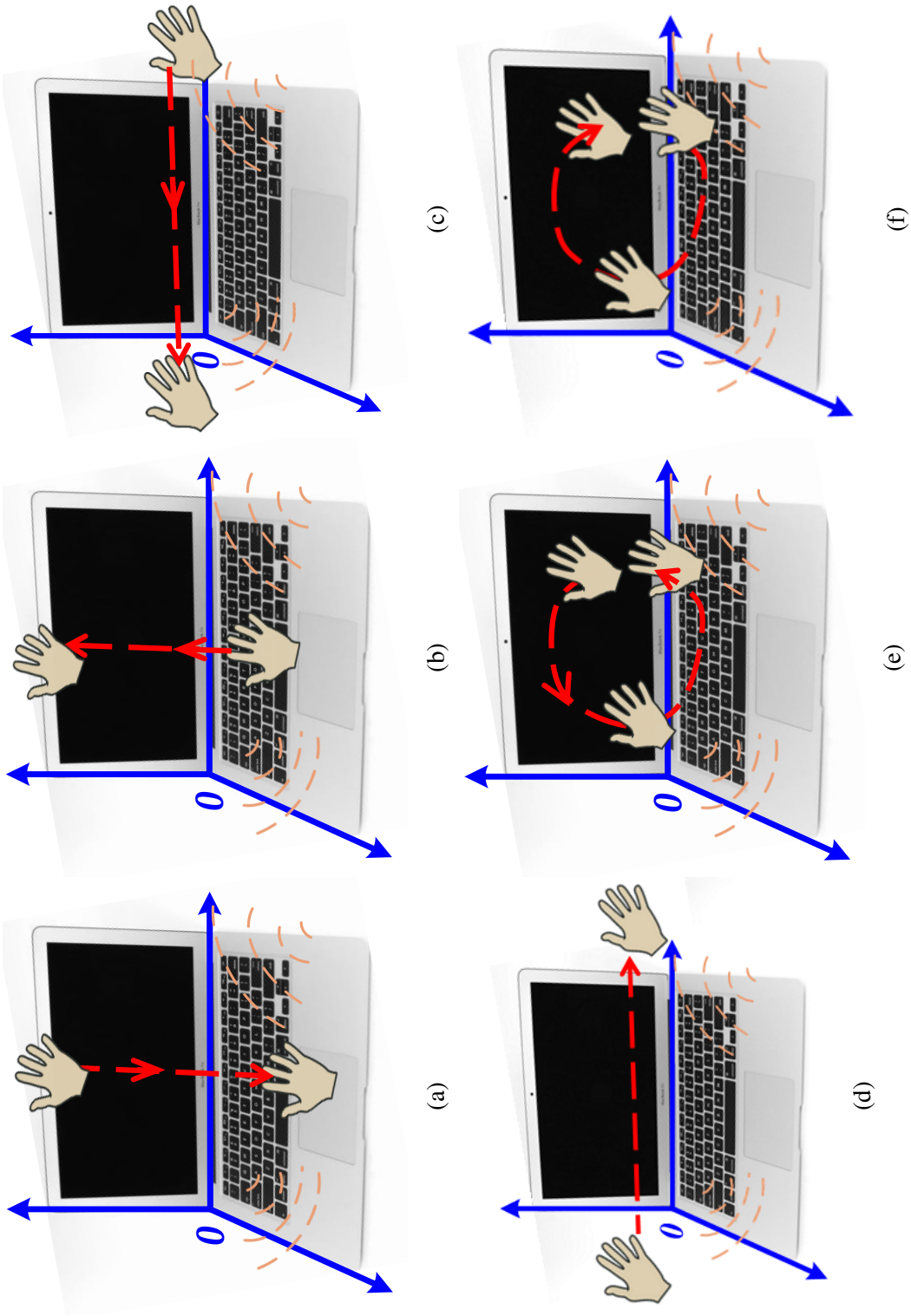


Fig. 8.12 Six hand waving scenarios: (a) Up-to-down hand waving; (b) Down-to-up hand waving; (c) Right-to-left hand waving; (d) Left-to-right hand waving; (e) Anticlockwise hand circling; (f) Clockwise hand circling

Algorithm 5: Pseudocode of AudioGest system

Input: Initializing System; Setting System Parameters
Result: Hand Speed-Ratio, Hand In-Air Time,
Hand Waving Range-Ratio and Basic Gesture Type

```
Speaker emits 19kHz sinuous sound-waves
Microphone records sound signals
while AudioGest not turn off do
    Reset Sound Frame Size
    while Sound Frame Size < FrameSizeParameter do
        wait
    end
    Do FFT Normalization based on Section 8.1      // remove the signal
     drifts caused by device deversity and time elapse
    Do Squared Continuous Frame Subtraction based on Section 6.2.1 // enhance
     frequency shifts and weak background noise
    Do Gaussian Smoothing based on Section 8.2.2     // connect shifted
     frequency-bins and augment it into a zone
    Segment Frequency Shift Area based on Section 8.2.2   // set threshold
     to conduct image binarization
    5 if Shift Area Size < AreaSizeParameter then
        No Hand Motion Detected
        continue       // eliminate false frequency shifts not
         caused by hand's motion and go to next loop
    else
        Estimate Hand's Basic Gesture Types based on Section 8.5.1
         // recognize gesture by decoding the direction
         squence of hand towards the microphone
        Estimate Hand In-Air Duration based on Section 8.5.2
        Estimate Hand Speed-Ratio based on Section 8.5.2   // transfer hand
         velocity into a speed without the direction
        Estimate Hand Waving Range-Ratio based on Section 8.5.3
    end
end
Speaker stops
Microphone stops
```

Putting Things Together

Now we put all the pieces together and concretely illustrate how our AudioGest system works through the pseudo-code, shown as Algorithm 5.

Fig. 8.13 The three mobile devices used for testing

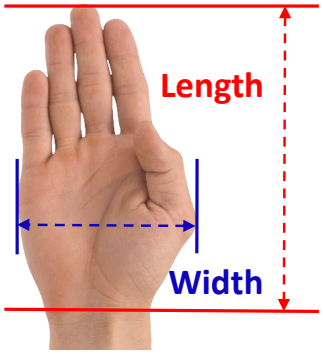
8.6 Evaluation

We start with micro-benchmark experiments in a lab environment and then conduct the in-**situ** tests in four real-world places - Living Room, Bus, Cafe, and HDR Office. We conduct the testing on three typical mobile devices: laptop (MacBook Air laptop), tablet (GALAXY Tab-2 tablet), and mobile phone (GALAXY S4 smartphone) without any hardware modification. We name the three devices as *D1*, *D2* and *D3* for simplicity.

8.6.1 Hardware

For the MacBook Air laptop, we run AudioGest on the computer using Audio System Toolbox[14][15]. For the GALAXY tablet and smartphone, we design the AudioGest system in the Simulink8.6 that provides a library of Simulink blocks for accessing the devices speaker and microphone[16].

[14] mathworks.com/hardware-support/audio-ast.html
[15] developer.apple.com/library/mac/documentation/MusicAudio/ Conceptual/CoreAudioOverview/
[16] mathworks.com/hardware-support/android-programming-simulink.html



User ID	Gender	Age	Hand Length	Hand Width
User 1 (U1)	Male	29	17.1 cm	9.2 cm
User 2 (U2)	Female	29	16.4 cm	8.5 cm
User 3 (U3)	Male	27	18.5 cm	10 cm
User 4 (U4)	Female	13	14.7 cm	7.5 cm
User 5 (U5)	Male	23	17.4 cm	9.5 cm

Fig. 8.14 The illustration of handsize measurement and participant information

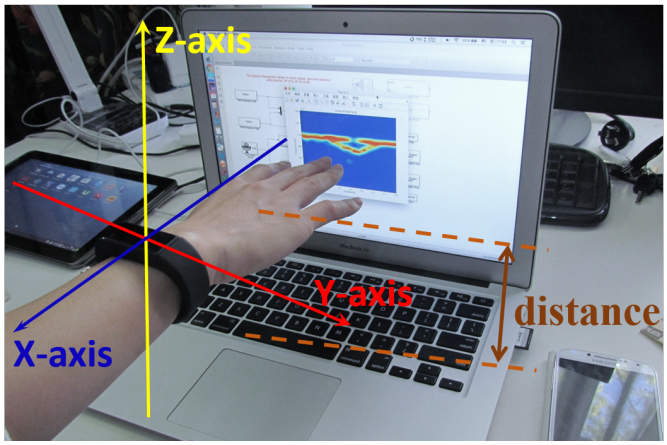


Fig. 8.15 The 3-axis accelerometer in smartwatch

8.6.2 Testing Participants

Five participants join the experiments. AudioGest decodes the hand gesture via analyzing the reflected audio signal from hands. Intuitively, a bigger hand generates a stronger echo signal. Thus we measure the handsize of each participant. Figure 8.13 shows the hardware devices used in the experimental studies. The five users are marked as *U1*, *U2*, *U3*, *U4* and *U5*.

8.6.3 Collection of Ground Truth

We use the 3-axis MEMS accelerometer in a smart-watch for collecting the ground truth. Generally, the 3-axis accelerometer records acceleration readings along three orthogonal

axises. We set the sampling rate as 24Hz. In this chapter, we decode two types of hand gestures: *i)* linear movement (e.g., waving from *up to down* or *left to right*); and *ii)* circle movement (e.g., waving in *clockwise circle* or *anticlockwise circle*). For the first case, we measure the acceleration of the corresponding direction (remove the gravity if in z-axis, same goes the followings) to calculate the hand in-air time, average hand speed (*i.e.,* $\bar{v} = 1/2at$) and waving range (*i.e.,* $r = 1/2at^2$), then we set a same baseline of waving speed and range as AudioGest to calculate the speed-ratio and range-ratio. For the second case, we keep the hand downward and do the circling movement. Then we can estimate the total acceleration based on the recorded three ones (*i.e.,* $a_{total} = \sqrt{a_x^2 + a_y^2 + a_z^2}$). Finally, we conduct the same calculation to get the ground truth.

8.6.4 Evaluation Metrics

We adopt four typical evaluation metrics to evaluate our methods: *i)* Detection Rate (or True Detection Rate): the ratio of correctly detected hand gesture to overall testing hand gestures, measuring whether our system can efficiently detect a hand gesture when a hand waving happens; *ii)* False Detection Rate: the ratio of wrongly detected hand gestures to overall detected hand gestures, evaluating whether our system is too “sensitive" by recognizing a non-handgesture as a hand gesture; *iii)* Gesture Classification Accuracy: the rate that system can correctly classify the gesture type among all the detected hand gestures; *iv)* Detection Accuracy: the rate that system can correctly classify the gesture types as well as the categories of the in-air duration, average speed and waving range.

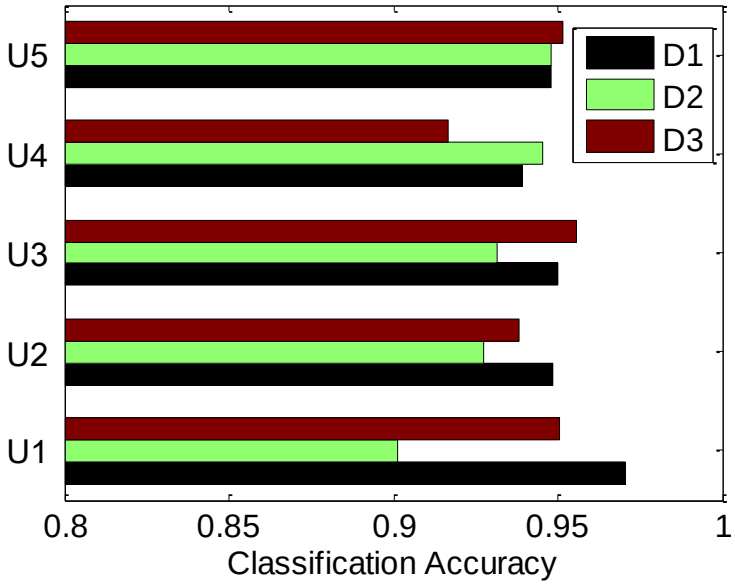


Fig. 8.16 The average gesture classification accuracy for different mobile devices and users

8.6.5 Micro-Test Benchmark

We conduct some micro-benchmarks in a lab environment. We ask the five participants to perform each hand gesture 30 times for each device[17], hence we test 2,700 hand gestures by collecting around 4.5 hours audio data.

Gesture Recognition

Fig. 8.16 shows the gesture classification accuracies of five users for three devices. AudioGest achieves 94.15% gesture type recognition accuracy. In particular, subject U5 can get average 95% accuracy, but U1 achieves 90.15% mean accuracy using the tablet. From its confusion matrix (shown in Fig. 8.17), we can observe that most errors happen in distinguishing *Right-Left/Front-Behind* and *Left-Right/Behind-Front*. Detecting the hand gestures is done by decoding the hand-microphone angle sequence and its corresponding duration. For device D1 (*i.e.,* MacBook Air laptop), its microphone locates in the left side, which results in different

[17]The participants can freely wave with any speed or range, but have to be within the category of the defined gesture types. The collection time spans over two weeks based on their available time. We also require the minimum time-interval of two hand gestures is $> 1s$.

Actual Gestures \ Classified Gestures	Up-Down	Down-Up	Right-Left/ Front-Behind	Left-Right/ Behind-Front	Clock-wise	AntiClockwise	Non-Action
Up-Down	434	0	5	2	1	0	8
Down-Up	0	430	8	9	1	2	0
Right-Left / Front-Behind	3	1	396	23	7	8	12
Left-Right / Behind-Front	1	4	19	411	0	4	11
Clockwise	3	2	6	13	426	0	0
Anti-Clockwise	2	1	8	14	0	425	0

Fig. 8.17 The Confusion Matrix for the gesture classification

duration time of two angle categories for *Right-Left* and Left-Right waving. But we cannot distinguish hand waving from *Front-Behind* or *Behind-Front* due to the block of the computer screen. However, for D2 and D3 (*i.e.,* Galaxy tablet and smartphone), their microphones locate in the bottom of the device, which substantially enable *Right-Left* and Left-Right hand movement generating the same angle category sequence (*i.e.,* $[0, \pi/2] \rightarrow [\pi/2, \pi]$) and roughly same durations. Hence we cannot distinguish such two directions, but we can recognize the *Front-Behind* or *Behind-Front*. Due to the same reason, for recognizing *Right-Left/Front-Behind* and *Left-Right/Behind-Front*, we can only depend on the difference of angle durations, making it less reliable as other directions. Moreover, to better illustrate the idea of multi-modal hand detection, we depict several real-world examples in APPENDIX C.

Waving Attributes Estimation

Fig. 8.18-8.20 show the results of estimation errors[18] of the hand in-air duration, moving speed-ratio and range-ratio respectively. The bar charts indicate both average error and its standard derivation. In particular, AudioGest can estimate the three gesture context

[18]Namely, the distance between estimated value with the ground truth (≥ 0).

