Title	スマート環境向けの人体EMIベースのアクティビティ認識と予測
Author(s)	都，業剛
Citation	
Issue Date	2020-03-25
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/16652
Rights	
Description	Supervisor:丹　康雄，情報科学研究科，博士

Doctoral Dissertation

Human Body EMI-based Activity Recognition and Prediction for Smart Environments

Yegang Du

Supervisor: Professor Yasuo Tan

School of Information Science
Japan Advanced Institute of Science and Technology

March, 2020

Abstract

This dissertation proposed a novel interaction-based HAR approach. The approach can be scalable enough to be applied to different indoor scenarios. In this dissertation, we chose smart home and smart library as the main scenarios. The dissertation makes a deep analysis on indoor human activity, and finally determines to utilize the interaction between human and objects to infer the human activity. Because only in this way, the proposed approach can be scalable, flexible, and avoid of "cold start" problem which appears commonly in data-driven approaches. The proposed approach utilized passive RFID technology to recognize the interactions between human and the objects of daily life. And with the help of machine learning and deep learning, the proposed approach can combine the recognized low level activities to infer the high level activities. So that the proposed approach can be applied to recognize all grain activities including both fine-grain and coarse-grain activities.

Smart Homes are generally considered the final solution for living problem, especially for the health care of the elderly and disabled, power saving, etc. Human activity recognition of smart homes is the key to achieving home automation, which enables the smart services to automatically run according to the human mind. Recent research has made a lot of progress of this field; however, most of them can only recognize default activities, which is probably not needed by smart homes services. In addition, low scalability makes such research infeasible to be used outside the laboratory. In this study, we unwrap this issue and propose a novel framework to not only recognize human activity but also to predict it. The framework contains three stages: recognition after the activity, recognition in progress, and activity prediction in advance. Furthermore, using passive RFID tags, the hardware cost of our framework is sufficiently low to popularize the framework. In addition, the experimental result demonstrates that our framework can realize good performance in both activity recognition and prediction with high scalability.

In the library, recognizing the activity of the reader can better uncover the reading habit of the reader and make books management more convenient. In this study, we present the design and implementation of a reading activity recognition approach based on passive RFID tags. By collecting and analyzing the phase profiling distribution feature, our approach can trace the reader's trajectory, recognize which book is picked up, and detect the book misplacement. We give a detailed analysis of the factors that can affect phase profiling in theory and combine these factors of relevant activities. The proposed approach recognizes the activities based on the amplitude of the variation in phase profiling, so that the activities can be inferred in real time through the phase monitoring of tags. We then implement our approach with off-the-shelf RFID equipment, and the experiments show that our approach can achieve high accuracy and efficiency in activity recognition in a real-world situation. We conclude our work and further discuss the necessity of a personalized book recommendation system in future libraries.

Keywords: Wireless sensing, human activity recognition, activity prediction, activity of daily living, passive RFID system

Acknowledgments

First of all, the author takes this opportunity to express the sincerest gratitude to his parents, who give him the life, gives him great love and fully understanding through all these years, which is his strongest motivation to continue his study and pursue his life goals.

The author wishes to express his sincere gratitude to his principal supervisor Professor **Yasuo Tan** and Associate Professor **Yuto Lim** of Japan Advanced Institute of Science and Technology for their constant encouragement, inspiring instruction and patient guidance throughout this research endeavor.

The author also would like to thank Professor **Keqiu Li** and Professor **Wenyu Qu** of Tianjin University for giving him the opportunity to internship as a researcher, and their insightful comments on his research. Also thanks to all of the members in Tianjin Key Laboratory of Advanced Networking (TANK Lab), who provide him so much support on both research and life.

The author would like to thank Dr. **Xiulong Liu** for his technical expertise and dedicated support, which help the author complete his minor research and inspire his main research.

The author also would like to express his gratitude to the members of examination committee of his doctoral dissertation, Professor **Yoichi Shinoda** and Associate Professor **Ken-ichi Chinen**, for their valuable comments and suggestions to improve the quality of this thesis.

In addition, the author would like to express to his gratitude to China Scholarship Council and Japan Advanced Institute of Science and Technology for providing him financial support for this work. Also thanks to NTT, HONDA, NEC, and Ministry of Internal Affairs and Communications of Japan for their kind financial support.

Last but not least, the author wishes to appreciate all partners of "Tan and Lim Lab" and all the staff in Japan Advanced Institute of Science and Technology for their help and cooperation.

Contents

Abstract **i**

Acknowledgments **iii**

1 Introduction **1**

1.1 Smart Environments and Human 1

1.1.1 The Trend of IoT 2

1.1.2 Human-oriented Service in Smart Environments 3

1.1.3 Human State Sensing 5

1.2 Statement of Problems 5

1.2.1 The Gap Between Human and Computer 5

1.2.2 The Characteristic of Human Activity 6

1.2.3 The Hierarchical Human Activity 8

1.3 Dissertation Purpose and Objectives 9

1.4 Structure of the Dissertation 10

2 Background and Motivations **12**

2.1 Introduction 12

2.2 Human Activity Recognition (HAR) Systems 13

2.2.1 Visual-Based HAR 13

2.2.2 Sensor-Based HAR 17

2.2.3 Data-Driven HAR 20

2.2.4 Knowledge-Driven HAR 21
2.3 From Action to Activity 23
2.4 Interaction-based HAR 25
2.5 Summary 29

3 Low-level Activity Recognition 31
3.1 Introduction 31
3.2 RF-Switch: A Novel Wireless Controller in Smart Home 32
3.2.1 Home Control Systems 32
3.2.2 Architecture of Wireless Controller 33
3.2.3 On-Off Control 34
3.2.4 Volume Control 35
3.3 Experiments and Results 37
3.4 Summary 38

4 High-level Activity Recognition and Prediction 39
4.1 Introduction 39
4.2 Off-line Activity Recognition 43
4.2.1 Definition of High-Level Activity 43
4.2.2 Object Usage Detection 44
4.2.3 Activity Fusion and Segmentation 47
4.3 On-line Activity Recognition 49
4.4 Activity Prediction with LSTM Model 52
4.5 Experiments and Results 54
4.5.1 First Stage 55
4.5.2 Second Stage 56
4.5.3 Third Stage 60
4.6 Summary 61

5 Reading Activity Recognition in Smart Library **63**
5.1 Introduction 63
5.2 Preliminaries and Inspiration 65
5.2.1 RF Phase Values 65
5.2.2 Rotation of Tags 66
5.2.3 Multi-path Effect 69
5.3 System Design 71
5.3.1 Activity Definition 73
5.3.2 Data Collection 74
5.3.3 Phase Dispersion of Activity 75
5.3.4 Thresholding 79
5.3.5 Postprocessing 80
5.4 Experiment and Evaluation 81
5.5 Misplaced Book Detection 82
5.6 Summary 84

6 Conclusion **86**
6.1 Discussion and Conclusion 86
6.2 Directions and Future Works 88

Bibliography **100**

Publications **101**

List of Figures

1-1 A simplified architecture of smart home 6

1-2 Comparison between the existing human activity recognition (HAR) approaches and the ground truth in the real world. (**a**) Result of the existing HAR approaches; (**b**) Activities in the real world. 7

1-3 Maslow's hierarchy of needs. 9

2-1 The categories of human activity recognition. 14

2-2 Modification of activity theory. 24

2-3 Examples of phase change and the corresponding actions. 27

3-1 The architecture of RF-Switch. Note that RF-Switch can be accessed into different kinds of smart home platform. 34

3-2 The structure of tag. 35

3-3 Sweep a finger from left to right on the surface of a tag. 36

4-1 Three-stage framework to recognize and predict human activity in a smart home. 43

4-2 Interactions and the corresponding phase changes. (**a**) Passing by; (**b**) Picking up. 45

4-3 Object-usage state vector changes with time, and we can generate "On queue" and "Off queue", respectively. 47

4-4 We propose two strategies to determine the start time and end time. O_1 and O_2 are objects that belong to one activity, and O_3 belongs to some other activity. (**a**) No interruption between two objects; (**b**) Interruption between two objects. 49
4-5 Activity sequence and recurrent neural network (RNN) model. 54
4-6 Accuracy of Naive Bayes and long short-term memory (LSTM) solution. 61

5-1 Measuring the phase of a tag rotating along each axis. 67
5-2 Measured phase of a tag rotating along three axes. 68
5-3 Measuring the phase under multi-path effect. 69
5-4 The phase changes while L is set from 10 cm to 110 cm. The phase in the red frame is affected by the multi-path effect. 70
5-5 The architecture of activity recognition approach. F1, F2, F3 and F4 in the second picture represent four features respectively. The upper line is an example of walking before a book and the under line is an example of picking up a book. 72
5-6 The white antenna is placed at a 45° angle to the ground to ensure it can cover as much region as possible, because of its fan-shaped radiation regions. 74
5-7 Half-wave effect and the regularization. The blue circle is the original phase value and the red star is the phase after regularization. 76
5-8 The sequence of time slot that caused by reader's movement reflects the order of books. 83