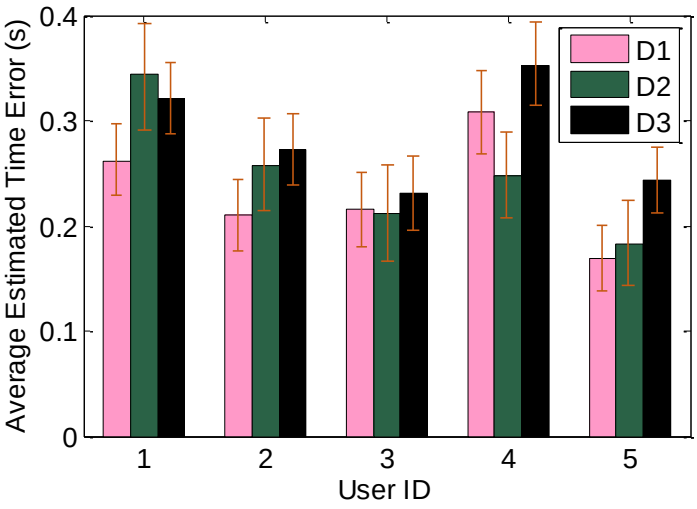


Fig. 8.18 The hand in-air duration estimation error for different mobile devices and users

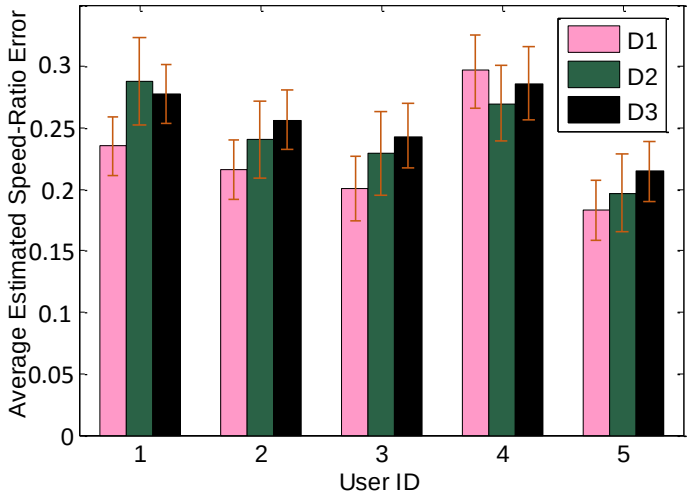


Fig. 8.19 The average speed-ratio estimation error of hand moving for mobile devices and users

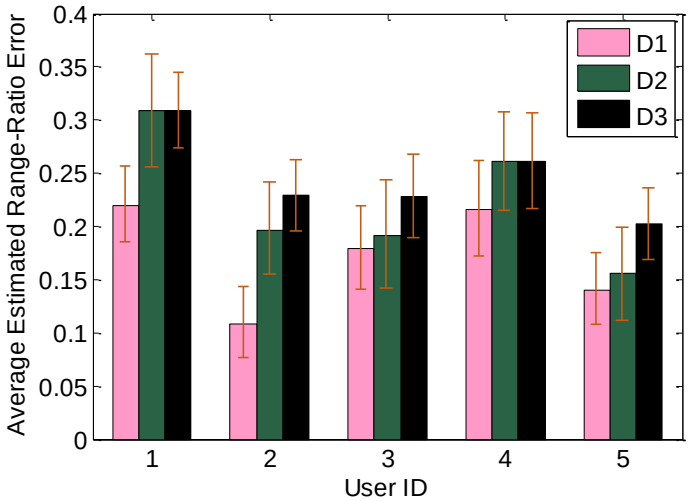


Fig. 8.20 The average range-ratio estimation error of hand moving for different users

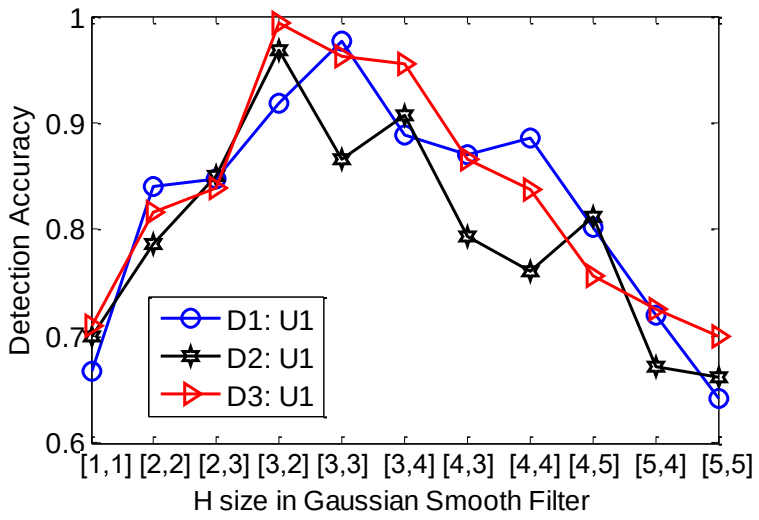


Fig. 8.21 The gesture detection accuracy with parameter *H-size*

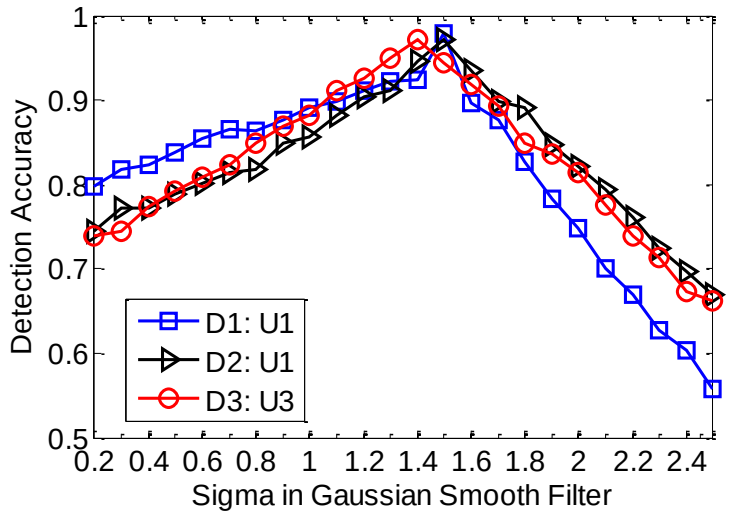


Fig. 8.22 The gesture detection accuracy with parameter σ

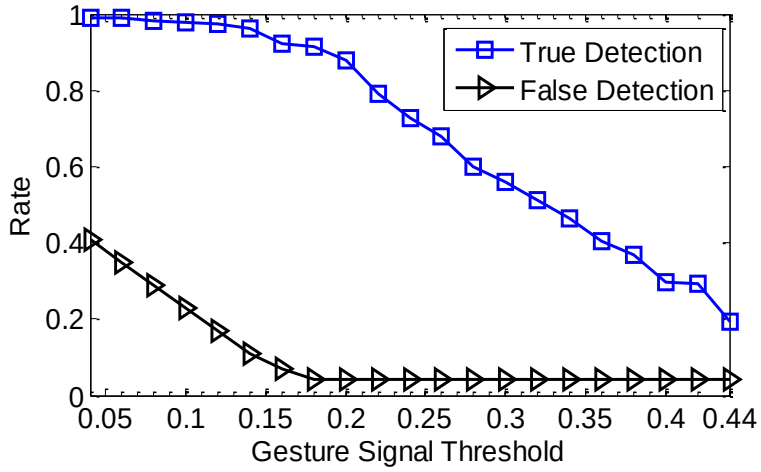


Fig. 8.23 The gesture detection accuracy with gesture signal threshold

Table 8.1 Calculation time and resolution vs. frame sizes

Frame Size	Resolution of FFT (Hz)	Calculation Time (s)	Resolution of Speed (m/s)	In-Air Time Resolution (s)
256	172.27	2.767	3.110	0.0058
1024	43.07	0.733	0.777	0.0232
2048	21.53	0.396	0.389	0.0464
4096	10.77	0.226	0.194	0.0929
8192	5.38	0.134	0.097	0.1858

information with average 0.255s in-air duration, 0.242 speed-ratio and 0.2138 range-ratio error respectively. It is worth to mention that, among 5 subjects, U5 achieves a better result in both the gesture classification and the context estimation, which mainly lie in the fact that U5 has a slightly bigger hand, which enhances the audio signal reflection.

Parameters Chosen

Fig. 8.21-8.23 illustrate how three key parameters influence the performance of our system. The parameter *H-size* specifies the number of rows and columns used in the gaussian filter (*i.e.,* $H_{size} = [x, y]$ in Eqn. 8.3). We test overall 11 different H-size when $[x = 3, y = 2]$ performs better. Parameter σ indicates the standard deviation in Gaussian function, which achieves the best accuracy at $\sigma = 1.5$. The last parameter *Gesture-Signal Threshold* determines whether a shift happens in a frequency bin, which plays an important role in AudioGest. We can see that the higher the value is, the more both true detection and false detection rates decrease. Hence we choose *Threshold = 0.16* to balance such two detection rates.

As Table 8.1 shows, we also measure the FFT resolution, calculation time, speed, and in-air time detection resolutions by using different signal frame sizes[19]. We find that for a smaller frame size, we need to calculate more FFTs within a second and get a smaller frequency bin, which in turn produces a finer speed resolution but a coarser time resolution.

[19]We apply FFT with a same length as the signal frame, *e.g.,* for a frame size of 256 samples, we adopt an FFT with 256-points

To balance the speed and time resolution as well as to maintain a reasonable calculation time, we choose 2,048 as the frame size and as the FFT points. Note that the speed resolution is also equivalent to the lower-boundary of hand speed that we can detect (*e.g.,* if the hand speed is extremely slow such as less than $0.389m/s$, our HGR system cannot detect it). However, our system focuses on a multi-modal hand gesture recognition, in which we categorize the hand speed into three levels: slow, medium, and fast (see Fig. 8.5). A speed resolution of 0.389m/s is accurate enough to serve the purpose of this system because this resolution has a good trade-off among the calculation time, speed resolution and time resolution, especially it can filter out some false alarms caused by finger movements (those movements usually produce gentle frequency shifts which can be captured by a sensitive speed resolution).

Please note that we can also use a 4,096-point frame size that can reach $0.194m/s$ speed resolution for a more fine-grained hand gesture detection (*e.g.,* we can categorize the hand-speed into 4 or more ranges so that HGR system can provide more control commands). The choice of frame size mainly depends on the real-world applications (*e.g.,* whether it requires a smaller delay, more fine-grained speed and in-air time detection) and the calculation capacities. The decision also relates to the sampling rate that a mobile device can support. For example, if the hardware supports a higher sampling rate (*e.g.,* 192kHz in SUMSUANG Galaxy S6 smart-phone), we can choose 1024-point or even 512-point frame size to achieve a better or comparable speed resolution as 2048-point size in 44.1kHz sampling rate but with a better time resolution. In AudioGest, for generality, we set its sampling rate as 44.1kHz. With this sampling rate, we choose 2,048-point frame size, which is acceptable for multi-module hand-gesture detection.

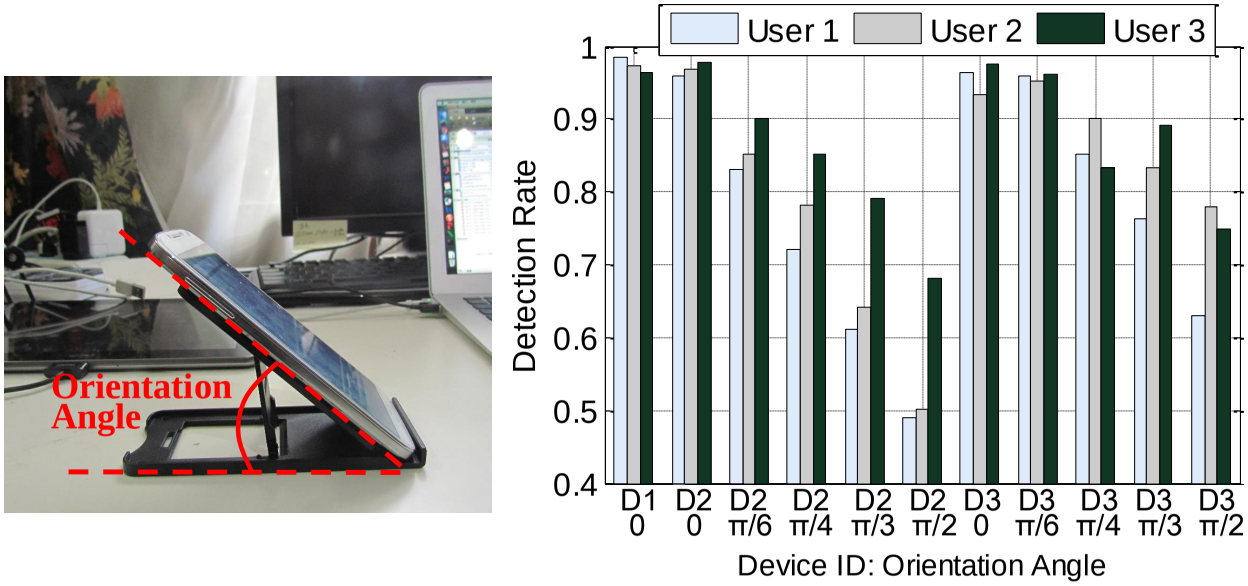


Fig. 8.24 The device orientation angle with its detection accuracy

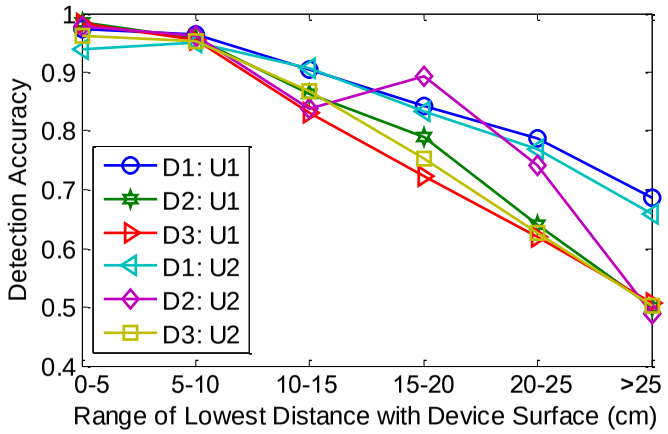


Fig. 8.25 The device-hand distance with its detection accuracy

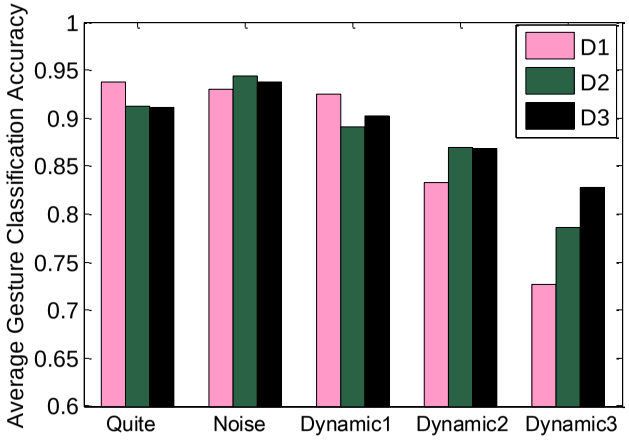


Fig. 8.26 The average detection accuracy for different scenarios

System Robustness

We evaluate the robustness of AudioGest in four ways:

- *Orientation Angle*: as Fig. 8.24 shows[20], AudioGest performs well when the orientation angle is less than $\pi/4$. Under a $\pi/2$ circumstance, its accuracy greatly decreases to around 60%, which we will leave for further work.

- *Hand-Device Distance*: we test the system when the hand waves in different categories of hand-device distance[21]. AudioGest achieves satisfied accuracy when the distance is below 10cm (which is the typical using scenario for most users). We also observe its performance decreases when the hand waves in a far distance from the device (the COTS microphone cannot capture the echo-sound due to its capability limitation).

- *Environmental Motion*: as Fig. 8.26 shows, we test our system under five environmental motion circumstances - Quiet (no audible noise and human motion), Noisy (playing music loudly), Dynamic1 (with human walking back and forth in around 4 meters away the device), Dynamic2 (with human walking back and forth in around 2 meters away) and Dynamic3 (with human walking back and forth nearby, around 0.5 meter). We can see AudioGest works well under first three cases (especially, it is nearly unaffected by human noise).

- *Time Elapse*: we also test its performance under different elapsed time periods - 6 hours, 1 day, 3 days, 1 week, and 10 days, without tuning parameters. We conduct a comparison experiment to study the performance of the system adopting and not adopting the proposed signal denoising method (*i.e.,* FFT normalization). In other words, the audio data collected under a same experiment (*i.e.,* same participants, gestures and same mobile devices under a same testing surrounding environment) is

[20] we mainly test D2 and D3 from 0 to $\pi/2$, since laptop normally lie flat on the surface.

[21] It is difficult for us to accurately control/measure how hand close to the device while waving, but controlling the lowest hand-device distance within a range is possible.

fed into two HGR systems – one contains the denoising processing and the other does not. As the results shown in Fig. 8.27, by applying FFT normalization, AudioGest achieves about 35% to 70% performance increase when dealing with the signal drifting challenge, which demonstrates the effectiveness of our denoising approach.

In summary, AudioGest performs accurately under normal circumstance, and is robust to the human noise and signal drifting.

8.6.6 In-suit Experiments

Fig. 8.28-8.30 show the system performance in some typical daily-living environments. Two subjects (U1 and U2) participate in the test. We ask the subjects to use three mobile device in a living room ($5m \times 3.5m$), on a bus, in a Cafe, and in an HDR (Higher Degree by Research) space (around $15m \times 10m$, contains > 20 students). We collect 1,200 hand gestures (Living Room: 360, Bus: 240, Cafe: 240, HRD Space: 360). The in-situ testing spans around two weeks upon participants' time availability. Under the living room and HDR office, AudioGest performs similarly to our micro-benchmark since such testing scenarios are usually with less environmental motion inferences. When coming to the bus (the most dynamic environment but also where people usually use the mobile devices), the performance is degraded to an average 89.67% accuracy, and the segmentation (*i.e.,* hand in-air duration) and speed-ratio accuracy also decrease, which is mainly caused by the narrow space and unpredictable motion influences on the bus.

To summarize,the results from both micro-benchmark and in-suit experiments suggest that *AudioGest* provides an enabling primitive that can device-free recognize hand gestures, as well as accurately estimate hand movement speed and range.

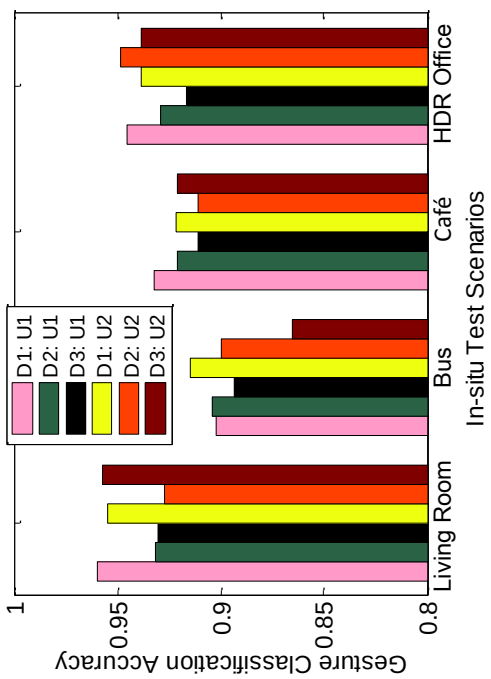


Fig. 8.28 The average gesture classification accuracy for in-suit test

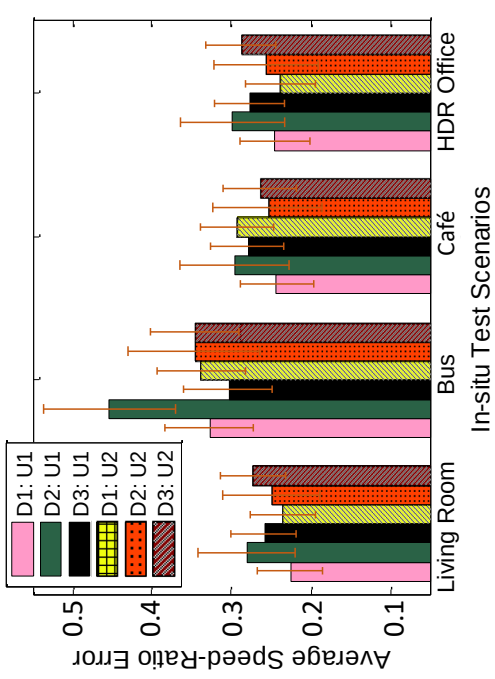


Fig. 8.30 The average speed-ratio estimation error of hand movement for in-suit test

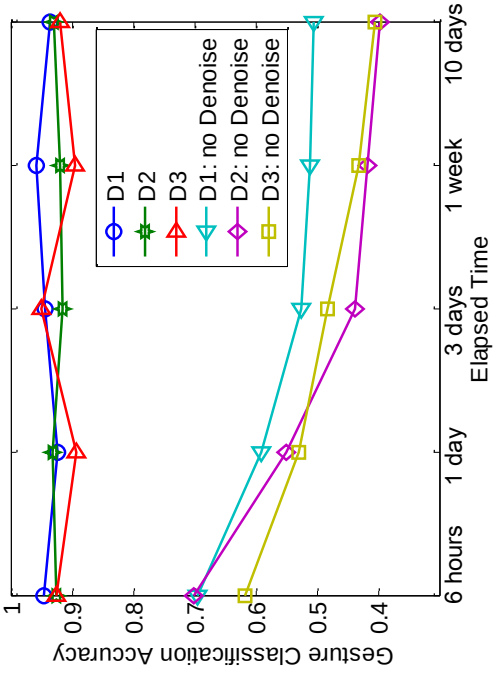


Fig. 8.27 The detection accracy with and without denoising

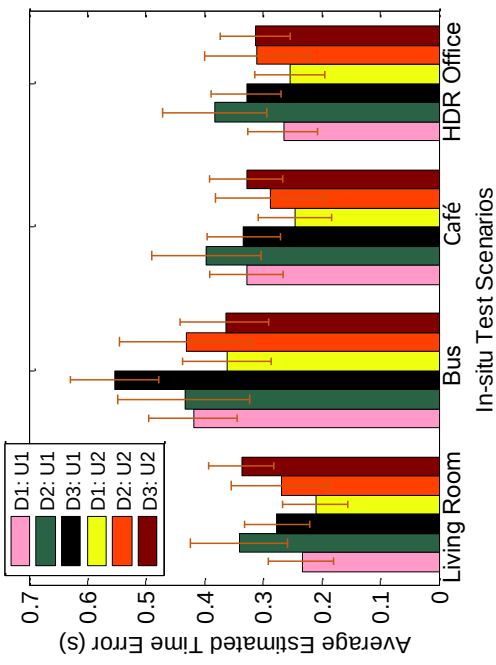


Fig. 8.29 The average estimation error of hand in-air duration for in-suit test

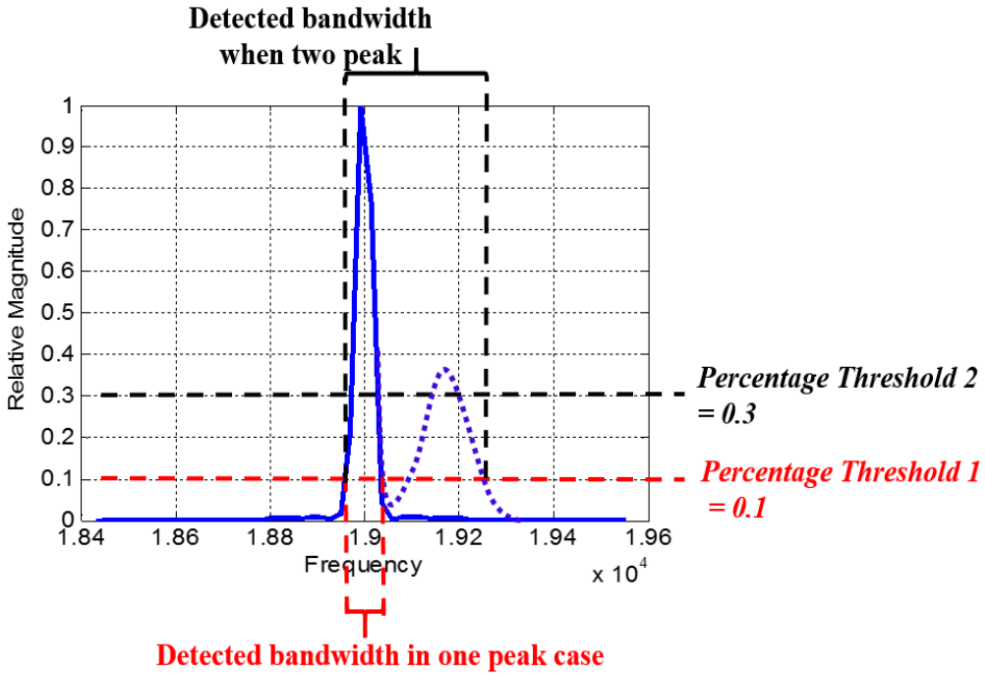


Fig. 8.31 SoundWave detects the frequency shift based on a percentage-threshold method. *For one peak case, it detects the bandwidth of the amplitude drops below 10% of the tone peak. For a large frequency shift casing two peaks, it performs a second scan (if the second peak $\geq$ 30%) and repeats the first scan to find the bandwidth drops from the second peak.*

8.6.7 Comparing with the State-of-the-Art

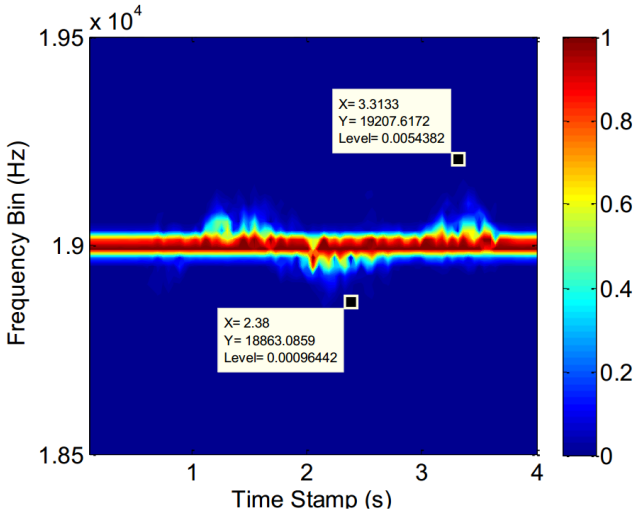
This section compares our AudioGest with seven state-of-the-art HGR systems in terms of detection mechanism, hardware, testing environment, system training requirement and detection capacity/resolution as well as the accuracy, shown in Table 2.1. Briefly, except for SoundWave [111], other HGR systems mainly exploit *Radio Frequency* (RF) signals to recognize hand motions. Those RF signals are either from COTS or modified WIFI and GSM infrastructures (*e.g.,* WiGest [103], WiSee [107] and SideSwipe [196]), or radars (*e.g.,* FineGesture [108] and RadarGesture [197]), or generated by specialized hardwares (*e.g.,* AllSee [109]). While bearing many advantages, they are either built upon extra hardwares or available WIFI signals, which may be impractical under some circumstances (see discussions in Sec. 1).

Unlike the above HGR systems, SoundWave is one pioneering work to exploit the Doppler effect of sound wave reflected by hands, sharing the same hand gesture recognition mechanism as AudioGest. SoundWave can recognize five hand gestures: Two Handed Pull, Back Flick, Quick Taps, Slow Taps, with an average 94.5% accuracy. It mainly adopts a percentage-threshold based dynamic peak tracking method to capture the frequency shifts. Different from SoundWave, our system achieves a multi-modal hand gesture recognition, which not only can recognize basic hand gestures but also aims to quantitatively measure the hand waving speed, rang and in-air time (see Fig. 8.5). More importantly, we provide a mathematical model for interpreting Doppler Effect into hand motion (see Sec. 8.2.1 and 8.5.4) by linking the equation of Doppler Frequency shift and Newton's law of motion of hand gestures. As far as we know, neither SoundWave nor other HGR systems can achieve this.

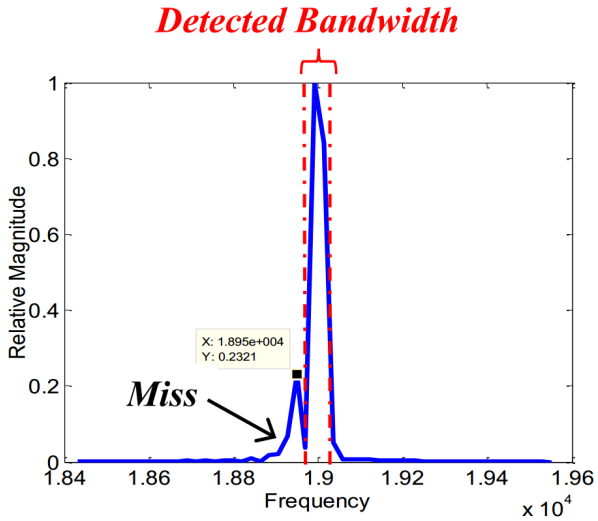
Moreover, since SoundWave cannot achieve a multi-module hand gesture detection, and gesture types, testing experiments, and participants are also very different, it is hard to conduct fair benchmark experiments. We thus compare our work with SoundWave and other HGR systems in a high-level of view, shown in Table 2.1. Accurately detecting the frequency shifts is the foundation of HGR systems based on Doppler Effect. Without a good performance in capturing frequency shifts, both our system and Soundwave cannot achieve an accurate hand gesture recognition. To this end, we compare AudioGest with SoundWave by two experimental cases in terms of the performance of detecting frequency shifts. Fig. 8.31 depicts how SoundWave detects the bandwidth of shifted frequency. When four or more FFT frames (i.e., 2048-point segmentation) in succession are detected with frequency shifts, SoundWave will consider a hand motion is happened.

Table 8.2 Comparison of typical device-free HGR systems

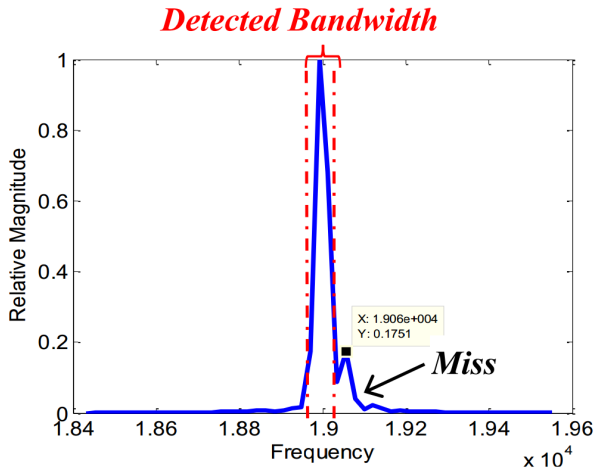
Comparison Items	WiGest [103]	FineGesture [108]	AllSee [109]	SoundWave [111]
Measured Signal	RSSI	RSS, CSI, Phase	RF signal	Audio
Need extra hardware?	No	Yes	Yes	No
Test in dynamic environments? (*e.g.*, bus)	No	Yes	No	No
Need training?	No	Yes (kNN)	No	No
Sense gesture contexts? (*e.g.*, speed, range)	Yes (speed)	No	No	No
Accuracy / Gesture Resolution	96% / 36	92% / 25	97% / 8	94.5% / 5
Comparison Items	SideSwipe [196]	RadarGesture [197]	WiSee [107]	*AudioGest*
Measured Signal	GSM signal	FMCW Rada	OFDM radio	Audio
Need extra hardware	Yes	Yes	Yes	No
Test in dynamic environments? (*e.g.*, bus)	No	No	No	Yes
Need training?	Yes (SVM)	No	No	No
Sense gesture contexts? (*e.g.*, speed, range)	No	Yes (speed, range)	No	Yes (relative speed & range)
Accuracy	87.2%	N/A (hand track)	94%	95.1%
Gesture Resolution	14	N/A (hand track)	9	54 (randomly choose two attributes)



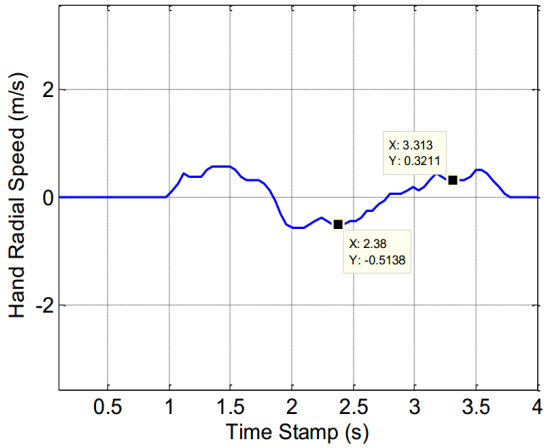
(a) Echo signal's spectrogram after FFT normalization



(b) Detected bandwidth of frequency shift at $t = 2.38s$ (*i.e.*, FFT Frame 51) by SoundWave

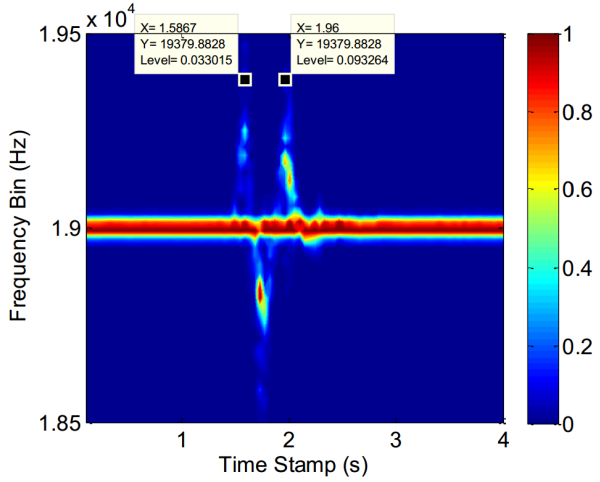


(c) Detected bandwidth at $t = 3.313s$ (*i.e.*, FFT Frame 71) by SoundWave

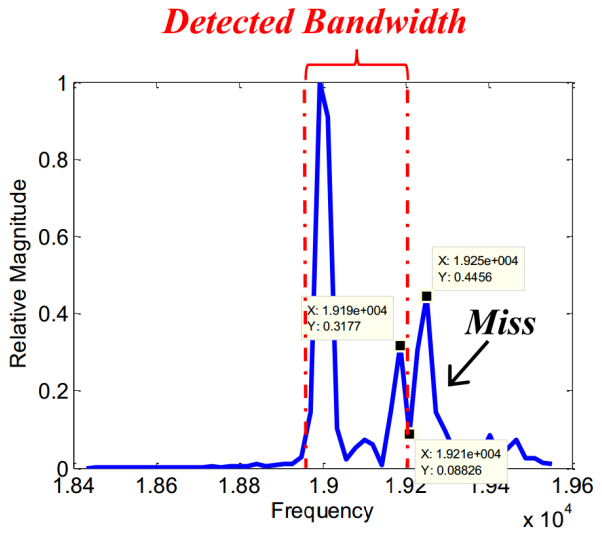


(d) The real-time hand radical velocity detected by *AudioGest*

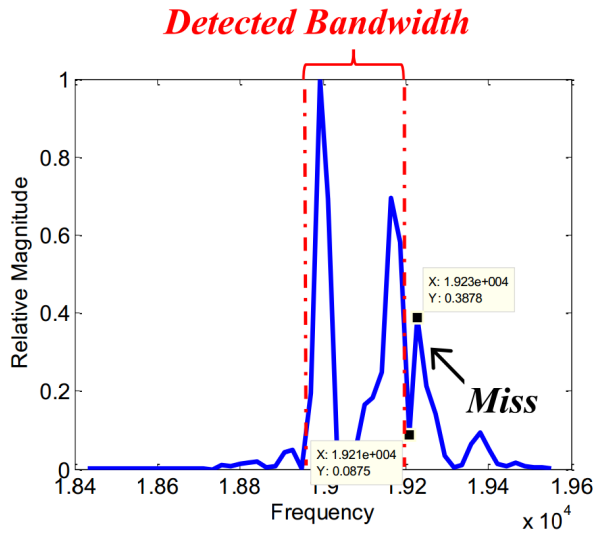
Fig. 8.32 Experimental Case 1: a slow-speed clockwise hand circling