List of Tables

2.1 Contrast of state of the art wireless sensing technologies for action sensing. 25

3.1 The performance of RF-Switch on different materials. 37

4.1 Object-usage detection via interaction. 46

4.2 Result of object–usage detection. 55

4.3 Example of verification matrices of TP, FN and FP. TP represents the quantity of true positive; FN represents the quantity of false negative; and FP represents the quantity of false positive. 57

4.4 Confusion matrix of recognized activities. 59

5.1 Performance of detecting walking past. 81

5.2 Performance of detecting picking up. 82

Chapter 1

Introduction

1.1 Smart Environments and Human

In the past decades, the relevant technologies of edge computing, communication, cloud computing have achieved speedy development. The Micro-electromechanical Systems (MEMS) have greatly reduced the size weight of various devices. And in the meanwhile, the wireless communication technology enables such edge devices to transmit the information to the remote servers without irritating cables. The improvement of computing power promotes the applications of machine learning and deep learning algorithms. All the above technologies make it possible for the Internet of Things (IoT) to obtain data, analyze data, and provide services.

Along with the development of IoT, the population of the world also grows rapidly in recent years. According to the report of United Nations, the world population will reach 9.7 billion in 2050, and more than 11 in 2100. The huge population brings not only opportunities but also great challenges to different countries. Especially today when the population structure and social form has changed dramatically, people become more and more independent than ever. Actively or passively, people are surrounded by electronic equipments.

Thus, we have to think about how to make those equipments and devices improve our quality of daily life. The answer in this dissertation is to build human-oriented smart environments framework to give devices the ability to understand human, and future completely change the relation between devices and human.

1.1.1 The Trend of IoT

The term of IoT was first mentioned by Kevin Ashton in 1999 to propagate the supply chain management[1]. However, the definition has changed to be more inclusive in the past decade[2]. Here, the definition of "Things" includes far more than the hardware devices. All kinds of knowledge and information that could be sensed and conveyed by computer belong to "Things". And the goal of IoT is not only to connect the "Things" but also to analyze them and make the whole system run automatically.

IoT comes from the concept of ubiquitous computing, which is a network of smart devices. Thanks to the development of embedded systems, Wireless Sensor Networks (WSN), real-time analytics, control systems, IoT has been out of the laboratory and widely applied in different areas.

In industrial area, IoT is used to acquire the data from the connected equipment to regulate and monitor the industrial systems. Several years before, such industrial IoT systems can only get the data from the equipments embedded with sensors, and transmit the data to human. The IoT systems are highly dependent on the experience of workers and not robust enough. While when Industrial Internet of Things (IIoT) hugs intelligence, that enables content-aware data processing, the IIoT systems will greatly improve the efficiency of manufacturing and save more energy[3]. Here, industrial big data plays an important role in IIoT[4]. The intelligent IIoT manufacturing systems will not rely on human any more, and realize self-configure for resilience, self-adjust for variation, self-optimize for disturbance.

In agriculture area, IoT applications are used to collect the data of temperature, rainfall, humidity, wind speed and so on. The data are utilized to provide automatic decision in the farm management. For example, Farmers now can precision fertilization programs with the help of agriculture IoT system[5].

In infrastructure area, IoT applications are usually utilized to build smart city, which can continuously monitor and automatically control the infrastructures like railway, bridges, parking lots, etc[6]. Another well-known application of IoT is smart grid, which is used to balance the power usage and energy generation to optimize energy consumption in the real time[7].

From the above applications, we can see that intelligence and autonomous control are the key trends of current IoT. Although they are not in the original concept of IoT, Ambient intelligence (AI) and autonomous control greatly promote the development of IoT. IoT systems equipped with AI real-time analysis module can realize autonomous behavior by collecting and analyze the context information. Thus, the smart environment systems in the future will not act as a tool of human, but a reliable partner of human.

1.1.2 Human-oriented Service in Smart Environments

Apart form the above areas, a major application of IoT in consumer area is home automation, which is so-called “smart home”. All of the technologies are invented to serve humans, and IoT is no exception. IoT applications in the above areas are used to improve the efficiency of manufacturing or save energy, which is relative far from the daily life of ordinary person. While smart home is very closed to everyone’s life, which make it a most promising application of IoT.

Note that, the “home” here is a generalized concept rather than “house”. In this dissertation, we define “home” as indoor space where human exists in the daily life, e.g. house, office, hospital, library, and gymnasium. Smart home systems can be implemented in these places to provide human-oriented services, which then become “smart environment”. The services aim to reduce the inconvenience of human and improve the quality of daily life.

The concept of building a home that can automatically control and manage the resources for the inhabitants has been put forward many years before. Smart environment can be considered to be a kind of context-aware system that provides services based on the knowledge it sensed in home environment. To achieve this goal, smart home systems typically utilize a home gateway to connect the controlled devices and sensors. And now the users can control the system through different kinds of remote terminals, such as tablet, Web, computers, and mobile phones.

Such smart home platforms enable home devices to collect surrounding data and communicate with central processing module in bidirectional way. Many companies have produced their own smart home platforms to take the lead, e.g. Apple HomeKit[8], Samsung SmartThings[9], Google Nest[10], etc. A typical scene of such smart home platforms is to allow users to remote control the home devices through their smart terminals or voice assistant. For example, users can ask "Siri" to turn on the light in living room, while they do not neet to get up from the bed. Obviously, it is not intelligent enough to be called smart home.

Fortunately, smart home can do much more than this in the view of researchers. For instance, the health care of inhabitants is an essential function of smart home, especially for the disabled and elder people. In this part, different health care services represented by fall down detection draw a lot of attention of researchers. The vision is to realize timely remote health assistance by monitoring the physical conditions of users[11]. The living conditions control is another function of smart home. Smart home can automatically control the overall devices to keep the home environment comfortable, including light, temperature, humidity, etc[12]. Personalized recommendation is also a potential service for users. Based on the behavioral habit, smart home can provide recommendation and advices to improve the quality of life. Power saving in smart home is that the electric appliances can be automatically controlled according to the occupancy of users.

Services including but not limited to the above functions make a real smart home, compared with existing smart home systems. However, Such services all rely on the knowledge of the human state. So, the lack of human state sensing becomes the core restriction that obstructs the current smart home to be real smart.

1.1.3 Human State Sensing

To achieve real smart home, we need to know the state of human. We divide the human state into two parts: inner state and external state. The difference between the two kinds of states is that if the state can change according to human mind.

Such inner state of human can already be monitored in different appropriate ways. For example, heartbeat is a typical inner state, which can not be directly controlled by human mind. Smart watches represented by Apple Watch can monitor not only heartbeat but also electrocardiogram (ECG)[13]. Adib et al. have proposed a wireless sensing technology to monitor the breathing and heart rate without contact. Panasonic has produced EW-NK63 to measure the calories consumed in everyday[14].

The external state of human contains more information than the inner state, which is more important to the smart home. However, monitoring the external state of human, in other words, human activity in home environment is still a challenge that requires more research. The existing work is introduced in Chapter 2.

1.2 Statement of Problems

Now, we already know that indoor human activity recognition is a key bottleneck to make a real smart home. However, existing research does not work well in this area because of several reasons. In this section, we provide a deep analysis of the reason why it is hard for computer to monitor the human activity at home.

1.2.1 The Gap Between Human and Computer

As depicted in Figure 1-1, the knowledge of home environment contains two parts: surroundings and inhabitants. Due to the development of sensing technology, it is not difficult for smart home to know the state of the surroundings. With those information, smart home can achieve some degree of automation. However, to provide

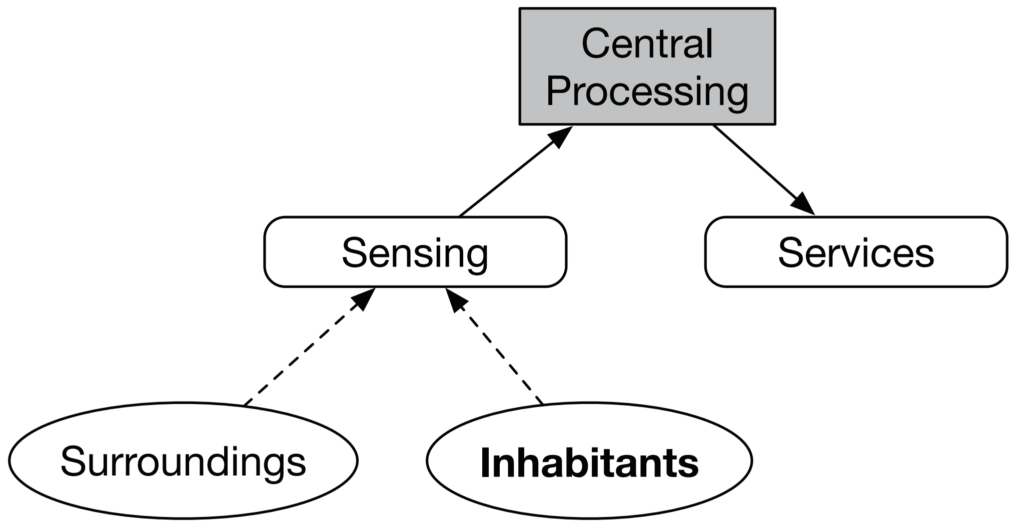


Figure 1-1: A simplified architecture of smart home

human-oriented services, the activity of human has to be taken into consideration.