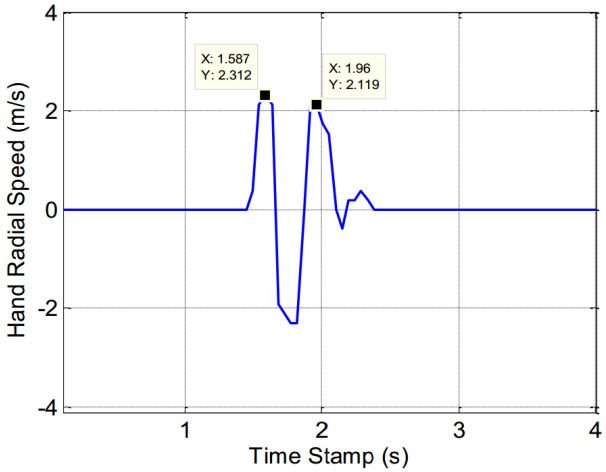
(a) Echo signal's spectrogram after FFT normalization



(b) Detected bandwidth of frequency shift at $t = 1.586s$ by SoundWave



(c) Detected bandwidth at frame $t = 1.96s$ by SoundWave



(d) The real-time hand radical velocity detected by *AudioGest*

Fig. 8.33 Experimental Case 2: a fast-speed clockwise hand circling

Case 1: Detecting Slow-speed Clockwise Hand Circling

Fig. 8.32 (b)∼(d) compare the detection results of SoundWave and AudioGest for a *Slow-Speed* clockwise circling case. In Fig. 8.32 (a), we observe that the hand is currently moving away from the microphone at $t = 2.38s$, and towards the microphone at $t = 3.3133s$. However, SoundWave cannot accurately detect frequency shifts in such two FFT frames (see Fig. 8.32 (b) and (c)) since both the second peaks are less than a threshold 30% and the lower point is below 10%, thus leading the recognition of "no motion". Fig. 8.32 (d) shows the result of our method, in which we first utilize *Squared Continuous Frame Subtraction* and *Gaussian Smoothing* to get the shifted frequency area and then transfer it into a hand radial speed curve. Both frequency shifts as well as hand speed in these two frames are successfully detected and estimated.

Experimental Case 2

Fig. 8.33 (b)∼(d) illustrate another detection results for a *Fast-Speed* clockwise circling case. Similarly, although SoundWave can successfully detect the happening of hand motion, it still fails to accurately estimate shifted bandwidth (missing the third peak), which results in incorrect hand speed estimation. Actually, those two FFT frames represent two peak speeds during the hand waving. Fig. 8.33 (d) shows our result, which correctly quantifies the bandwidth of the shifted frequency and captures the peak speeds.

To summarize, from the perspective of technique and methodology, percentage threshold-based dynamic peak tracking in SoundWave is a promising and efficient method that can deal with the hardware diversities and signal drifts. The FFT Normalization in our chapter actually serves the same purpose. However the rest techniques introduced by our system including *Squared Continuous Subtraction*, *Gaussian Smoothing* and *Hand Radial Speed Transformation* make AudioGest free of percentage threshold chosen and more accurate in quantifying shift frequency bandwidth.

8.7 Discussion

This section will discuss the limitations of our work that are left for the future work.

8.7.1 Separation of the Speaker and Microphone

In AudioGest, we focus on multi-modal hand gesture recognition with only one pair of microphone and speaker. Our system requires that the microphone and speaker are placed in different places. The rational of speaker-microphone separation lies on *i)* reducing the self-inference from the speaker (*i.e.,* preventing the microphone recording the sound of the speaker on the same device); *ii)* increasing the performance of microphone (*i.e.,* sound from outside is neutralized by the sound-wave emits from speaker with a higher possibility if speaker and microphone are close); and *iii)* limited deployment space in a mobile device. Interestingly, commercial mobile devices (e.g., laptops, mobile phones and tablets) on the market perfectly meet this requirement of AudioGest.

8.7.2 Gesture Trajectory

By making sensing of Doppler Effect, AudioGest can recognize six types of pre-defined basic gestures regardless of other hand motion attributes. The starting and stopping points of those gestures are quite flexible. AudioGest can interpret the spectrogram to the hand movements as long as the Doppler frequency shifts are captured. However, it is possible that two different gestures generate a same spectrogram, in which we cannot distinguish these two gestures. This is the reason that AudioGest needs to pre-define the hand moving trajectories. In other words, AudioGest does not care about the starting or ending point of hand movements, as long as two hand waving trajectories do not generate the same spectrogram.

8.7.3 Noise Disturbance to Human

Considering normal human hearing scope of 55∼18kHz, AudioGest emits a 19kHz single tone sound-wave. At the same time, to largely reduce the possible disturbance brought by the sound, we adjust the sound volume into a very low intensity since our system aims to detect hand movements around the vicinity of a mobile device. We adjust the volume of 19kHz into a very low-intensity level, which is difficult to be heard by people who is 0.5 meter away from the device. A 13-year-old young female participates in our experiments and she does not feel any uncomfortable while using our HGR system. That is also why we give up a plan using more advantaged soundwave signals (such as FMCW), which although has a better speed and range detection ability, it still generates small hear-able sound even the frequency of the signal is designed to be higher than 18kHz. In addition, we could choose a 20kHz sound-wave in our system, which would be safer and unobtrusive for human users.

8.8 Conclusion

To summarize, this chapter has shown how one single pair of microphone and speaker can achieve a multi-modal hand motion detection. AudioGest thoroughly exploits the Doppler frequency shift from hand movement and accurately interprets the spectrogram of echo signals into the multi-modal hand motion attributes. Our system only uses a single pair of COTS speaker & microphone without any extra hardware, and it is capable of accurately recovering hand's real-time radial velocity, thus decodes the hand moving direction, waving speed, and in-air range. By deeply interpreting the Doppler effect and hand motion, AudioGest is free from labor-intensive data labeling and supervised model training. It can provide more control commands for various applications by co-recognizing hand gestures and their associated motion attributes. The real-world experiments demonstrate the feasibility and effectiveness of

our system, which marks an important step toward enabling accurate and ubiquitous gesture recognition.

Chapter 9

Conclusion and Future Work

From six concrete research perspectives, this Ph.D. thesis has systematically explored how to utilize the cost-effective and maintenance-free passive RFID sensor technology to build a smart living-assistive system that can enable a healthy and safe independent life for the elderly in a residential home. From Chapter 3 to Chapter 8, for each research issue, we present a device-free, cost-effective and innovative approach that advances existing similar works. In this chapter, we will conclude this thesis and point out some promising future research issues, as well as give some illuminative solutions.

9.1 Conclusions

With the continuously growth of the aging population, the shortage of health-care service, and the importance that people want to remain independent and safe at their own homes, the demand on developing a novel living-assistive system can support the elderly live longer independently and safely in their own homes is becoming increasingly urgent. This Ph.D. thesis thus attempts to develop such a living-supportive system that can enable a healthy, safe, cost-effective independent living for the elderly in a residential home. Briefly, comparing to other related works, our proposed system built on passive RFID tags and cheap sensors

bears three major merits - *device-free*, *intelligent* and *maintenance-free*. In particular, we decompose the system into six specific research problems. For each issue, we provide a novel, device-free and cost-effective solution by taking recent advances of sensor technologies and state-of-the-art machine learning techniques.

First of all, this thesis has thoroughly reviews state-of-the-art related works from five research facets in Chapter 2, which substantially corresponds the six research issues we intend to solve. Concretely, in the hardware layer, we discuss the recent research efforts on missing data recovery, especially thoroughly compare the pros and cons between matrix completion and tensor completion techniques. In the discovery layer, we review the indoor human localization and activity recognition approaches from wearable device and device-free based techniques, especially intensively discuss the latter one by categorizing it into WIFI-based and RFID-based schemes. In the monitoring layer, we primarily focus on reviewing the fall detection systems. In the application layer, we discuss the recent hand gesture recognition systems and highlight the advantages of our HGR system.

Then, from Chapter 3 to Chapter 8, we elaborate the technical details of our solutions for those six research problems.

To deal with the challenge of sensor reading loss in passive RFID tags, Chapter 3 proposes tensor completion based method. It can accurately recover the missing readings given partial observed corrupted sensor data. The main novelty of this solution lies on that it can naturally capture the two-dimensional geographic dependency among sensors by formulating sensor readings with strong spatio-temporal correlations as a *multi-dimensional tensor*.

Upon this sensor reading recovery solution, Chapter 4 then presents a passive RFID-based device-free indoor localization and tracking system. Firstly, the localization problem is tackled by introducing a series of data-driven models to quantify the RSSI distributions when a user appears at various locations within a monitored area. These approaches enable our system to localize a subject by maximizing the posteriori probability given RSSI observations.

By transferring the pattern learned in localization, we further propose the multivariate GMM-based HMM and *k*NN-based HMM methods to deal with the human tracking problem.

However, the pure passive RFID based system cannot achieve satisfying localization and tracking accuracy in a cluttered living environment (*e.g.,* a a fully-furnished residential home) because RSSIs from passive tags are heavily obstructed by furniture or metallic appliances, significantly compromising the system's performance. As a result, in Chapter 5, we develop an enhanced RFID-based localization and tracking system that exploits human object interaction events to facilitate the traditional RFID-based localization under a rigid probabilistic framework. It is based on an intuition that, in a residential environment, the HOI events caused by daily living activities, detected by pervasive sensors, can potentially reveal people's transient locations, such as watching TV, opening the fridge door. The real-world experiments demonstrate the feasibility and effectiveness of our system, which marks an important step toward enabling an accurate and practical device-free human localization and tracking in a real residential home.

Apart from the resident's location context, recognizing the user's daily activity is also essential for our living-assistive system. Thus, Chapter 6 proposes a novel device-free human activity recognition approach by deploying the passive RFID tags as an tag-array. Such tag-array significantly improves the capability of sensing human activities comparing to regular passive RFID tags. Our HAR system can accurately classify 12 orientation-sensitive activities in a lab environment as well as in a residential environment. We also explore the optimal tag selection and placement that enable an more accurate activity recognition but with less passive tags.

Considering that falls are among the leading causes of hospitalization for the elderly, Chapter 7 introduces a device-free, fine-grained fall detection (FD) approach based on the same passive RFID hardware. Our FD system can simultaneously identify regular activities and detect a fall event by the proposed *p*-partially Angle-based Outlier Detection method.

More importantly, our system can distinguish different falling orientations, which we believe is a valuable context for the care-givers. Comparing to the current fall detection solutions, our approach relaxes the requirement of tuning parameters and provides more fine-grained contexts regarding fall events.

At last, to conveniently interact with those electronic appliances equipped in a residential environment (*e.g.,* automated window curtain, TV and air conditioner). Chapter 8 designs *AudioGest* , a novel hand gesture recognition system that uses one single pair of microphone and speaker to achieve a multi-modal, fine-grained hand motion detection. Our system is able to accurately recover hand's real-time radical velocity, and then decode the hand moving direction, waving speed and in-air range. It is also training-free and can provide up to 54 gesture control commands by randomly choosing two hand motion attributes.

Overall, this thesis provides a series of novel solutions for six research challenges. Those six parts are closely related and together compose a device-free, intelligent and maintenance-free living-assistive system that can enable an independent, low-cost and safe living for the elderly. Given the aging of the population, the cost of health care, and the importance that people want to remain independent and safe at their own homes, we believe the proposed innovative technologies in this thesis will be extremely valuable to both government and society in the era of global aging.

9.2 Open Issues for Future Work

In this section, we will identify some open research issues for the future work, which is listed as follows.

9.2.1 Sensor Data Recovery

In Chapter 3, we propose a robust tensor based model to recover missing readings for the sensor dataset with strong spatio-temporal correlations. Our method however still is based upon some assumptions that may be unpractical for some types of spatio-temporal sensor data.

One assumption is that, every mode of the tensor $\mathscr{X}$ is simultaneously low-rank. However, this assumption might be too strict to be satisfied in practice. For example, the original tensor is low-rank only in certain mode, such as our experiments in recovering RFID sensor data, where the low-rank may only exist in the third-mode for 4×4 sensor array. To deal with this issue, a straight-forward solution is to add a priori factor/weight to each unfolding matrix of the tensor thus penalize different modes with different low-rank priori knowledge (if we know in advance). Ideally, a method that can adaptively find the low-rank modes and only minimize the modes where low-rank exits may also be investigated in the future. Therefore, some other methods are proposed, such as the "Mixture" model for the tensor completion in [126, 125], which relaxed the each mode of tensor low-rank to the corresponding mode of the different component tensors low-rank. Based on this model, we can also explore the new representation method for the low-rank tensor in the future.

Moreover, in our model, we assume the noise is sparse, thus usually adopts l_1 norm that is optimal only for Laplacian noise. However, the real-world data may be Gaussian noise (Frobenius norm), or the mixture of different kinds of noise, even more complex noise, unknown noise. Therefore, it is necessary to consider a robust model to tackle much complex noise cases. The The relevant research efforts have emerged in matrix completion, such as a model to deal with mixture of Gaussian noise [198]. How to extend such model into a tensor case is also an interesting future work.

Lastly, in the optimization objective function, we used the Tucker rank (*i.e.,* multi-linear rank of tensor) in our model, which is not exactly equivalent to actual tensor rank. Similar to

matrix, estimation of the rank of tensor is a NP-hard problem [19], and the approximation of rank, *e.g.,* Tucker rank based on each unfolding matrix, might be omit some intrinsic information in some cases. The estimation of tensor rank can determined the accurate of the recovery, therefore, how to find the suitable rank of tensor worths an in-depth exploration in the future. In this chapter, we transform the tensor nuclear norm into the sum of nuclear norm in each unfolding matrix. However, we unavoidably need to compute many SVDs of matrices (possibly very large for some circumstances) at each iteration, leading to an expensive computational cost. As a result, some alternative tensor decomposition methods can be studied in the future.

9.2.2 Device-free Indoor Localization and Tracking

Chapter 4 introduces a series of methods from a data-driven viewpoint to deal with human localization and tracking problems. Comparing with physical models that leverage the backscatter propagation mechanism, it delivers many promising features including no requirement of tag pre-calibration, flexible deployment of RFID tags, a large monitoring area, and robustness in the face of multi-path effect[1]. However, a learning/training stage is necessary. Based on our experiments, for a $20m^2$ room, it requires about one-minute training data to reach 85% in accuracy. Future work in this regard will focus on the investigation of how to utilize the signal's backscatter propagation to facilitate our data-driven model for further reducing the learning overhead.

Our system in Chapter 4 targets to track a single resident in an indoor environment with an aim to support the elderly who live alone.. For the circumstance of several residents locating in a same residential room, the location-RSSI impacts from different persons will be tangled and coupled which require an expensive learning overhead, *i.e.,*, exponentially increasing with the number of residents. One way to address this problem is to retrieve

[1]Data-driven methods substantially learn the multi-path effect directly in the model-training stage instead of designing a delicate multipath propagation model in physical model based methods.

other information from the backscatter signals in RFID tags such as RF phase, RSSI reading rate, doppler frequency. Those signals can potentially serve as indicators of locations and reduce the pattern overlapping from multiple users, thus to ease the learning burden. The other possible solution is to build a delicate backscatter model to decompose the impact of different persons to RSSIs. In the future, we will investigate this idea in details.

9.2.3 Device-free Human Activity Recognition

In Chapter 6, we directly use raw RSSI as the features, since the feature extraction methods popularly used in inertial sensor based activity recognition do not work well on RSSI. We investigate 13 general feature extraction methods widely used in inertial sensor based activity recognition including mean average value, kurtosis, correlation. However, the performance are not good even very bad by using the features. We will keep exploring suitable features of RSSI in recognizing activities. Our next exploration will be deriving deeper relations between tags correlation in terms of temporal and spatial features, *e.g.,* strength variation and coverage, which can be exploited to build a more robust tag coverage model for accurate recognitions.

The HAR work presented in this thesis is the first step to recognize high-level human activities. There are three types of human activities generally: *i)* actions, which consist of multiple activities for a single person with temporal dimension, *e.g.,* walking, cooking; *ii)* interactions, which are activities that involve two or more persons, *e.g.,* two people are shaking hands; and *iii)* group activities, which are activities performed by groups of people, *e.g.,* a group of people having a meeting. Identifying these complex human activities is another main goal of our future work. Another promising research direction is how to develop a location-aware activity recognition system based on pure passive RFID tags, in which we not only can monitor the elderly people's activities but also locate their fine-grain locations, *e.g.,* we will know where the person is when our system detect his fall.

9.2.4 Device-free Fall Detection

In Chapter 7, we present a device-free and fine-grained fall detection system based on passive RFID tag-array. One of limitations is that the current system is designed for and tested with only a single resident. We believe that this is an important use case, particularly in an aging-in-place setting, which aims to ensure that a single person can live in his/her home and community safely and independently regardless of age and ability level. However, the number of profiles needed with multiple persons would increase exponentially. A more promising approach therefore would be to find techniques that can isolate concurrent activities in separate space from each other and match them against profiles separately, which we will consider in our future work. Another limitation is that labeling profiling data is time-consuming and labor-intensive, which is also an issue shared by other fall detection systems. In the Profile Construction phase, we have to use a camera to record the daily living activities, and then synchronize the camera and RSSI reading based on the time stamp, finally label and segment data streams into different action categories to build a labeled profile dataset based on the video records. So how to reduce the human labeling burden is a big challenge that worths us a further investigation.

Moreover, we use a standard, commercial RFID system with passive tags in our work. The passive tags are more cost-effective and, due to their simple structure and protective encapsulation, more robust than the sensor nodes. Passive tags operate without batteries. Once deployed, no further maintenance is required. The devices that require power in our sensing system is the RFID reader and antenna. But recent technical trends show that low-cost, low-power RFID readers are becoming commonly available by integrating into the smart phones, making our work potentially beneficial to more users in the future. Moreover, more advanced passive tags are emerging recently, such as WISP that can sense quantities such as light, temperature, acceleration, strain, liquid level. Such new RFID technology

will enable us to build more capable smart space, which is also a very promising research direction.

9.2.5 Device-free Hand Gesture Recognition

In Chapter 8, we develop a novel hand gesture recognition system *AudioGest* by transforming a mobile device into an active mini sonar system. According to our experiments, *AudioGest* can provide up to much more fine-grained control commands for applications by combining the hand-gesture types, hand in-air duration, average speed and waving range. It, however, can only distinguish overall eight hand gestures accurately. The main reason lies in that we only utilize one microphone and depend on the Doppler frequency shift to interpret the echo audio signals. In the future, we will investigate this from two ways: *i)* mining other features from the spectrogram of reflected signals to facilitate our physical model in order to recognize more hand gestures (it may bring some burden of labeling training data); and *ii)* adopting two or more microphones to enable a real-time hand motion tracking.

The work in Chapter 8 belongs to the area of *Hand Gesture Recognition* which focuses on recognizing pre-defined hand gestures. Another similar technique, *Hand Tracking*, however mainly aims to real-time recover hand moving trajectory. Generally, hand tracking needs more dedicated hardware support and efficient signal processing techniques. Given a pair of speaker and microphone, *AudioGest* 's upper limitation is to recognize several pre-defined hand movements and their associated motion features as illustrated in Chapter 8. However, with one more microphone (in a location that is different to the first one), *AudioGest* can substantially achieve a hand tracking purpose, which will be part of our future work.

As the system robustness evaluation shows, *AudioGest* 's performance decreases for some challenging scenarios such as the device orientation greatly changes ($> \pi/4$) and human motions at the vicinity of device ($< 0.5m$). However, such issues can be addressed by two possible ways: *i)* exploring the built-in 3-axis accelerometer to detect the orientation of the

device, then real-time updating parameters and hand-gesture recognition rules accordingly; *ii)* borrowing the idea from radar to transmit MFSK (multiple frequency shift keying) audio signal, enabling multiple-target range sensing, hence distinguishing the nearby environmental motion and hand movement.

References

[1] Jimeng Sun, Dacheng Tao, and Christos Faloutsos. Beyond streams and graphs: dynamic tensor analysis. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 374–383. ACM, 2006.

[2] Tamir Hazan, Simon Polak, and Amnon Shashua. Sparse image coding using a 3d non-negative tensor factorization. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 50–57. IEEE, 2005.

[3] Kim-Chuan Toh and Sangwoon Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(615-640):15, 2010.

[4] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[5] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Constantine Caramanis. Low-rank matrix recovery from errors and erasures. *Information Theory, IEEE Transactions on*, 59(7):4324–4337, 2013.

[6] Hui Ji, Sibin Huang, Zuowei Shen, and Yuhong Xu. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences*, 4(4):1122–1142, 2011.

[7] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.

[8] Yudong Chen, Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust matrix completion with corrupted columns. *arXiv preprint arXiv:1102.2254*, 2011.

[9] Olga Klopp, Karim Lounici, and Alexandre B Tsybakov. Robust matrix completion. *arXiv preprint arXiv:1412.8132*, 2014.

[10] Junbo Zhang Tianrui Li Xiuwen Yi, Yu Zheng. St-mvl: Filling missing values in geo-sensory time series data. In *IJCAI 2016*, 2016.

[11] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2114–2121. IEEE, 2009.

[12] Wenjie Ruan, Peipei Xu, Quan Z Sheng, Nickolas JG Falkner, Xue Li, and Wei Emma Zhang. Recovering missing values from corrupted spatio-temporal sensory data via robust low-rank tensor completion. In *International Conference on Database Systems for Advanced Applications*, pages 607–622. Springer, 2017.

[13] Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

[14] Nadia Kreimer, Aaron Stanton, and Mauricio D Sacchi. Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction. *Geophysics*, 78(6):V273–V284, 2013.

[15] Marco Signoretto, Lieven De Lathauwer, and Johan AK Suykens. Nuclear norms for tensors and their use for convex multilinear estimation. *Linear Algebra and Its Applications*, 43, 2010.

[16] Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

[17] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[18] Evrim Acar, Daniel M Dunlavy, Tamara G Kolda, and Morten Mørup. Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems*, 106(1):41–56, 2011.

[19] Boris Alexeev, Michael A Forbes, and Jacob Tsimerman. Tensor rank: Some lower and upper bounds. In *Computational Complexity (CCC), 2011 IEEE 26th Annual Conference on*, pages 283–291. IEEE, 2011.

[20] Curt Da Silva and Felix J Herrmann. Hierarchical tucker tensor optimization-applications to tensor completion. In *Proc. 10th International Conf. on Sampling Theory and Applications*, 2013.

[21] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location-support system. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom 2000)*, pages 32–43, 2000.

[22] L.M. Ni, Yunhao Liu, Yiu Cho Lau, and AP. Patil. LANDMARC: indoor location sensing using active rfid. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, pages 407–415, 2003.

[23] 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 2014)*, pages 237–248, 2014.

[24] Baoding Zhou, Qingquan Li, Qingzhou Mao, Wei Tu, and Xing Zhang. Activity sequence-based indoor pedestrian localization using smartphones. *IEEE Transactions on Human-Machine Systems*, 45(5):562–574, 2015.

[25] H. Xie, T. Gu, X. Tao, H. Ye, and J. Lu. A reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone. *IEEE Transactions on Mobile Computing*, PP(99):1–1, 2015.

[26] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. Challenges: Device-free passive localization for wireless environments. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking (MobiCom 2007)*, pages 222–229, 2007.

[27] B. Yang, Y. Lei, and B. Yan. Distributed multi-human location algorithm using naive bayes classifier for a binary pyroelectric infrared sensor tracking system. *IEEE Sensors Journal*, PP(99):1–1, 2015.

[28] Xufei Mao, Shaojie Tang, Jiliang Wang, and Xiang Yang Li. ilight: Device-free passive tracking using wireless sensor networks. *IEEE Sensors Journal*, 13(10):3785–3792, 2013.

[29] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1820–1833, 2011.

[30] Huan Dai, Zhao-Min Zhu, and Xiao-Feng Gu. Multi-target indoor localization and tracking on video monitoring system in a wireless sensor network. *Journal of Network and Computer Applications*, 36(1):228–234, 2013.

[31] T. Helten, M. Muller, H.-P. Seidel, and C. Theobalt. Real-time body tracking with one depth camera and inertial sensors. In *Proceedings of IEEE International Conference on Computer Vision (ICCV 2013)*, pages 1105–1112, 2013.

[32] Yunhao Liu, Yiyang Zhao, Lei Chen, Jian Pei, and Jinsong Han. Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays. *IEEE Transactions on Parallel and Distributed Systems*, 23(11):2138–2149, 2012.

[33] Daqiang Zhang, Jingyu Zhou, Minyi Guo, Jiannong Cao, and Tianbao Li. Tasa: Tag-free activity sensing using rfid tag arrays. *IEEE Transactions on Parallel and Distributed Systems*, 22(4):558–570, 2011.

[34] J. Wilson and N. Patwari. Radio tomographic imaging with wireless networks. *IEEE Transactions on Mobile Computing*, 9(5):621–632, 2010.

[35] Pablo Najera, Javier Lopez, and Rodrigo Roman. Real-time location and inpatient care systems based on passive rfid. *Journal of Network and Computer Applications*, 34(3):980–989, 2011.

[36] Moustafa Seifeldin, Ahmed Saeed, Ahmed E. Kosba, Amr El-keyi, and Moustafa Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing*, 12(7):1321–1334, 2013.

[37] Chenren Xu, Bernhard Firner, Robert S. Moore, Yanyong Zhang, Wade Trappe, Richard Howard, Feixiong Zhang, and Ning An. SCPL: Indoor Device-free Multi-subject Counting and Localization Using Radio Signal Strength. In *Proceedings of the*

12th International Conference on Information Processing in Sensor Networks (IPSN 2013), pages 79–90, 2013.

[38] A Saeed, AE. Kosba, and M. Youssef. Ichnaea: A low-overhead robust wlan device-free passive localization system. *IEEE Journal of selected Topics in Signal Processing*, 8(1):5–15, 2014.

[39] J. Han, C. Qian, X. Wang, D. Ma, J. Zhao, W. Xi, Z. Jiang, and Z. Wang. Twins: Device-free object tracking using passive tags. *IEEE/ACM Transactions on Networking*, PP(99):1–13, 2015.

[40] Wei Li, Jorge Portilla, Félix Moreno, Guixuan Liang, and Teresa Riesgo. Multiple feature points representation in target localization of wireless visual sensor networks. *Journal of Network and Computer Applications*, 57:119–128, 2015.

[41] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-person localization via rf body reflections. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation (NSDI 2015)*, pages 279–292, 2015.

[42] Du Yuanfeng, Yang Dongkai, Yang Huilin, and Xiu Chundi. Flexible indoor localization and tracking system based on mobile phone. *Journal of Network and Computer Applications*, 69:107–116, 2016.

[43] Zheng Yang, Zimu Zhou, and Yunhao Liu. From rssi to csi: Indoor localization via channel response. *ACM Computing Survey*, 46(2):25:1–25:32, 2013.

[44] Zheng Yang, Chenshu Wu, Zimu Zhou, Xinglin Zhang, Xu Wang, and Yunhao Liu. Mobility increases localizability: A survey on wireless indoor localization using inertial sensors. *ACM Computing Survey*, 47(3):54:1–54:34, 2015.

[45] Chenren Xu, Bernhard Firner, Yanyong Zhang, Richard Howard, Jun Li, and Xiaodong Lin. Improving rf-based device-free passive localization in cluttered indoor environments through probabilistic classification methods. In *Proceedings of the 11th International Conference on Information Processing in Sensor Networks (IPSN 2012)*, pages 209–220, 2012.

[46] B. Wagner, N. Patwari, and D. Timmermann. Passive rfid tomographic imaging for device-free user localization. In *Proceedings of the 9th Workshop on Positioning Navigation and Communication (WPNC 2012)*, pages 120–125, 2012.

[47] Lei Yang, Qiongzheng Lin, Xiang-Yang Li, Tianci Liu, and Yunhao Liu. See through walls with rfid systems! In *Proceedings of the 21th Annual International Conference on Mobile Computing and Networking (MobiCom 2015)*, 2015.

[48] Narayanan C Krishnan and Sethuraman Panchanathan. Analysis of low resolution accelerometer data for continuous human activity recognition. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3337–3340. IEEE, 2008.

[49] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *Proceedings of 2nd International Conference on Pervasive Computing (PERVASIVE)*, pages 1–17. Springer, 2004.

[50] Paul Lukowicz, Holger Junker, Mathias Stäger, Thomas von Bueren, and Gerhard Tröster. Wearnet: A distributed multi-sensor system for context aware wearables. In *Proceedings of ACM International Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 361–370. Springer, 2002.

[51] Wenjie Ruan, Quan Z Sheng, Lina Yao, Nguyen Khoi Tran, and Yu Chieh Yang. Preventdark: Automatically detecting and preventing problematic use of smartphones in darkness. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*, pages 1–3. IEEE, 2016.

[52] Nicky Kern, Bernt Schiele, Holger Junker, Paul Lukowicz, and Gerhard Tröster. Wearable sensing to annotate meeting recordings. *Personal and Ubiquitous Computing*, 7(5):263–274, 2003.

[53] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12(2):74–82, 2011.

[54] Nicholas D Lane et al. Bewell: A smartphone application to monitor, model and promote wellbeing. In *Proc. of 5th Intl. ICST Conference on Pervasive Computing Technologies for Healthcare*, pages 23–26, 2011.

[55] Narayanan C Krishnan and Diane J Cook. Activity recognition on streaming sensor data. *Pervasive and Mobile Computing*, 10:138–154, 2014.

[56] Maja Stikic et al. Adl recognition based on the combination of rfid and accelerometer sensing. In *Proc. of Intl. Conference Pervasive Computing Technologies for Healthcare*, 2008.

[57] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In *Proc. of 11th ACM Intl. Conference on Ubiquitous Computing (UbiComp)*, pages 51–60, 2009.

[58] Liang Wang, Tao Gu, Hongwei Xie, Xianping Tao, Jian Lu, and Yu Huang. A wearable rfid system for real-time activity recognition using radio patterns. In *Proc. of the 10th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*, 2013.

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

[60] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. Challenges: device-free passive localization for wireless environments. In *Proc. of 13th ACM Intl. Conference on Mobile Computing and Networking (MobiCom)*, 2007.

[61] Jihoon Hong and Tomoaki Ohtsuki. Ambient intelligence sensing using array sensor: device-free radio based approach. In *Proc. of ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, 2013.

[62] Stephan Sigg, Markus Scholz, Shuyu Shi, Yusheng Ji, and Michael Beigl. Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals. *IEEE Transactions on Mobile Computing (TMC)*, 13(4):907–920, 2014.

[63] Muhammad Mubashir, Ling Shao, and Luke Seed. A survey on fall detection: Principles and approaches. *Neurocomputing*, 100(0):144 – 152, 2013.

[64] N. Noury, A. Fleury, P. Rumeau, A.K. Bourke, G.O. Laighin, V. Rialle, and J.E. Lundy. Fall detection - principles and methods. In *Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE*, pages 1663–1666, 2007.

[65] J.K. Lee, S.N. Robinovitch, and E.J. Park. Inertial sensing-based pre-impact detection of falls involving near-fall scenarios. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 23(2):258–266, March 2015.

[66] Wen-Chang Cheng and Ding-Mao Jhan. Triaxial accelerometer-based fall detection method using a self-constructing cascade-adaboost-svm classifier. *Biomedical and Health Informatics, IEEE Journal of*, 17(2):411–419, March 2013.

[67] Qiang Li, J.A. Stankovic, M.A. Hanson, A.T. Barth, J. Lach, and Gang Zhou. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In *Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on*, pages 138–143, June 2009.

[68] Mars Lan, Ani Nahapetian, Alireza Vahdatpour, Lawrence Au, William Kaiser, and Majid Sarrafzadeh. Smartfall: An automatic fall detection system based on subsequence matching for the smartcane. In *Proceedings of the Fourth International Conference on Body Area Networks (BodyNets '09)*, pages 8:1–8:8, 2009.

[69] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 8(6):883 – 899, 2012.

[70] O. Aziz, C.M. Russell, E.J. Park, and S.N. Robinovitch. The effect of window size and lead time on pre-impact fall detection accuracy using support vector machine analysis of waist mounted inertial sensor data. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 30–33, Aug 2014.

[71] Yabo Cao, Yujiu Yang, and Wenhuang Liu. E-falld: A fall detection system using android-based smartphone. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on*, pages 1509–1513, May 2012.

[72] Lih-Jen Kau and Chih-Sheng Chen. A smart phone-based pocket fall accident detection, positioning, and rescue system. *Biomedical and Health Informatics, IEEE Journal of*, 19(1):44–56, Jan 2015.