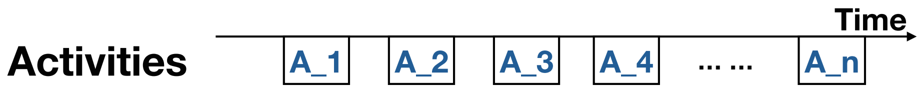
Unfortunately, there is a natural gap between human and computer. Millions years ago, human start use tool in hunting and cultivation to rule the world. Today, human invents computer systems as a tool in all aspects of daily life. The role of computer is still somehow the same with the stone. Although many kinds of interface between computer and human have been proposed to help human use computer with less troubles. One thing is not changed for many years. The computer systems still work in a positive way, which restricts computer to wait for the order from human.

Computer systems rely on data, which is impossible to get from human directly. That is to say, the activity of human will not produce data for computer to understand human mind. And this leads to the gap between computer and human.

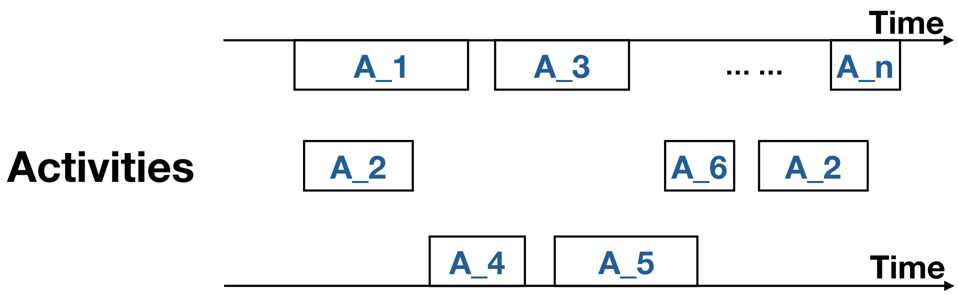
1.2.2 The Characteristic of Human Activity

The characteristics of human activity also bring troubles to computers to understand the activity. Here, we clarify them to better solve the problems.

To recognize a human activity, we have to know the habit of the human at home. Here, we explore the features of human activity at home as follows, and they may



(a)



(b)

Figure 1-2: Comparison between the existing human activity recognition (HAR) approaches and the ground truth in the real world. (**a**) Result of the existing HAR approaches; (**b**) Activities in the real world.

be the reasons why it is difficult to build a general HAR approach.

- **Concurrency.** This is the biggest gap between the result of the HAR approach and the ground truth in the real world. A human usually performs different activities concurrently. As depicted in Figure 1-2b, activities A_1 and A_2 are simultaneously in progress—while the existing HAR approaches are basically discriminative, meaning they can only recognize one result at a moment, as depicted in Figure 1-2a.

- **Multiplicity.** It means that the sort of activities could be different as seen from different granularity. For example, the activity of "cooking" may involve sub-activities such as "washing vegetables", "cutting ingredients", "frying" and so on. It is difficult to recognize a full-grain-size activity, as the HAR approach does not even know what kind of activity the smart-home platform needs to know. In this study, we grade the activities according to the semantic meaning they have. A high-level activity has more semantic meaning than that of lower ones [15].

- **Complexity.** An object or device could be used in different activities, and an activity may also use several devices. This means it is not impossible to infer the activity directly from object usage.

- **Diversity.** Owing to the difference in cultural, race, region, and even character, individuals may perform the same activity in completely different ways. This difference causes huge trouble to the training-based approach, as the trained model in the laboratory could not be applied to real houses.

- **Randomness.** Although the individuals have specific behavioral habits, it is not easy to excavate such patterns. This is because large randomness in activities exists in daily living, and because the next activity is decided not only by the current activity, but also by the earlier ones.

The above-mentioned five features render the existing research works impractical in smart homes. To monitor human activity, we have to face the above five features to make the solution work well in real world.

1.2.3 The Hierarchical Human Activity

Maslow proposed his theory in psychology about humans' innate curiosity in his 1943 paper[16]. As depicted in Figure 1-3, Maslow's hierarchy of needs is usually portrayed in a shape of pyramid, where most fundamental needs are at the bottom and the need of self actualization is at the top.

Inspired by Maslow's theory, we believe the activities of human also follow a hierarchy structure. Because, human performs the activities according to their mind. And the mind is from the human need, at the same time. Thus, similar with hierarchy of needs, different activities of human belong to different hierarchies.

Actually, several researchers have paid attention in this area and form activity theory. Based on activity theory, an activity is seen as a system of human "doing" whereby a subject works on an object in order to obtain a desired outcome. Ac-

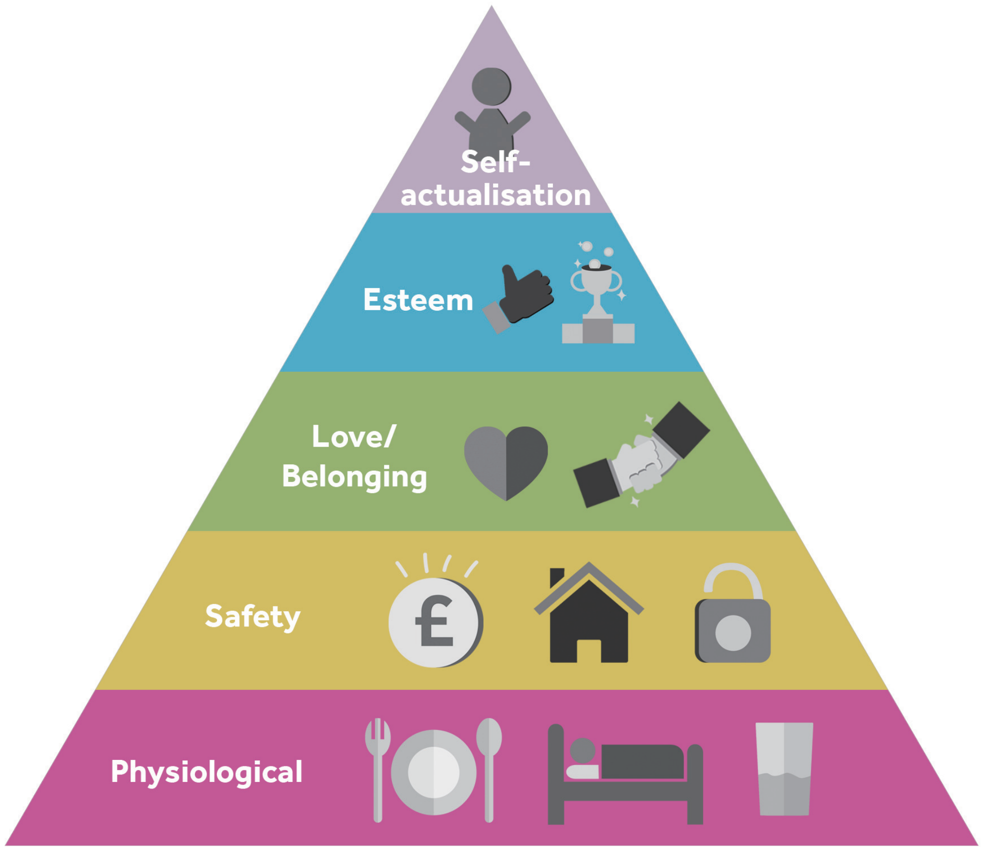


Figure 1-3: Maslow's hierarchy of needs.

cordingly, Saguna proposed a context-driven activity theory to show that complex activity can be inferred by simple activities[17].

1.3 Dissertation Purpose and Objectives

In this dissertation, I am motivated to fulfill the gap between computer and human, by giving the computer the ability to understand human. And this will subvert the traditional relation between computer and human, enabling real smart home systems. Smart environment systems will be able to run in an active way to provide human-oriented services without human order. To achieve this goal, this dissertation is to propose scalable and practical frameworks to recognize human activity. Here, this dissertation focuses on two scenes: house and library.