[73] Zhengming Fu, E. Culurciello, P. Lichtsteiner, and T. Delbruck. Fall detection using an address-event temporal contrast vision sensor. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, May 2008.

[74] H. Foroughi, A. Naseri, A. Saberi, and H.S. Yazdi. An eigenspace-based approach for human fall detection using integrated time motion image and neural network. In *Signal Processing, 2008. ICSP 2008. 9th International Conference on*, pages 1499–1503, Oct 2008.

[75] E.E. Stone and M. Skubic. Fall detection in homes of older adults using the microsoft kinect. *Biomedical and Health Informatics, IEEE Journal of*, 19(1):290–301, Jan 2015.

[76] Xin Ma, Haibo Wang, Bingxia Xue, Mingang Zhou, Bing Ji, and Yibin Li. Depth-based human fall detection via shape features and improved extreme learning machine. *Biomedical and Health Informatics, IEEE Journal of*, 18(6):1915–1922, Nov 2014.

[77] Tracy Lee and Alex Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection. *Journal of telemedicine and telecare*, 11(4):194–198, 2005.

[78] Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau. Fall detection from human shape and motion history using video surveillance. In *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*, volume 2, pages 875–880. IEEE, 2007.

[79] Peter Hevesi, Sebastian Wille, Gerald Pirkl, Norbert Wehn, and Paul Lukowicz. Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 141–145, 2014.

[80] Yun Li, K.C. Ho, and M. Popescu. A microphone array system for automatic fall detection. *Biomedical Engineering, IEEE Transactions on*, 59(5):1291–1301, May 2012.

[81] H. Rimminen, J. Lindstrom, M. Linnavuo, and R. Sepponen. Detection of falls among the elderly by a floor sensor using the electric near field. *Information Technology in Biomedicine, IEEE Transactions on*, 14(6):1475–1476, Nov 2010.

[82] Chunmei Han, Kaishun Wu, Yuxi Wang, and L.M. Ni. Wifall: Device-free fall detection by wireless networks. In *INFOCOM, 2014 Proceedings*, April 2014.

[83] B.Y. Su, K.C. Ho, M.J. Rantz, and M. Skubic. Doppler radar fall activity detection using the wavelet transform. *Biomedical Engineering, IEEE Transactions on*, 62(3):865–875, March 2015.

[84] L.M. Ni, Dian Zhang, and M.R. Souryal. Rfid-based localization and tracking technologies. *Wireless Communications, IEEE*, 18(2):45–51, April 2011.

[85] Wenjie Ruan, Lina Yao, Quan Z. Sheng, Xue Li, and Nicholas J.G. Falkner. Tagtrack: Device-free localization and tracking using passive rfid tags. In *Proc. of the 11th Intl. Conf. on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MOBIQUITOUS '14, 2014.

[86] Jiahui Wu, Gang Pan, Daqing Zhang, Guande Qi, and Shijian Li. Gesture recognition with a 3-d accelerometer. In *Ubiquitous intelligence and computing*, pages 25–38. Springer, 2009.

[87] Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. Humantenna: using the body as an antenna for real-time whole-body interaction. In *The ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*, pages 1901–1910, 2012.

[88] Sandip Agrawal, Ionut Constandache, Shravan Gaonkar, Romit Roy Choudhury, Kevin Caves, and Frank DeRuyter. Using mobile phones to write in air. In *The International Conference on Mobile Systems, Applications, and Services (MobiSys'11)*, pages 15–28, 2011.

[89] Taiwoo Park, Jinwon Lee, Inseok Hwang, Chungkuk Yoo, Lama Nachman, and Junehwa Song. E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices. In *The ACM Conference on Embedded Networked Sensor Systems (SenSys'11)*, pages 260–273, 2011.

[90] Hamed Ketabdar, Peyman Moghadam, Babak Naderi, and Mehran Roshandel. Magnetic signatures in air for mobile devices. In *The ACM 14th international conference on Human-computer interaction with mobile devices and services companion*, pages 185–188, 2012.

[91] Zhiyuan Lu, Xiang Chen, Qiang Li, Xu Zhang, and Ping Zhou. A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices. *IEEE transactions on human-machine systems*, 44(2):293–299, 2014.

[92] Gurashish Singh, Alexander Nelson, Ryan Robucci, Chintan Patel, and Nilanjan Banerjee. Inviz: Low-power personalized gesture recognition using wearable textile capacitive sensor arrays. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 198–206. IEEE, 2015.

[93] Zheng Li, Ryan Robucci, Nilanjan Banerjee, and Chintan Patel. Tongue-n-cheek: non-contact tongue gesture recognition. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 95–105. ACM, 2015.

[94] Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N Patel. Tongue-in-cheek: Using wireless signals to enable non-intrusive and flexible facial gestures detection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 255–258. ACM, 2015.

[95] Jacob O Wobbrock, Andrew D Wilson, and Yang Li. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 159–168. ACM, 2007.

[96] Yang Li. Protractor: a fast and accurate gesture recognizer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2169–2172. ACM, 2010.

[97] Thad Starner and Alex Pentland. Real-time american sign language recognition from video using hidden markov models. In *Motion-Based Recognition*, pages 227–243. Springer, 1997.

[98] Sy Bor Wang, Ariadna Quattoni, Louis-Philippe Morency, David Demirdjian, and Trevor Darrell. Hidden conditional random fields for gesture recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1521–1527, 2006.

[99] Nasser H Dardas and Nicolas D Georganas. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Transactions on Instrumentation and Measurement*, 60(11):3592–3607, 2011.

[100] Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Kari Pulli. Multi-sensor system for driver's hand-gesture recognition. In *IEEE Conference and Workshops on Automatic Face and Gesture Recognition (FG'15)*, volume 1, pages 1–8, 2015.

[101] Xin Zhao, Xue Li, Chaoyi Pang, Xiaofeng Zhu, and Quan Z Sheng. Online human gesture recognition from motion data streams. In *The ACM international conference on Multimedia (MM'13)*, pages 23–32, 2013.

[102] Han Ding, Longfei Shangguan, Zheng Yang, Jinsong Han, et al. Femo: A platform for free-weight exercise monitoring with rfids. In *The ACM Conference on Embedded Networked Sensor Systems (SenSys'15)*, pages 141–154, 2015.

[103] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *IEEE Conference on Computer Communications (INFOCOM'15)*, pages 1472–1480, 2015.

[104] Fadel Adib and Dina Katabi. See through walls with wifi! In *The ACM SIGCOMM 2013 Conference (SIGCOMM'13)*, pages 75–86, 2013.

[105] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):219, 2015.

[106] Lina Yao, Quan Z. Sheng, Wenjie Ruan, Tao Gu, Xue Li, Nick Falkner, and Zhi Yang. Rf-care: Device-free posture recognition for elderly people using a passive rfid tag array. In *The international Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous'15)*, pages 120–129, 2015.

[107] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *The international conference on Mobile computing & networking (MobiCom'13)*, pages 27–38, 2013.

[108] Pedro Melgarejo, Xinyu Zhang, Parameswaran Ramanathan, and David Chu. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *The ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*, pages 541–551, 2014.

[109] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. Bringing gesture recognition to all devices. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI'14)*, pages 303–316, 2014.

[110] Kaustubh Kalgaonkar and Bhiksha Raj. One-handed gesture recognition using ultrasonic doppler sonar. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)*, pages 1889–1892, 2009.

[111] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave: using the doppler effect to sense gestures. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*, pages 1911–1914, 2012.

[112] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *The ACM international Conference on Mobile Systems, Applications, and Services (MobiSys'15)*, pages 45–57, 2015.

[113] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *The ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'16)*, pages 1515–1525, 2016.

[114] Stephen P Tarzia, Robert P Dick, Peter A Dinda, and Gokhan Memik. Sonar-based measurement of user presence and attention. In *The ACM international conference on Ubiquitous computing (UbiComp'09)*, pages 89–92, 2009.

[115] Li Da Xu, Wu He, and Shancang Li. Internet of things in industries: a survey. *Industrial Informatics, IEEE Transactions on*, 10(4):2233–2243, 2014.

[116] Yasushi Sakurai, Yasuko Matsubara, and Christos Faloutsos. Mining and forecasting of big time-series data. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 919–922. ACM, 2015.

[117] Hsun-Ping Hsieh, Shou-De Lin, and Yu Zheng. Inferring air quality for station location recommendation based on urban big data. In *Proc. of the 21th ACM SIGKDD Intl. Conference on Knowledge Discovery and Data Mining*, pages 437–446, 2015.

[118] Peipei Xu, Wenjie Ruan, Quan Z. Sheng, Tao Gu, and Lina Yao. Interpolating the missing values for multi-dimensional spatio-temporal sensory data: A tensor svd approach. In *Proc. of the 14th Intl. Conf. on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MOBIQUITOUS '17, 2017.

[119] Wenjie Ruan, Peipei Xu, Quan Z Sheng, Nguyen Khoi Tran, Nickolas JG Falkner, Xue Li, and Wei Emma Zhang. When sensor meets tensor: Filling missing sensor values through a tensor approach. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2025–2028. ACM, 2016.

[120] AJ Lawrance and PAW Lewis. An exponential moving-average sequence and point process. *Journal of Applied Probability*, pages 98–113, 1977.

[121] Mohinder S Grewal. *Kalman filtering*. Springer, 2011.

[122] Anthony M Norcia, Maureen Clarke, and Christopher W Tyler. Digital filtering and robust regression techniques for estimating sensory thresholds from the evoked potential. *IEEE Engineering in Medicine and Biology Magazine*, 4(4):26–32, 1985.

[123] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[124] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[125] Ryota Tomioka, Kohei Hayashi, and Hisashi Kashima. Estimation of low-rank tensors via convex optimization. *arXiv preprint arXiv:1010.0789*, 2010.

[126] Donald Goldfarb and Zhiwei Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.

[127] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[128] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[129] Wenjie Ruan, Lina Yao, Quan Z Sheng, Nickolas Falkner, Xue Li, and Tao Gu. Tagfall: Towards unobstructive fine-grained fall detection based on uhf passive rfid tags. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 140–149, 2015.

[130] Lina Yao, Wenjie Ruan, Quan Z Sheng, Xue Li, and Nicholas JG Falkner. Exploring tag-free rfid-based passive localization and tracking via learning-based probabilistic approaches. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1799–1802. ACM, 2014.

[131] Wenjie Ruan, Quan Z Sheng, Lina Yao, Xue Li, Nicholas J.G. Falkner, and Lei Yang. Device-free human localization and tracking with uhf passive rfid tags: A data-driven approach. *Journal of Network and Computer Applications*, 2017.

[132] Wenjie Ruan, Quan Z Sheng, Lina Yao, Tao Gu, Michele Ruta, and Longfei Shangguan. Device-free indoor localization and tracking through human-object interactions. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2016 IEEE 17th International Symposium on A*, pages 1–9. IEEE, 2016.

[133] Chengwen Luo, Hande Hong, Long Cheng, Mun Choon Chan, Jianqiang Li, and Zhong Ming. Accuracy-aware wireless indoor localization: Feasibility and applications. *Journal of Network and Computer Applications*, 62:128–136, 2016.

[134] Chenshu Wu, Zheng Yang, Yunhao Liu, and Wei Xi. WILL: Wireless indoor localization without site survey. In *Proceedings of the 31st IEEE International Conference on Computer Communications (INFOCOM 2012)*, pages 64–72, 2012.

[135] T. Liu, Y. Liu, L. Yang, Y. Guo, and W. Cheng. Backpos: High accuracy backscatter positioning system. *IEEE Transactions on Mobile Computing*, PP(99):1–1, 2015.

[136] Kaishun Wu, Jiang Xiao, Youwen Yi, Dihu Chen, Xiaonan Luo, and L.M. Ni. Csi-based indoor localization. *IEEE Transactions on Parallel and Distributed Systems*, 24(7):1300–1309, 2013.

[137] 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.

[138] Wenjie Ruan, Lina Yao, Quan Z. Sheng, Nickolas Falkner, and Xue Li. Device-free localization and tracking using passive rfid tags. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2014)*, pages 80–89, 2014.

[139] Francesco Rizzo, Marcello Barboni, Lorenzo Faggion, Graziano Azzalin, and Marco Sironi. Improved security for commercial container transports using an innovative active rfid system. *Journal of Network and Computer Applications*, 34(3):846–852, 2011.

[140] Leonardo A Amaral, Fabiano P Hessel, Eduardo A Bezerra, Jerônimo C Corrêa, Oliver B Longhi, and Thiago FO Dias. ecloudrfid–a mobile software framework architecture for pervasive rfid-based applications. *Journal of Network and Computer Applications*, 34(3):972–979, 2011.

[141] Daniel M Dobkin. *The RF in RFID: UHF RFID in Practice, 2nd Edition*. Newnes, 2012.

[142] Lei Yang, Yi Guo, Tianci Liu, Cheng Wang, and Yunhao Liu. Perceiving the slightest tag motion beyond localization. *IEEE Transactions on Mobile Computing*, 14(11):2363–2375, 2015.

[143] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[144] Lina Yao, Wenjie Ruan, Quan Z Sheng, Xue Li, et al. Exploring tag-free rfid-based passive localization and tracking via learning-based probabilistic approaches. In *Proceedings of 23rd ACM International Conference on Information and Knowledge Management (CIKM 2014)*, pages 1799–1802, 2014.

[145] Wenjie Ruan, Quan Z Sheng, Lina Yao, Lei Yang, and Tao Gu. Hoi-loc: Towards unobstructive human localization with probabilistic multi-sensor fusion. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*, pages 1–4. IEEE, 2016.

[146] Jizhong Xiao, S.L. Joseph, Xiaochen Zhang, Bing Li, Xiaohai Li, and Jianwei Zhang. An assistive navigation framework for the visually impaired. *IEEE Transactions on Human-Machine Systems*, 45(5):635–640, 2015.

[147] Shuai Tao, M. Kudo, Bing-Nan Pei, H. Nonaka, and J. Toyama. Multiperson locating and their soft tracking in a binary infrared sensor network. *IEEE Transactions on Human-Machine Systems*, 45(5):550–561, 2015.

[148] Jinsong Han, Chen Qian, Xing Wang, Dan Ma, Jizhong Zhao, Pengfeng Zhang, Wei Xi, and Zhiping Jiang. Twins: Device-free object tracking using passive tags. In *Proceedings of the 33rd IEEE International Conference on Computer Communications (INFOCOM 2014)*, pages 469–476, April 2014.

[149] L.M. Ni, Dian Zhang, and M.R. Souryal. Rfid-based localization and tracking technologies. *IEEE Wireless Communications*, 18(2):45–51, April 2011.

[150] 2014 state of the smart home, http://www.icontrol.com/blog/2014-state-smart-home/.

[151] Liyanage C. De Silva, Chamin Morikawa, and Iskandar M. Petra. State of the art of smart homes. *Engineering Applications of Artificial Intelligence*, 25(7):1313 – 1321, 2012.

[152] Lina Yao, Quan Z Sheng, Anne H.H. Ngu, and Byron Gao. Keeping you in the loop: Enabling web-based things management in the internet of things. In *Proceedings of 23rd ACM International Conference on Information and Knowledge Management (CIKM 2014)*, November 2014.

[153] B.N. Schilit and M.M. Theimer. Disseminating active map information to mobile hosts. *IEEE Network*, 8:22–32, 1994.

[154] Ping Wei, Yibiao Zhao, Nanning Zheng, and Song-Chun Zhu. Modeling 4d human-object interactions for event and object recognition. In *Computer Vision (ICCV '13), 2013 IEEE International Conference on*, pages 3272–3279.

[155] Vincent Delaitre, Josef Sivic, and Ivan Laptev. Learning person-object interactions for action recognition in still images. In *Advances in neural information processing systems (NIPS '11)*, pages 1503–1511, 2011.

[156] Bangpeng Yao and Li Fei-Fei. Recognizing human-object interactions in still images by modeling the mutual context of objects and human poses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(9):1691–1703, 2012.

[157] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, (1–3):489 – 501, 2006.

[158] Tian He et al. Vigilnet: An integrated sensor network system for energy-efficient surveillance. *ACM Transactions on Sensor Networks (TOSN)*, 2(1):1–38, 2006.

[159] Lina Yao, Quan Z Sheng, Xue Li, Tao Gu, Mingkui Tan, Xianzhi Wang, Sen Wang, and Wenjie Ruan. Compressive representation for device-free activity recognition with passive rfid signal strength. *IEEE Transactions on Mobile Computing*, 2017.

[160] Matthai Philipose et al. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 3(4):50–57, 2004.

[161] DJ Cook and M Schmitter-Edgecombe. Assessing the quality of activities in a smart environment. *Methods of Information in Medicine*, 48(5):480, 2009.

[162] Wenjie Ruan. Unobtrusive human localization and activity recognition for supporting independent living of the elderly. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*, pages 1–3. IEEE, 2016.