In narrow sense of smart home, we want to not only recognize the human activity but also try to predict the activity. A new way to control the equipment at home will be proposed to realize more flexible home control. The final goal of this part is to make smart home could remember the habit of the inhabitant's daily life and guess what the inhabitant wants to do next. With help of this kind of knowledge, smart home can enable many kinds of functions to improve the quality of human life.

And in the scene of library, we would like to recognize the behaviors of reader and to further better excavate the reading habit of the readers and bring convenience to the book management. To accomplish these, we make effort to present the design and implementation of a reading activity recognition approach with off-the-shelf equipment.

1.4 Structure of the Dissertation

The remainder of this dissertation is organized as follows:

- Chapter 2. Background and Motivations

  In Chapter 2, a review of the existing research in the area of HAR is introduced, and we than introduce the motivation of our interaction-based HAR which utilizes passive RFID technology. We briefly introduce the applicability of using passive RFID technology to detect the human action in different scenarios.

- Chapter 3. Low-level Activity Recognition

  Then, Chapter 3 describes a way to recognize the low-level human activity at home. A novel wireless controller in smart home is proposed as a practical application. The wireless controller allows user to move the controller to where they want as easy as blowing off dust.

- Chapter 4. High-level Activity Recognition and Prediction

  In Chapter 4, a three-level framework is proposed to recognize and predict human activity in smart home. The framework enables smart home system to monitor human activity in real-time and predict what the user wants to do next at home.

- Chapter 5. Reading Activity Recognition in Smart Library
  After that, Chapter 5 introduces a novel framework to recognize the reading activity in future smart library. The framework can recognize which book is picked up and trace the reader at the same time.

- Chapter 6. Conclusion and Future Work
  Chapter 6 summarizes the dissertation and discusses the future trends in the area of HAR.

Chapter 2

Background and Motivations

2.1 Introduction

In the filed of human activity recognition, there are several categories of existing research work that are related to our work. And all the existing work inspires us to propose our work with more efficiency and effectiveness. In this chapter, I will discuss them in detail.

Inspired by the existing work, we proposed our interaction-based HAR approach using passive RFID technology. In this chapter, we also discussed the motivations of our approach, which are mainly about the relation and difference between action and activity. The main idea of our approach is to detect the action first and then infer the relative high level activity.

With the help of machine learning and deep learning algorithms, the proposed approach can not only recognize the full grain human activities in real time, but also predict the next possible activity at smart home. Moreover, the proposed approach can be applied to other indoor scenarios. We take library as an example, and introduce briefly about the interaction-based HAR in smart library.

2.2 Human Activity Recognition (HAR) Systems

Several years ago, the concept of smart home has been proposed with the realization of the great technology development[18]. Now, we say smart home as a general representative of smart environment that covers different places. Even though the places are different, the purposes of smart home are the same, which is to provide an Ambient Assisted Living (AAL) to realize an active interactive environment for human to improve the quality of life. For example, smart hospitals can monitor the health state of patient and provide prompt treatment without delay. No mater where the place is, one basic knowledge has to be acquired by computer systems to realize smart home, that is the activity of human.

To obtain the reliable and accurate knowledge of human activity, a lot of research have been done in different ways. However, this is not an easy task. The reasons has been introduced in the last chapter. While, in this section, we will show the effort from different researchers in recognizing human activity. For a typical HAR system, the first task is to obtain the data from different sensors, and then different algorithms will be utilized to analyze the data to decide what the user is doing. As depicted in Figure 2-1, if we divide the existing work according to the type of sensors, there are two kinds of approaches of HAR: vision-based HAR and sensor-based HAR. While if we divide the existing work based on the type of activity model, there are also two kinds of HAR: data-driven HAR and knowledge-driven HAR. In this section, we will introduce them respectively.

2.2.1 Visual-Based HAR

The most widely used way to recognize human activity in the early stage is to utilize visual data. Usually, visual-based HAR approaches use video cameras to monitor the actor's behavior and the relative environmental changes. The video data are video sequences or digitized image data. The visual-based approaches exploit computer vision techniques to distinguish different activities. They utilize feature extrac-

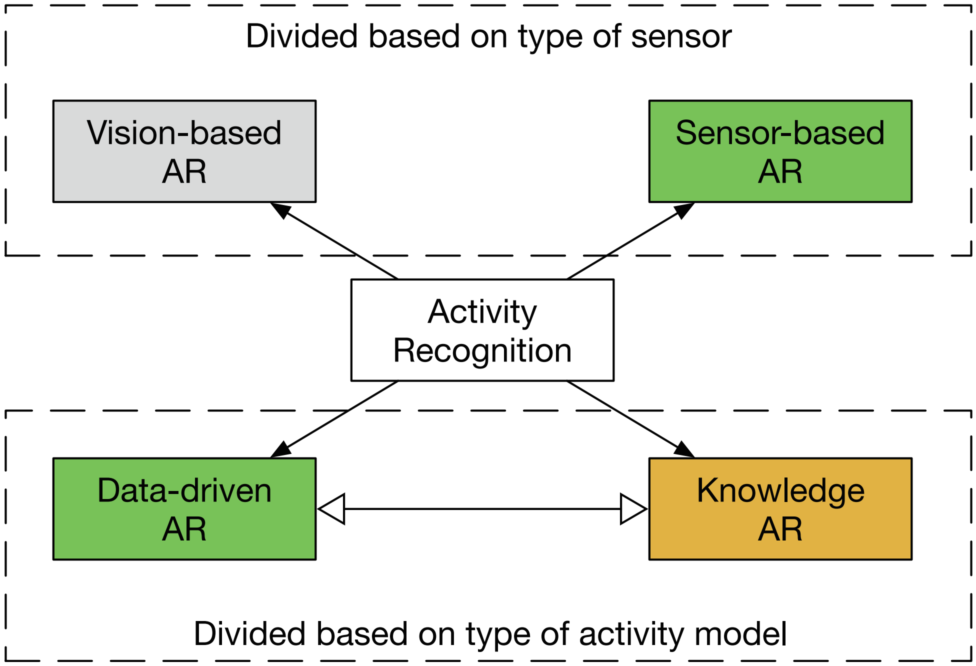


Figure 2-1: The categories of human activity recognition.

tion, structural modeling, movement segmentation, action extraction, and movement tracking to analyze visual data to realize patten recognition[19].

The advantages of visual-based HAR include several aspects. The first is that cameras are easy to be implemented in different places. Then, the price of cameras is not that high, so that the cameras could be equipped in a huge amount. The last is that computer vision in the area of pattern recognition has been well developed in many years. These advantages make visual-based HAR approaches achieve success in several scenarios, such as surveillance, robot learning, and security.

To recognize the human activities using video data, one basic idea is to isolate the human body from the background. The procedure is called back-ground subtraction or fore ground extraction. After obtaining the silhouette, the 2D shape of the silhouette can be used to recognize the activity of human. Veeraraghavan et al. emphasized the effectiveness of shape features, they utilize the shape features to recognize the human motion[20]. To alleviate the influence of view-point changes, multiple cameras are used to extract silhouettes in different viewpoints in the research of Xu et al[21]. Weinland et al. derived Motion History Volumes (MHVs) by stacking 4D silhouettes using four orthogonal cameras.

Optical flow is another way to extract silhouettes from a dynamic background. Activity is represented as a sequence of posture according to the research proposed by Lu et al [22]. And each key posture was recorded in a key frame. After recognizing the key posture, they matched the posture sequences with the defined activity to confirm the specific activity.

An activity video could be treated as a series of images that contain activity sequences. By concatenating the frames along the time, we can form the 3D Space-Time Volume (STV). Such STV contains not only spatial information but also time information. Shechtman et al. proposed an approach to compare volumes based on their patches. The method requires no prior modeling, while it can still handle the complex dynamic scenes and detect multiple activities[23].

Instead of extracting the silhouette or STV and encoding them as a whole, local representations process activity video as a collection of local descriptors. Scovanner et al. proposed a 3D Scale-Invariant Feature Transform (SIFT) descriptor and utilize it in HAR[24]. The videos are described as a bag of spatiotemporal words using the 3D SIFT descriptor. Moreover, a feature grouping histogram which groups the co-occurred words out of the original one is used to build a more discriminative action video representation and finally used for classification.

Lu et al. utilized principle component analysis (PCA) to project the original histogram of oriented gradients (HOG) descriptor to a linear subspace. Then, the descriptor was utilized to track and recognize the activity simultaneously[25].

Apart from the early stage visual-based HAR approaches using RGB cameras, approaches using depth cameras can achieve more information for recognition. Low-cost depth cameras have been well developed. And we can easily achieve depth map with off-the-shelf devices such as Microsoft Kinect[26]. With Kinect SDK, researchers can directly obtain the skeletal joint positions in real-time. Both the depth map and skeletal information can help a lot in HAR problems.

Li et al. introduced the action graph model, which represents activities using several salient postures serving as nodes in action graph[27]. All activities share same posture sets and each posture is characterized as a bag of 3D points from the depth maps. However, involving all the 3D points is computationally expensive; thus, a simple and effective method to sample the representative 3D points is proposed, achieving over 90% recognition accuracy by sampling approximately 1% points according to their report.

Wang et al. proposed a depth-based feature called local occupancy pattern (LOP) to describe the occupancy of the neighborhood of each point. A data mining solution is also leveraged to select the active joint feature subset for a particular activity. Then, actionle ensemble which is linear combination of actionlets is achieved to represent each activity[28].

Although, the existing research to recognize the activity with visual data are popular in the early stage. The shortcomings of these approaches greatly restrict the appliance in current smart home. The most important issue is privacy problem. As introduced above, smart home tend to provide services to human and build an active interaction. Visual data contain too much sensitive information of the user, and no one want to be monitored by cameras all the time. No mater in the public places or private place, the privacy of user need to be well protected. Even though some research have been done to encrypt the visual data, there is still great risk to disclose the privacy when the computer systems are hacked and the cameras are taken control by hackers. Also, the visual-based approaches usually perform well when they recognize simple activities, which we say low-level activities or actions. When we deal with high-level activities, visual data will include too much noisy and lead to mis-recognition[29]. Besides, visual-based approaches rely on high computational complexity, which makes them hardly to be utilized in different places.

2.2.2 Sensor-Based HAR

The concept to utilize sensors to recognize the activities has been proposed since the late 1990s. It was firstly pioneered by the research of the Neural Network house to enable home automation[30]. After that, extensive research has been undertaken to explore the use of sensors to create smart appliances and activity recognition in various scenarios. Thanks to the development of ubiquitous and mobile computing, HAR with sensors keep receiving attention from researchers till now.

Wearable Device-Based HAR

In the early stage, most research made use of wearable devices to monitor the activity of users. They attach devices to human bodies, or utilize the portable devices like mobile phones, to obtain sensor data. In that stage, they can mainly recognize the physical activities, which are mostly like simple actions, such as walking and running. Since the wearable devices can be embedded into clothes, shoes, eyeglasses, and mobile devices, they can easily achieve the information such as body position and movement. Then researchers can utilize such information to infer the relative activities.

Accelerometer sensors are the most frequently used wearable sensor to recognize the human activity. Lee and Mase proposed a dead-reckoning method to determine the user's location and recognize the user's behaviors such as sitting, standing, and walking, by measuring the acceleration and angular velocity through accelerometers and gyroscopes[31]. Lukowicz et al. proposed a way to recognize workshop activities utilizing accelerometers and body worn microphones[32].

Biosensors are also more and more popular in the area of HAR aiming to monitor the vital signs. Different biosensors can obtain different body information such as blood pressure, heart rate, EEG, ECG, etc. Researchers have found that when the human perform different activities, the biological information will change accordingly. Harms et al. identify body posture utilizing the information from a smart garment[33]. Similar work also includes [34] and [35].

Now, most smart phones equipped with multiple sensors become the best intermediary to obtain the compound data of human. For example, Anguita et al. provided a public dataset for human activity recognition using smart phones[36]. They utilized smart phones with inertial sensors to collect context information of the user, and then used Support Vector Machine to recognize different actions. Su et al summarized the HAR approaches with smart phones in their research[37].

Wearable devices-based HAR approaches achieved great success in monitoring the actions and gestures of human. However, they suffer from limitations at the same time. Even though the microelectronics technology has been well developed for several years, the size of wearable devices still makes troubles to users, especially to elder and babies. Also, wearable devices have to face the issues include ease of use, battery life and the effectiveness in the real-world scenarios. Besides, how to persuade human to ware the additional devices is another practical problem. Last but not the least, wearable sensors only work well when they are used to recognize low-level activities. Wearable devices are not suitable in most cases, when we want to recognize different complex and high-level activities. Because the simple physical movements information is not sufficient to infer activities that contain mixed physical motions.

Dense Sensing-Based HAR

To overcome the shortage of wearable devices, dense sensing-based HAR approaches have become the main trend. Using dense sensing, a large number of sensors, normally low-cost low-power, and miniaturized, are deployed in a range of objects or locations within an environment for the purpose of monitoring movement and behavior. As dense sensing-based monitoring embeds sensors within environments, this makes it more suitable to create ambient intelligent applications such as smart homes[19].

Sensors in smart homes can monitor the inhabitant's movements and environmental changes, so that the undergoing activities can be inferred by the sensor observations, thus providing in-time context-aware ADL information. Extensive research has been done in this issue to investigate the applicability of the approach in terms of sensor types, modalities, and applications.

Tapia et al. used environmental state-change sensors to collect information of the interaction between human and the objects[38]. Then, they recognized different activities such as toileting, bathing, and grooming. Wilson and Atkeson used four kinds of anonymous and binary sensors to simultaneous tracking and recognizing the activities[39]. Srivastava et al. exploited wireless sensor network to develop smart learning environment for young children[40].

Although the dense sensing HAR approaches do not require the uses to carry any devices and they attach abundance of different types of sensors to the objects or furniture to monitor the change of the environment to obtain the ambient information. The size of the sensors still need to be taken into consideration. Thus, RFID tags become the best choice to work as dense sensors. However, since the range of RFID antenna in the early stage is limited, RFID-based HAR approaches require users to ware a glove or bracelet that equipped with RFID reader and near field antenna to detect the object usage, while attaching RFID tags to the ambient objects. For example, Philipose et al. developed iGlove, working as wearable RFID reader that can detect the usage of specific object with unobtrusive tags[41]. And Fishkin et al. also developed the similar thing called iBracelet to detect the usage of tagged objects[42]. In these cases, the wearable RFID readers cooperate with dense RFID tags to sense the usage information and then infer the related activities.

To obtain the object usage information first, then infer the activity draw more and more attention in HAR. Patterson et al. proposed fine-grained activity recognition by aggregating abstract object usage[43]. Hodges and Pollack proposed a way to identify individuals from their behavior based on their way to interact with the object they use in daily activities[44]. Moreover, to obtain optimal recognition performance, wearable devices and dense sensors could work together. For example, Gu et al.

proposed an approach to collect multi-modal sensor information with both wearable sensors and object sensors[45]. They could recognize sequential, interleaved, and concurrent activities using a pattern-based algorithm.

Note that the dense sensing-based HAR approaches usually have to face different types of sensors data. To combine those multiple types of data, much more effort need to be done and more computation ability is required. The RFID-based approaches can only detect the usage of objects that operated by hands. While there are several objects do not interact with human in that way. In that cases, the existing work will fail to detect the object usage. Besides, the dense sensors also have to suffer some problems such as size, weights, price, communication and battery life. The above shortages restrict the existing approaches to work only in the laboratory of some specific scenarios.

2.2.3 Data-Driven HAR

Data-driven HAR can be classified into two main categories: generative and discriminative. In the generative approach, one attempts to build a complete description of the input or data space, usually with a probabilistic model such as a Bayesian network. In the discriminative approach, one only models the mapping from inputs (data) to outputs (activity labels). Generally speaking, discriminative approaches perform better than generative approaches, but are less interpretable[46].

The naive Bayes classifier (NBC) is the simplest possible generative approach, which has been proved to perform well in HAR. NBC model all observations as arising from a common causal source: the activity, as given by a discrete label. The dependence of observations on activity labels is modeled as a probabilistic function that can be used to identify the most likely activity given a set of observations. For example, Maurer et al. proposed a wearable sensor platform to recognize human activity and the location with NBC and achieved good performance[47]. There are also different kinds of generative approaches such as Hidden Markov Model (HMM)[48], Dynamic Bayesian Networks (DBNs)[49], etc. A drawback of the generative approach is that enough data must be available to learn the complete probabilistic representations that are required.

Perhaps, the simplest discriminative approach is Nearest Neighbor (NN), in which a novel sequence of observations is compared with a set of template sequences in a training set, and the most closely matching sequences in the training set vote for their activity labels. Bao and Intille proposed a method to use NN along with other base-level classifiers to recognize the human activity with accelerometer data[50]. Recent years, deep learning methods achieve great success in different pattern recognition problems. Researchers also utilized deep learning to recognize human activity with different kinds of sensor data. For example, Ordonez et al. proposed a LSTM-based model to recognize several human activities with wearable sensor data[51]. Ibrahim et al. proposed a hierarchical deep temporal model with visual data to solve group activity recognition problem[52].