[163] Nicky Kern, Bernt Schiele, and Albrecht Schmidt. Multi-sensor activity context detection for wearable computing. In *Proc. of the 1st European Symposium on Ambient Intelligence (EUSAI)*, pages 220–232. 2003.

[164] Stephen Intille et al. Using a live-in laboratory for ubiquitous computing research. In *Proc. of Intl. Conf. on Pervasive Computing (PERVASIVE)*. 2006.

[165] Jamie A Ward et al. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(10):1553–1567, 2006.

[166] David Minnen, Thad Starner, Irfan Essa, and Charles Isbell. Discovering characteristic actions from on-body sensor data. In *Proc. of 10th IEEE International Symposium on Wearable Computers (ISWC)*, pages 11–18, 2006.

[167] Wenjie Ruan, Leon Chea, Quan Z Sheng, and Lina Yao. Recognizing daily living activity using embedded sensors in smartphones: A data-driven approach. In *Advanced Data Mining and Applications: 12th International Conference, ADMA 2016, Gold Coast, QLD, Australia, December 12-15, 2016, Proceedings 12*, pages 250–265. Springer, 2016.

[168] Moustafa Seifeldin, Ahmed Saeed, Ahmed E Kosba, Amr El-Keyi, and Moustafa Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing (TMC)*, 12(7):1321–1334, 2013.

[169] Dian Zhang et al. Rass: A real-time, accurate and scalable system for tracking transceiver-free objects. In *Proc. of IEEE Intl. Conference on Pervasive Computing and Communications (PerCom)*, 2011.

[170] Jinsong Han, Chen Qian, Dan Ma, Xing Wang, Jizhong Zhao, Pengfeng Zhang, Wei Xi, and Zhiping Jiang. Twins: device-free object tracking using passive tags. In *Proc. of IEEE Intl. Conference on Computer Communications (INFOCOM)*, 2014.

[171] Stephan Wagner et al. On optimal tag placement for indoor localization. In *Proc. of IEEE Intl. Conference on Pervasive Computing and Communications (PerCom)*, pages 162–170, 2012.

[172] Lina Yao, Quan Z Sheng, Xue Li, Sen Wang, Tao Gu, Wenjie Ruan, and Wan Zou. Freedom: Online activity recognition via dictionary-based sparse representation of rfid sensing data. In *Data Mining (ICDM), 2015 IEEE International Conference on*, pages 1087–1092. IEEE, 2015.

[173] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[174] Lina Yao, Quan Z Sheng, Wenjie Ruan, Xue Li, Sen Wang, and Zhi Yang. Unobtrusive posture recognition via online learning of multi-dimensional rfid received signal strength. In *Parallel and Distributed Systems (ICPADS), 2015 IEEE 21st International Conference on*, pages 116–123. IEEE, 2015.

[175] David M Blei, Michael I Jordan, et al. Variational inference for dirichlet process mixtures. *Bayesian analysis*, 1(1):121–143, 2006.

[176] Moustafa Seifeldin and Moustafa Youssef. A deterministic large-scale device-free passive localization system for wireless environments. In *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2010)*, page 51, 2010.

[177] Rein Tideiksaar. *Falls in older people: Prevention and management*. Health Professions Press, 2002.

[178] A. Sixsmith and N. Johnson. A smart sensor to detect the falls of the elderly. *Pervasive Computing, IEEE*, 3(2):42–47, April 2004.

[179] F. Bianchi, S.J. Redmond, M.R. Narayanan, S. Cerutti, and N.H. Lovell. Barometric pressure and triaxial accelerometry-based falls event detection. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(6):619–627, Dec 2010.

[180] Hans-Peter Kriegel, Matthias S hubert, and Arthur Zimek. Angle-based outlier detection in high-dimensional data. In *Proc. of the 14th ACM SIGKDD*, 2008.

[181] Nicholas Gillian, R Benjamin Knapp, and Sile O'Modhrain. Recognition of multivariate temporal musical gestures using n-dimensional dynamic time warping. In *Proc of the 11th Int'l Conf. on New Interfaces for Musical Expression*, 2011.

[182] Nathalie Mitton and David Simplot-Ryl. Is rfid dangerous?, 2011.

[183] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proc. of the ACM SIGMOD*, 2000.

[184] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011.

[185] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, 2009.

[186] Yanzhi Ren, Yingying Chen, Mooi Choo Chuah, and Jie Yang. User verification leveraging gait recognition for smartphone enabled mobile healthcare systems. *IEEE Transactions on Mobile Computing*, 14(9):1961–1974, 2015.

[187] Wenjie Ruan, Quan Z Sheng, Peipei Xu, Lei Yang, Tao Gu, and Longfei Shangguan. Making sense of doppler effect for multi-modal hand motion detection. *IEEE Transactions on Mobile Computing*, 2017.

[188] Lei Yang, Yi Guo, Xuan Ding, Jinsong Han, Yunhao Liu, Cheng Wang, and Changwei Hu. Unlocking smart phone through handwaving biometrics. *IEEE Transactions on Mobile Computing*, 14(5):1044–1055, 2015.

[189] Xin Zhao, Xue Li, Chaoyi Pang, Quan Z Sheng, Sen Wang, and Mao Ye. Structured streaming skeleton–a new feature for online human gesture recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(1s):22, 2014.

[190] Siddharth S Rautaray and Anupam Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54, 2015.

[191] Wenjie Ruan, Quan Z Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. Audiogest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 474–485. ACM, 2016.

[192] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. Vision-based hand-gesture applications. *Communication of ACM*, 54(2):60–71, 2011.

[193] David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *The ACM symposium on User interface software and technology (UIST'12)*, pages 167–176, 2012.

[194] Chong Wang, Zhong Liu, and Shing-Chow Chan. Superpixel-based hand gesture recognition with kinect depth camera. *IEEE Transactions on Multimedia*, 17(1):29–39, 2015.

[195] G Deng and LW Cahill. An adaptive gaussian filter for noise reduction and edge detection. In *IEEE 1993 Nuclear Science Symposium and Medical Imaging Conference*, pages 1615–1619, 1993.

[196] Chen Zhao, Ke-Yu Chen, Md Tanvir Islam Aumi, Shwetak Patel, and Matthew S Reynolds. Sideswipe: detecting in-air gestures around mobile devices using actual gsm signal. In *The ACM symposium on User interface software and technology (UIST'14)*, pages 527–534, 2014.

[197] Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Kari Pulli. Short-range fmcw monopulse radar for hand-gesture sensing. In *IEEE Radar Conference (Radar'15),*, pages 1491–1496, 2015.

[198] Qian Zhao, Deyu Meng, Zongben Xu, Wangmeng Zuo, and Lei Zhang. Robust principal component analysis with complex noise. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 55–63, 2014.

[199] Shiqian Ma, Lingzhou Xue, and Hui Zou. Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation*, 25(8):2172–2198, 2013.

APPENDIX A

Convergence Proof

Proof 1 *: To proof our convergence, we first define following notations:* $\bar{\mathscr{M}} = TArray(\mathscr{M}_1, \mathscr{M}_2, \mathscr{M}_3), \bar{\mathscr{N}} = TArray(\mathscr{N}, \mathscr{N}, \mathscr{N})$ *and* $\bar{\mathscr{O}} = TArray(\mathscr{O}, \mathscr{O}, \mathscr{O})$*, as well as* $F(\bar{\mathscr{M}}) = f(\mathscr{M}_1, \mathscr{M}_2, \mathscr{M}_3) := \sum_{i=1}^{3} \|M_{i,(i)}\|_*$ *and* $G(\bar{\mathscr{N}}) = 3g(\mathscr{N}) := 3\lambda_1 \|\mathscr{N}\|_1$ *where* $F(\cdot)$ *and* $G(\cdot)$ *are convex functions. As a result, we can write Equation 3.4 as:*

$$\begin{aligned} &\min_{\bar{\mathscr{M}}, \bar{\mathscr{N}}} F(\bar{\mathscr{M}}) + G(\bar{\mathscr{N}}) \\ &\text{s.t. } \bar{\mathscr{M}} + \bar{\mathscr{N}} = \bar{\mathscr{O}} \end{aligned} \tag{A.1}$$

And the $(k+1)$*-th iteration from Algorithm 1 is as follows:*

$$\begin{aligned} \bar{\mathscr{M}}^{k+1} &= \arg\min_{\bar{\mathscr{M}}} F(\bar{\mathscr{M}}) + \frac{1}{2\mu} \|\bar{\mathscr{M}} + \bar{\mathscr{N}}^k - \bar{\mathscr{O}} - \mu \bar{\mathscr{Y}}^k\|^2; \\ \bar{\mathscr{N}}^{k+1} &= \arg\min_{\bar{\mathscr{N}}} G(\bar{\mathscr{N}}) + \frac{1}{2\mu} \|\bar{\mathscr{N}} + \bar{\mathscr{M}}^{k+1} - \bar{\mathscr{O}} - \mu \bar{\mathscr{Y}}^k\|^2; \\ \bar{\mathscr{Y}}^{k+1} &= \bar{\mathscr{Y}}^k - \frac{1}{\mu}(\bar{\mathscr{M}}^{(k+1)} + \bar{\mathscr{N}}^{(k+1)} - \bar{\mathscr{O}}) \end{aligned} \tag{A.2}$$

To assist our proof, we introduce Lemma 1 below.

Lemma 1 *Assuming that* $\bar{\mathscr{M}}, \bar{\mathscr{N}}$ *are an optimal solution for Equation A.1, and* $\bar{\mathscr{Y}}$ *represents corresponding optimal dual variable according to the equality constraint* $\bar{\mathscr{M}} + \bar{\mathscr{N}} = \bar{\mathscr{O}}$*. Ob-*

viously, there exists $\eta > 0$ *making the sequence* $(\bar{\mathscr{M}}^k, \bar{\mathscr{N}}^k, \bar{\mathscr{Y}}^k)$ *obtained from Equation A.2 meets following relation.*

$$\|\mathscr{W}^k - \mathscr{W}^*\|_{\mathscr{D}}^2 - \|\mathscr{W}^{k+1} - \mathscr{W}^*\|_{\mathscr{D}}^2 \geqslant \eta \|\mathscr{W}^{k+1} - \mathscr{W}^k\|_{\mathscr{D}}^2 \tag{A.3}$$

Where $\mathscr{W} = TArray(\bar{\mathscr{M}}, \bar{\mathscr{Y}})$, $\mathscr{W}^k = TArray(\bar{\mathscr{M}}^k, \bar{\mathscr{Y}}^k)$, $\|\mathscr{W}\|_{\mathscr{D}} := \langle \mathscr{W}, \mathscr{D}\mathscr{W} \rangle$. *The inner product* $\langle \mathscr{W}, \mathscr{V} \rangle_{\mathscr{D}} := \langle \mathscr{W}, \mathscr{D}\mathscr{V} \rangle$, *where* $\mathscr{D} = \begin{pmatrix} \mu\mathscr{I} & 0 \\ 0 & \mu\mathscr{I} \end{pmatrix}$ *and* $\mathscr{V}$ *is anther tensor array with the same size as* $\mathscr{W}$. *The detail proof can be found in [199].*

Based on Lemma 1, we obtain following three properties: i) $\|\mathscr{W}^k - \mathscr{W}^{k+1}\|_{\mathscr{D}} \to 0$*;* ii) $\{\mathscr{W}^k\}$ *lies in a compact region, and* iii) $\|\mathscr{W}^k - \mathscr{W}^*\|_{\mathscr{D}}^2$ *is non-increasing monotonically so that it converges. As a result, we can obtain that the sequence*$\{\mathscr{W}^k, \mathscr{N}^k\}$ *has a subsequence that can converge to* $\{\hat{\mathscr{W}}^k, \hat{\mathscr{N}}^k\}$. *Based on the optimality conditions in the two subproblems of Equation A.2, any limit point* $\{\hat{\mathscr{W}}^k, \hat{\mathscr{N}}^k\}$ *in the sequence* $\{\mathscr{W}^k, \mathscr{N}^k\}$ *meets the KKT (Karush–Kuhn–Tucker) conditions for Equation A.1. As a result, any limit point of* $\{\bar{\mathscr{M}}, \bar{\mathscr{N}}\}$ *is an optimal solution for Equation 3.4. Similar to the proof of the robust matrix completion [123], the above proof for robust tensor completion is also valid for the partial observation case.*

APPENDIX B

Examples of Denoising and Segmentation in AudioGest

In this Appendix, we depict the spectrograms after denoising and our segmentation results for various hand gestures waving with different speeds. As we can see, the proposed segmentation method can accurately localize those areas where Doppler frequency shifts happen.

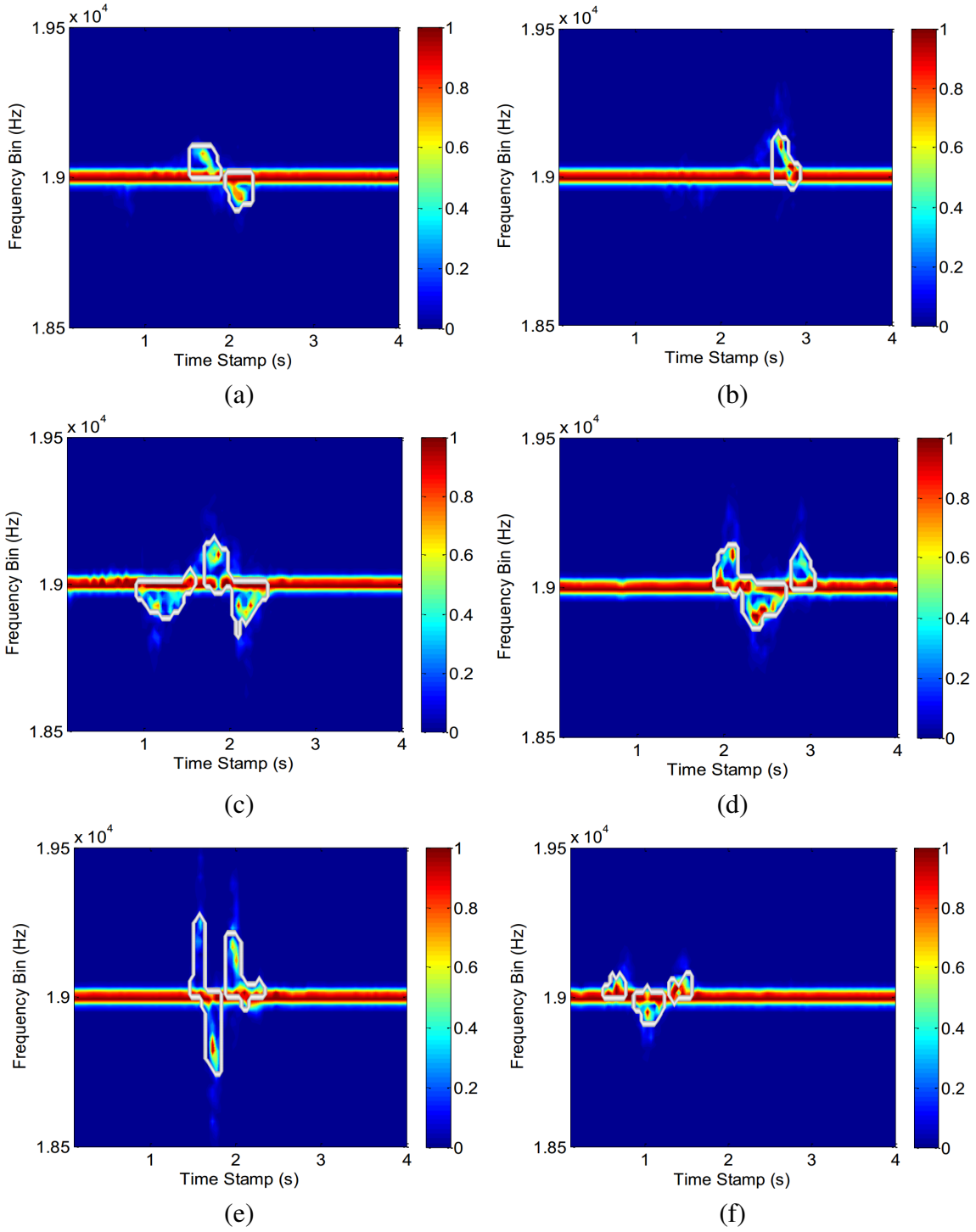


Fig. B.1 Denoised spectrograms of different hand gestures with various speeds and their segmentation results: waving hand (a) from Right to Left; (b) from Up to Down; (c) Anti-clockwise Circle; (d) Clockwise Circle; (e) Clockwise Circle with fast speed; (f) Clockwise Circle with slow speed

APPENDIX C

Multi-modal Hand Detection Examples

This Appendix illustrates some real examples of how our multi-modal hand detection works. Fig. C.1 and Fig. C.2 show the FFT-normalized spectrograms and their corresponding real-time hand radial-velocity curves detected and the in-air waving duration, speed-ratios and range-ratios measured by our system. Those four hand gestures are differentiated by their motion trajectories in AudioGest like other HGR systems.