Some work combine the above two kinds of approaches to achieve better performance. However, no mater generative approach or discriminative approach, they are all supervised machine learning approaches. They highly rely on the data with fine label. Unfortunately, it is very difficult to obtain well labeled data in the real world. Human tend to act the same activity in different ways. Thus, even when the HAR model is trained in the laboratory with the labeled data from the volunteers, the model can not be introduced to other scenes. Also, the data-driven HAR approach usually deal with fixed amount of activities. Because the model is fixed after training. This makes it impossible to extend the ability of recognizing more activities than the trained ones. The above shortages lead to that there are still no practical approaches that can be scalable to different scenes out of laboratory.

2.2.4 Knowledge-Driven HAR

Knowledge-driven HAR is motivated by real-world observations that for most ADL and working, the list of objects required for a particular activity is limited and functionally similar. Even if the activity can be performed in different ways, the type of the involved objects do not vary significantly. Knowledge-driven activity recognition is founded upon the observations that most activities, that take place in a relatively specific circumstance of time, location, and space. Knowledge-driven HAR approaches usually utilize the domain knowledge and heuristics in activity modeling and recognition.

There are several kinds of knowledge-driven HAR approaches that have been proposed. The first is mining-based approach. Perkowitz et al. proposed their idea to do activity modeling from the websites[53]. They generated a total 21300 activity models which is far more than what we need in smart home. Then Wyatt et al. continued Perkowitz's research and created DBN activity models[54]. Mining-based approaches face two main challenges. It is difficult to deal with synonymous problem which means that the synonymous words may lead misunderstands. Another shortcoming is they can not handle different situation and different persons. Mining-based approaches make use of public data sources, so that they can avoid the "cold start" problem. However, they are weak in dealing with idiosyncrasies of activities, since they assume that everyone follows the same way when they perform one activity. While data-driven approaches have the ability to generate personalized activity models. But the model generated by them has no reusability for different users.

Another way of knowledge-driven HAR approaches is to use logic. There are a number of logical modeling approaches and reasoning algorithms in this area. Generally speaking, they proposed their own activity description logic to model the activity according to the common sense of the activity, including the subset of the activity, the sequence of the actions that form the activity, the locations, etc. For example, Kautz et al. and their first-order axioms built a library of hierarchical plans of activities[55]. They treat the activity as a set of actions and set the logic between them to recognize the activity. Other research make use of Event Calculus (EC) to recognize the activity. The EC formalizes a domain using fluents, events, and predicates. In a short, they define the situation of the activity, and they take the temporal information into consideration. It means that they also define the relation between activities, so that they can somehow predict the activity. The work proposed by Chen and Nugent showed us a EC-based framework to recognize the activity according to the sensor activations[56].

Also, ontology-based approaches has gained growing interest. They can work with both vision data and dense sensor data, and achieve good performance, respectively. For example, Chen et al. proposed and developed an ontology-based approach to activity recognition[57]. They constructed context and activity ontologies for explicit domain modeling. Sensor activations over a period of time are mapped to individual contextual information and, then, fused to build a context at any specific time point. They made use of subsumption reasoning to classify the constructed context based on the activity ontologies, thus inferring the ongoing activity. Compared with logic-based approaches, ontology-based approaches have the same mechanisms for activity modeling and recognition. However, ontology-based approaches are supported by a solid technological infrastructure that has been developed in the semantic web and ontology-based knowledge engineering communities.

2.3 From Action to Activity

As we introduced above, there are a great deal of approaches that try to recognize the human activity according to different kinds of data. Most of them agree with that one activity consists of several actions, which is accord with activity theory. Actually, one activity may consists of several lower level activities, too. To simplify this issue, we just take the situation in Figure 2-2 as an example.

In Figure 2-2, we can see that the activity is defined by several sub-activities, which is also called actions. And the action is defined by the semantic information. For example, one action may imply “4W”, which are “who”, “where”, “when”, “what”. “4W” describe an objective action and determine the specific one action. And the actions contain very basic semantic information, while the higher level activities contain much more complex semantic information. It means that the lower actions

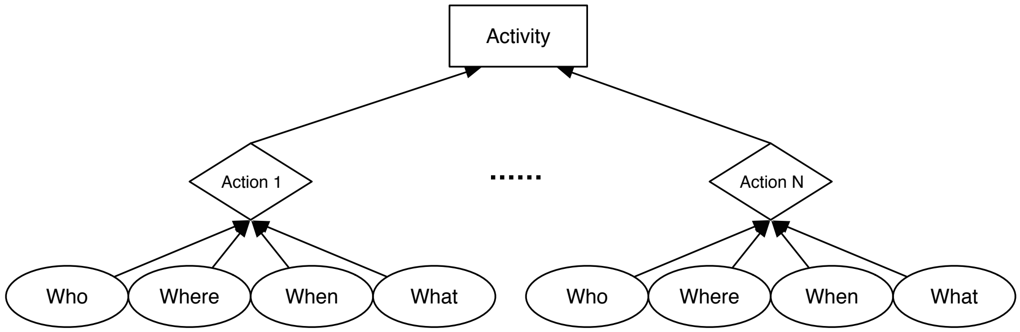


Figure 2-2: Modification of activity theory.

are more simple and easy to be recognized. Because, the lower actions contain less human personality. For example, even though people may prepare different food for their breakfast, they always turn on the gas with similar gesture. If we try to recognize that the user is making breakfast, the task may be difficult. But, if we try to sense the gas is being used, it will be much easier.

Thus, the HAR problem can be unwrapped into two problems. The first is how to sense the lower level actions, and the other is how to combine the actions to infer the higher level activity. For the first problem, the best way to enable the computer system to sense the actions is to utilize wireless sensing technology. Because, the existing work have shown us that approaches with vision data may lead essential privacy problem, especially in home environment. And the human may perform activities with so many objects around them. The objects are different from size and location. So wireless sensing technology can make sure that the system can cover as many objects as possible. One the other hand, we have to consider the relation between activity and its related actions. Since people may perform one activity in different ways, it is not appropriate to utilize fixed pattern to different users. The best way we believe is to apply machine learning or other data-driven algorithms to built personalized activity models to different users.

Table 2.1: Contrast of state of the art wireless sensing technologies for action sensing.

Characteristics	Wi-Fi	UWB	RFID
Low-cost	✔	✔	✔
Perceptive	✖	✔	✔
Multi-function	✔	✖	✔
Non-invasive	✔	✔	✔
Easy to implement	✖	✖	✔

2.4 Interaction-based HAR

After the analysis of activity and action, we propose our interaction-based HAR approach. As we introduced above, when a user perform one activity, he always interact with some objects. Our idea is to monitor the usage of these objects to further infer the action ongoing, and then combine the performed actions to infer the high level activities. The proposed approach is scalable in two spaces. The first is that the proposed approach can recognize all grain size activities from high level to low level. The other is that the proposed approach can be applied in different indoor scenarios. In this dissertation, we take home and library as examples.

We have introduced that the best way to sense the interaction between user and objects is to utilize wireless sensing technology, such as: Wi-Fi, Ultra-Wideband (UWB), RFID. With these technologies, researchers have proposed several work in action recognition. For example, Wisee [58], WiVi [59], and Witrack [60] recognized actions with WiFi signal. They succeed in distinguishing various kinds of gestures at the cost of sophisticated system implementation, and/or specialized and modified hardware. However, these systems work in short range, and usually require specific hardware environment. While, passive RFID is an alternative promising technology to recognize human action with objects.

We summarized the state of art wireless sensing technologies in action sensing, as shown in Table 2.1. According to Table 2.1, we will say passive RFID is more appropriate in sensing human action in indoor environment. Note that, we do not require user to carry any device in this dissertation, which is a great improvement than the existing work. We find that when user perform some basic action, the phase of RFID signal will change accordingly. For passive tag, phase is an attribute which is collected by the reader in every query. The interactive actions between users and tags will influence phase value in different aspects. Note that, specific action will cause the corresponding phase waveform. Thus, it is reasonable to detect the action by monitoring the phase change.This inspires us to utilize the feature of phase change to infer the interaction between human and objects with passive RFID tags. Due to RFID tags can label the objects with their inherent ability, system can easily distinct which object is being operated by the user. Thus, the usage of all the objects can be generated in the same way.

As depicted in Figure 2-3, these four pictures represent four actions that affect the phase value. To show the waveform more clearly, we use unwrapped phase rather than the original phase. Figure 2-3(a) represents that a hand covers the tag completely. Figure 2-3(b) represents a finger touches the tag. Figure 2-3(c) represents a finger sweep on the surface of the tag from one side to the other. Figure 2-3(d) represents a hand waves above the tag. The reason that different action causes different waveform will be discussed in later chapters of this dissertation.

After we generate the usage information of all the objects, we then utilize machine learning algorithm to model the activity with the weight of low level actions. For some simple activities which only include one object, it is quite easy to recognize the activity just based on the usage of the relative object. Unfortunately, most activities include far more than one object. One activity may include several actions interacted with several objects. Also, one action may belong to different activities, respectively. This brings great trouble when we try to determine what the user actually is performing. Because, only knowing the usage of current object is not enough to infer the high level activity.

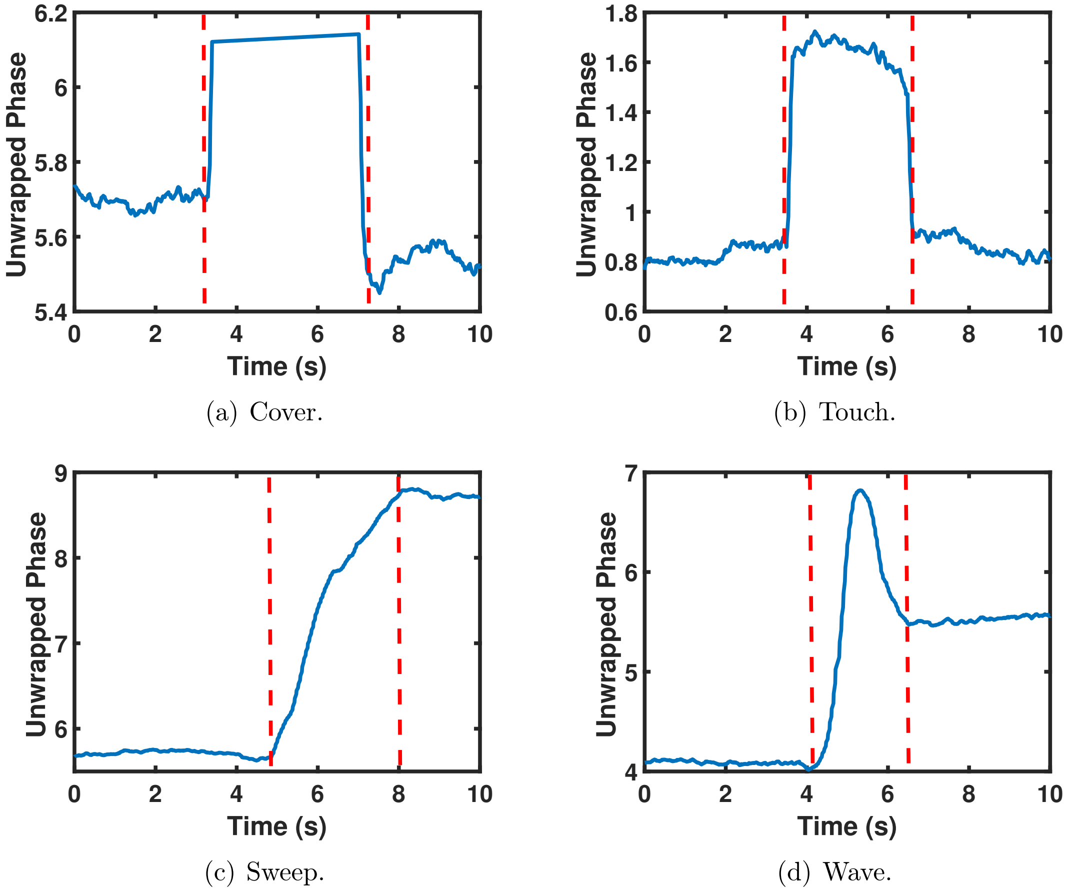


Figure 2-3: Examples of phase change and the corresponding actions.

Interaction-based HAR in smart home

In smart home scenario, there are different grains of activities. For simple activities, we can directly determine whether these activities are being performed or not by monitoring the state of RFID tag. We can use this find to replace existing way to control the home electrical appliance. So we introduce a new RF-Swith to replace the traditional switch on the wall. Because, simple interactions such as touch, sweep, are enough to be used to control the switch. And such simple interactions can be easily detected by passive RFID tags as introduced above. The detail of RF-Switch will be introduced in later chapter.

While for the complex activities, we introduce tf-idf weight to map the actions to activities. Even though one action may belong to different activities, we can calculate the possibility that one action belong to one specific activity. To generate the tf-idf weight, we first propose a off-line solution to record the activity log and the object usage log. With these two kinds of log, we calculate the tf-idf according to the frequency that one action occurs in one specific activity. The frequency actually suggests that the current action can be seen as a part of which activity. Thus, according to tf-idf weight, the proposed approach can realize on-line activity recognition.

Moreover, we further explore the habit of human activity. We believe that the human activity habit tends to follow a personal pattern. And this pattern differs from each other. Then, we make an assumption that the current activity that the user is performing has relation with the former several activities. This assumption according to the common sense that everyone has his own fixed schedule, especially for some specific occupations, such as students, white collar employees. They have their own pattern to perform different activities in specific order. What is more, the activities sequence can be seen as a time sequence. So, we can introduce LSTM to model this sequence. The reason why we utilize LSTM is that it can utilize both long and short information to determine the next activity, which just accord with our assumption. With the help of LSTM, our approach can predict the next possible activity that the user might perform. This is one of the contribution in the dissertation that has great improvement to smart home.

Interaction-based HAR in smart library

In the library situation, the proposed interaction-based HAR works in another form. In home environment, user interact with the object so that we can infer the action based on the usage. We only care about if the object is being used. While in the library, readers interact with the tagged books in different way. And this enables us to infer different activities according the way that the user interact with the book. In other words, we do not only care about which book is interacted by the reader, but also care about how it is interacted. Different ways of interactions suggest different different states of the readers. So that we can recognize the reading activities in the library with our proposed interaction-based approach.

What is more, we also explore more in smart library to try to solve the problem of book misplacement. Even the reader does not pick up any books, he will influence the RF single of the tags on the books. This enable us to make a reverse recognition of book state according to this tiny influence. If the book is right at where it should be, the book state should be the same with its neighbor book. If not, the book then can be most likely detected as a misplaced book. The whole task does not require any aware of reader, while the misplaced book can be detected when the reader walks before the bookshelf.

2.5 Summary

In this chapter, we introduced the background of different approaches in the area of HAR in detail. If we divide the existing work according to the type of sensors, there are two kinds of approaches of HAR: vision-based HAR and sensor-based HAR. While if we divide the existing work based on the type of activity model, there are also two kinds of HAR: data-driven HAR and knowledge-driven HAR. We introduced their characteristics and weakness, which inspire us to propose our interaction-based HAR approach.

Also, we briefly introduced the motivation of our proposed approach. We utilize passive RFID technology as a tool to sense the interaction between user and objects. Different interactions may cause different influence to the phase of RFID signal, which is a physical attribute of the RFID signal. Passive RFID technology enables us to monitor the actions without any requirement to the user, that can be seen as a device free approach to human. Our proposed approach can not only recognize the human activity in real time, but also predict the next possible activity in the home environment, which makes it a great improvement in home activity recognition.

For simple activity recognition, we proposed RF-Switch to work as a flexible home controller that greatly reduce the inconvenient of traditional fixed switch on the wall.

This part is introduced in detail in Chapter 3. And for complex activity recognition, we proposed RF-ARP to translate RF signal of tags on the daily objects to human activity, which will be introduced in detail in Chapter 4.

Besides, we also apply the idea to library to recognize the reading activities. This provides more knowledge to book management systems to enable several promising applications, such as personalized book recommendation and smart book replacement. This work is introduced in detail in Chapter 5.

Chapter 3

Low-level Activity Recognition

3.1 Introduction

In smart home environment, all of the devices are in the charge of unified platform. Through this, smart home can achieve the goal of home automation. The platform and devices are connected through different kinds of wired and wireless technologies, that the platform can easily get the state of the devices and control them at the same time. However, the interface between platform and human is far more complex than it is between platform and devices. Some research have been done in this area. Basically, they try to combine the smart home platform with human mind through audio[61], video[62] and action sensing[63]. Both audio-based and video-based interactions may bring privacy problems to users, especially in home environment. Action sensing methods usually require users to carry some additional equipments around or to act in specific region, which make it hard to implement in the real life. Thus, we turn to look for a feasible way to replace the current widely used controller: switch.

When it comes to traditional switch, it is not difficult to cite some examples of its advantages, such as stable, convenient and easy to carry out. While under the concept of smart home, the traditional switch seems not that appropriate in the future. The position of switch is fixed once it is installed which may cause troubles to elderly and children. Even if the inhabitant are not used to the place of the switch, the switch can hardly be moved. And human may get electric shock when touch the switch with a wet finger, not to mention the influence to the beauty of home decoration. Therefore, a exquisite, flexible and scalable switch is needed in future smart home.