Fig. C.3 and Fig. C.4 depict four types of clockwise hand circling with different waving speeds and ranges. Different to current HGR systems that can not distinguish those gestures, AudioGest can recognize hand gestures (a) and (b) in Fig. C.3 by the speed-ratio, and hand gesture (a) and (b) in Fig. C.4 by their different range-ratios even though they share the same waving trajectory. Those examples elaborate how AudioGest achieves the *multi-modal* hand motion detection.

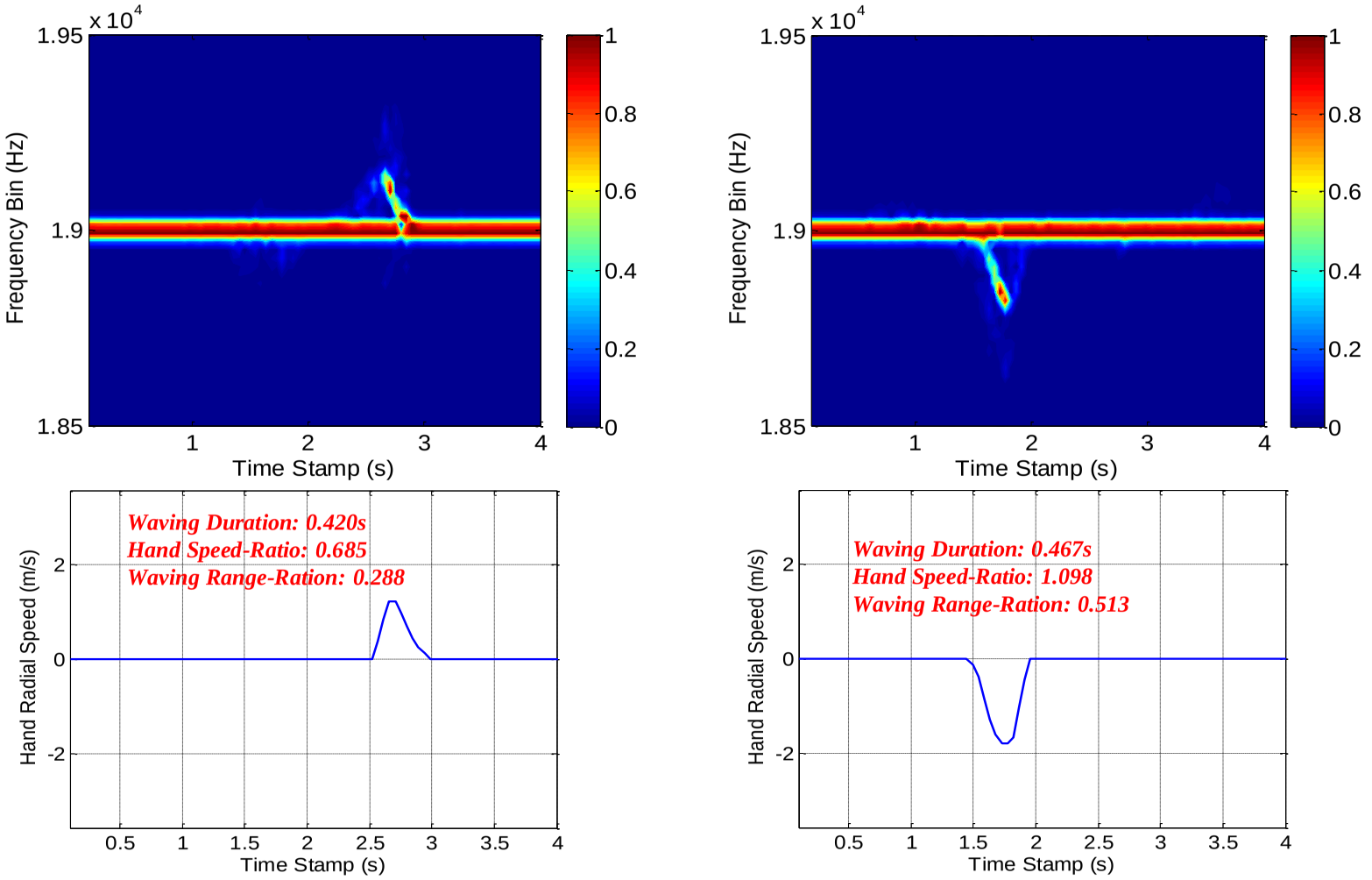


Fig. C.1 The echo spectrograms and the detected hand motion attributes: (a) Up-Down; (b) Down-Up. *We can distinguish different hand gestures via the waving directions, being similar to current hand-gesture recognition systems.*

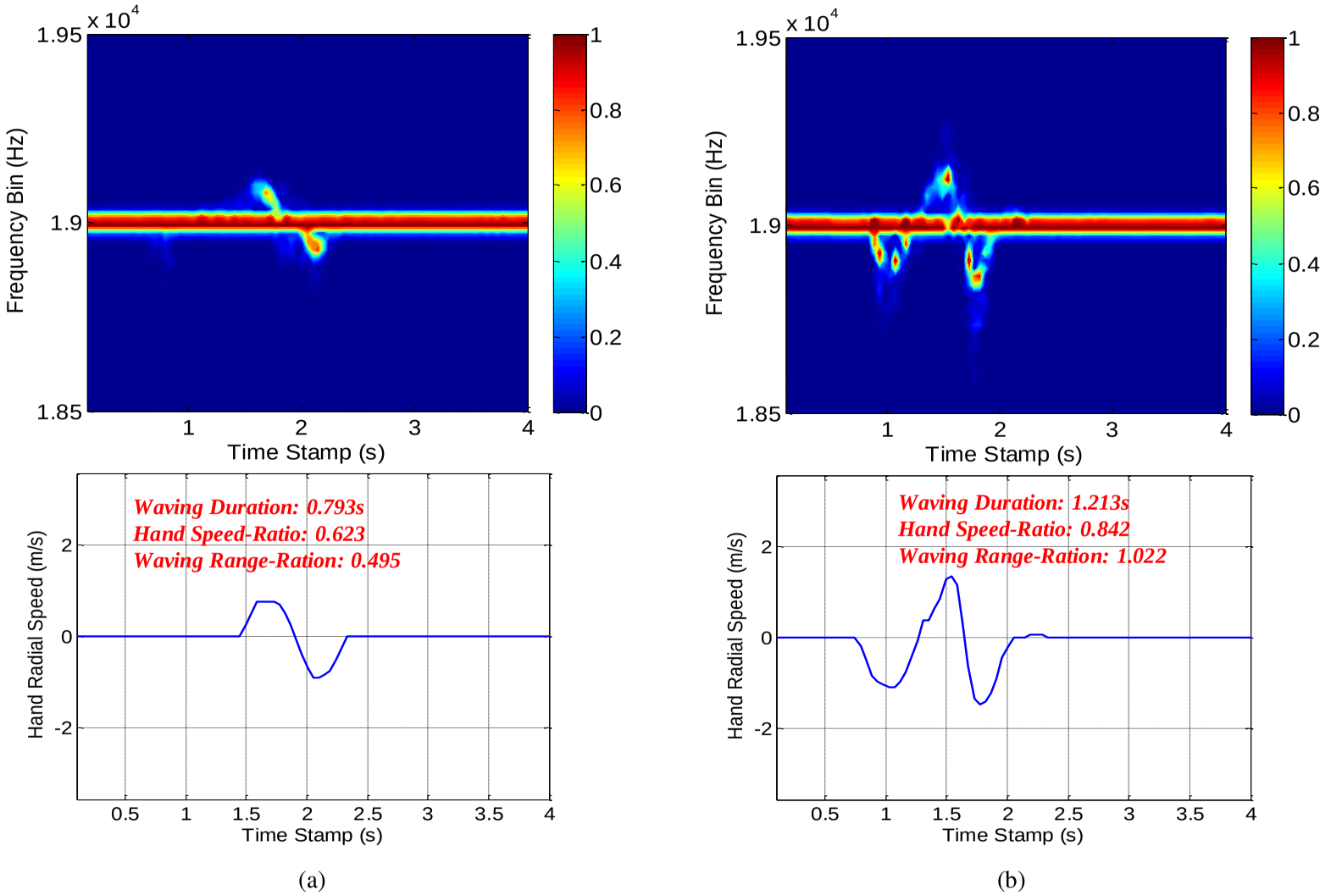


Fig. C.2 The echo spectrograms and the detected hand motion attributes: (a) Right-Left; (b) Anticlockwise Circle. *We can distinguish different hand gestures via the waving directions, being similar to current hand-gesture recognition systems.*

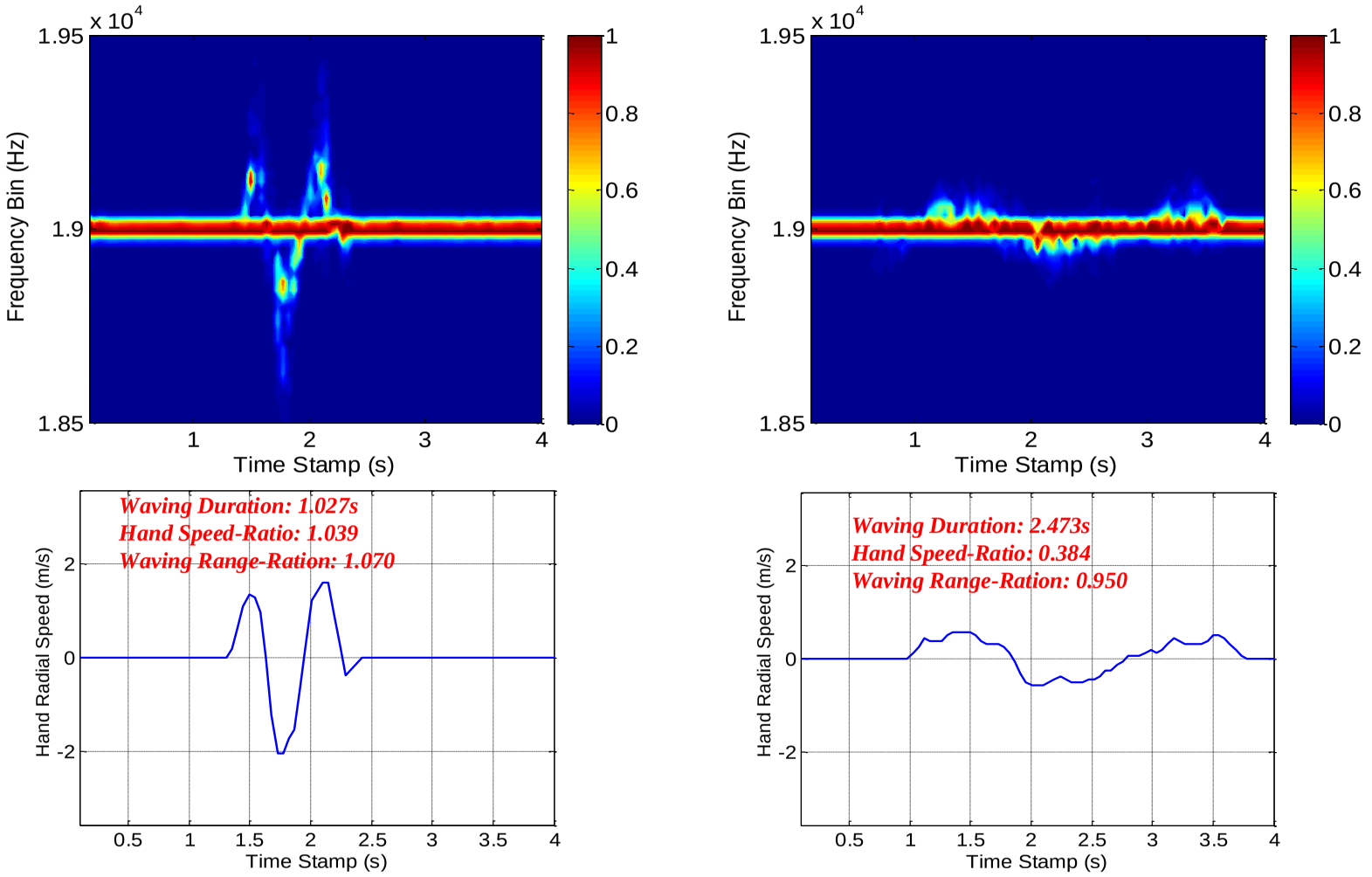


Fig. C.3 The echo spectrograms and the detected hand motion attributes for a same hand waving: (a) Fast-Speed Clockwise Circling; (b) Slow-Speed Clockwise Circling. *We can distinguish hand gestures (a) and (b) by the speed-ratios even though their waving trajectories are same.*

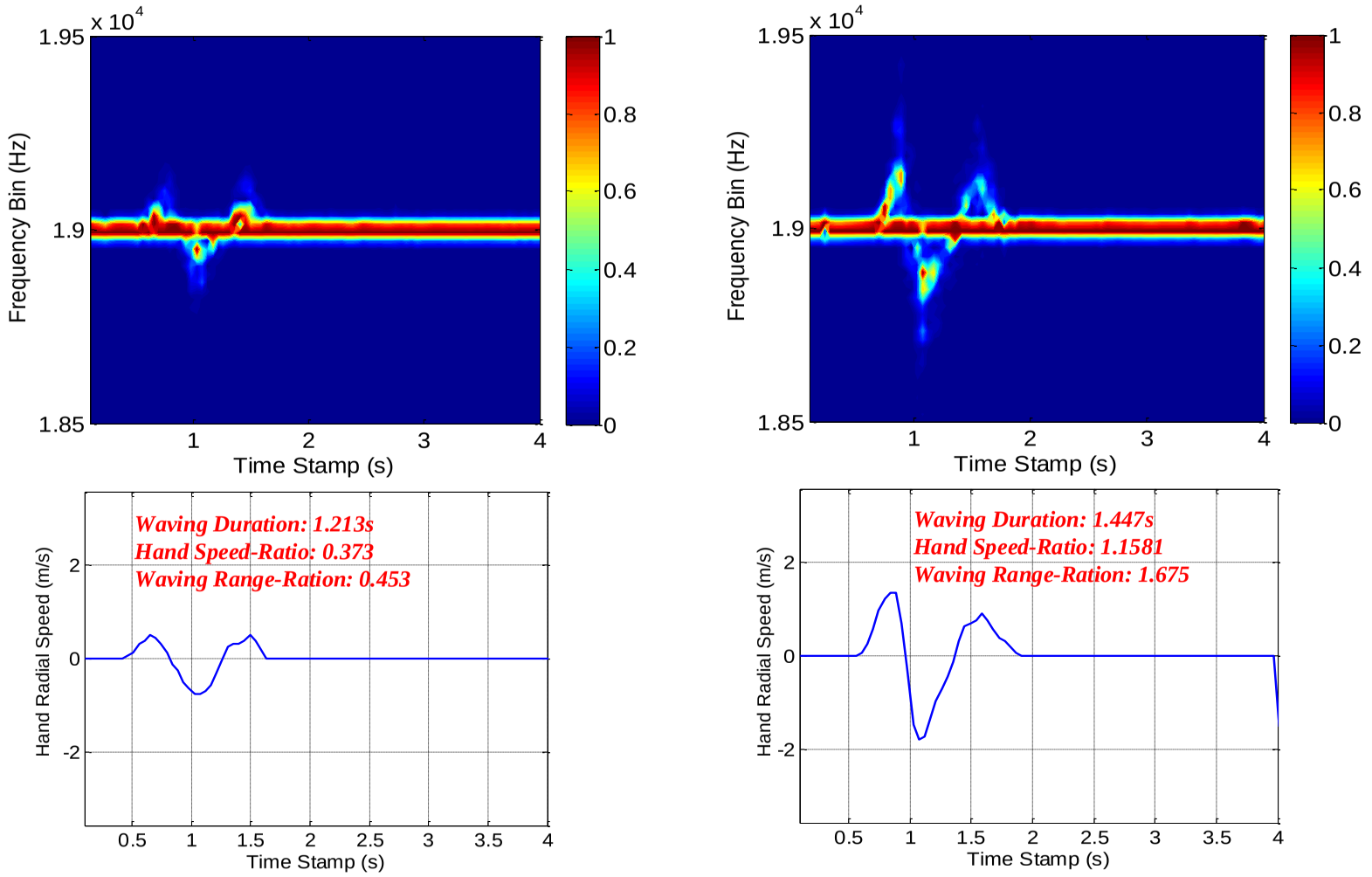


Fig. C.4 The echo spectrograms and the detected hand motion attributes for a same hand waving: (a) Small-Range Clockwise Circling; (b) Large-Range Clockwise Circling. *We can recognize hand gesture (a) and (b) by their range-ratios even though their waving trajectories are same, which enables our multi-modal hand motion detection and to advance current related systems.*