In this paper, we design a novel switch for smart home using passive Radio Frequency IDentification(RFID) tags. Passive RFID tags is completely battery-free, which make it possible to be quite thin and small, and can even be customized into arbitrary shape[64]. Thence, the tags can be attached everywhere without any brackets and can be moved easily. Besides, the tag ID is long enough to be encoded for every device in one's home. Futhermore, the Ultra High Frequency(UHF) RFID technology works on the frequency between 858MHz and 930MHz, that will not effect the other commonly used wireless communication in home environment like Wi-Fi. All the above characteristics make RF-Switch a qualified substitute of current switch and the design of RF-Switch will be introduced in the next section.

3.2 RF-Switch: A Novel Wireless Controller in Smart Home

3.2.1 Home Control Systems

To support more intelligent and adaptive home services, it has the potential to enable user to control consumer home devices used in everyday life in different channels. Now, more and more home control systems tend to build intelligent agents to control the home devices according to the inhabitants' command in different ways. For example, we can already talk with "Siri" to turn on the light in living room. And we can ask "Alexa" to turn down the power of the air conditioner whenever we want. Home control systems are always there to listen to the user command, and control the home device instead of the user.

Note that, these kind of intelligent home control systems always require the user to interact with them by audio. This indeed can reduce some effort to the user when they would like to talk with a machine. However, there is amount of people who do not like to talk with a machine agent in their daily life. Besides, in some specific scenes, it is not appropriate to talk loudly at home. For example, we do not want to disturb our parter when we get up at night.

Also, some work enable user to control the home devices remotely with their smart terminals, such as smart phones, smart watches, and laptops. However, we usually do not carry these smart terminals at home all the time. In this sense, we start to think about the revolution of traditional home controller on the wall. We build a home control agent that works as flexible controller that do not require the user to change their traditional habit to control the home devices.

3.2.2 Architecture of Wireless Controller

In this section, we present the architecture and design of RF-Switch in detail. The verification experiment is implemented using the ImpinJ R420 reader and Alien ALN-9654 tag.

As depicted in Fig.3-1, RF-Switch works with the help of smart home platform. RF-Switch is in charge of sensing the intention of inhabitant and sending the commands to smart home platform, then the platform is responsible for the control of devices in the smart home.

In this paper, we consider that all of the control commands consist of two basic commands: on-off control and volume control. For example, the ordinary light only has two state, thus it only needs on-off control. While the modern light can change the brightness, then it needs volume control. More complicated commands may need several RF-Switches work in collaborative way and this will be discussed in the future work. Next, we present our design of both on-off control and volume control separately.

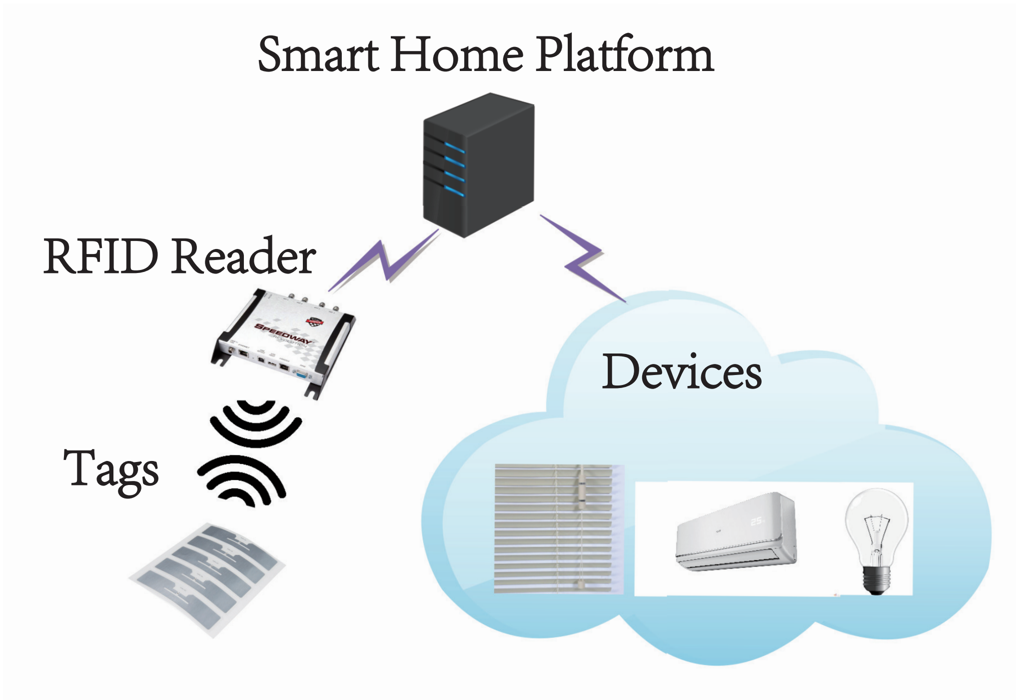


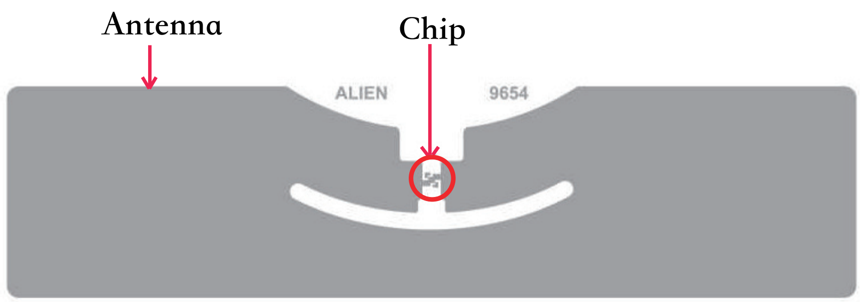
Figure 3-1: The architecture of RF-Switch. Note that RF-Switch can be accessed into different kinds of smart home platform.

3.2.3 On-Off Control

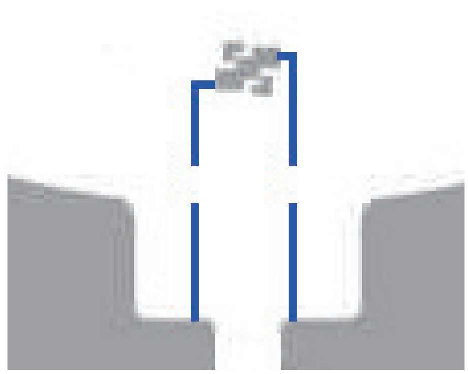
In this part, we introduce how RF-Switch can realize on-off control. Actually, it is because of the structure electric circuit that passive RFID tag does not need battery.

As shown in Fig.3-2(a), passive RFID tag is composed of RFID chip and tag antenna. The chip is in the red circle and the other part is antenna. Firstly, tag antenna receives the wireless signal from the antenna of reader, and the chip can get the most power from the tag antenna when the chip impedance and the reader antenna impedance are conjugately matched[65]. It means if the chip can not get power from the tag antenna, the tag will not response the reader's request even if it is covered by the reader's effective range.

Based on this, a simple but effective and easy to carry out idea comes to our mind. We remove the chip a little bit from its original position. Then add two leading wires to both the chip and antenna, as shown by the blue lines in Fig.3-2(b). We can see from the picture that the wires are close to each other but not connected



(a) The structure of original passive RFID tag.



(b) The redesigned tag structure.

Figure 3-2: The structure of tag.

with each other. This means in this state the tag can not be seen by the reader, thus this state can represent “off”. When a finger touch the middle of the tag, the skin will connect the four wires together. Since the skin of human can be treated as conductor, the chip will receives energy from the antenna and send the response back to the reader. In this situation, the tag can be seen by the reader, thus this state can represent “on”. In this way, RF-Switch can works as a on-off controller.

3.2.4 Volume Control

In this part, we introduce how RF-Switch can realize volume control. To control the volume, RF-Switch should have the ability to send continuous mutative commands whose value can cover a range from the minimum to the maximum.

$$\theta = mod(2\pi\frac{2d}{\lambda} + \theta_T + \theta_R + \theta_{TAG}, 2\pi) \tag{3.1}$$

In a basic RFID system, the reader transmits continuous-wave signals to the tags, and then receive backscattered signals from the tags. The phase value of RF signals calculated by Eq.5.1 describes the offset between the transmitted and received signals, which depends on the round-trip(2d) and hardware-specific factors[66]. Usually, the switch is fixed while working. Therefore, the round-trip distance is fixed. Phase can be seen as a continuous feature of tag.

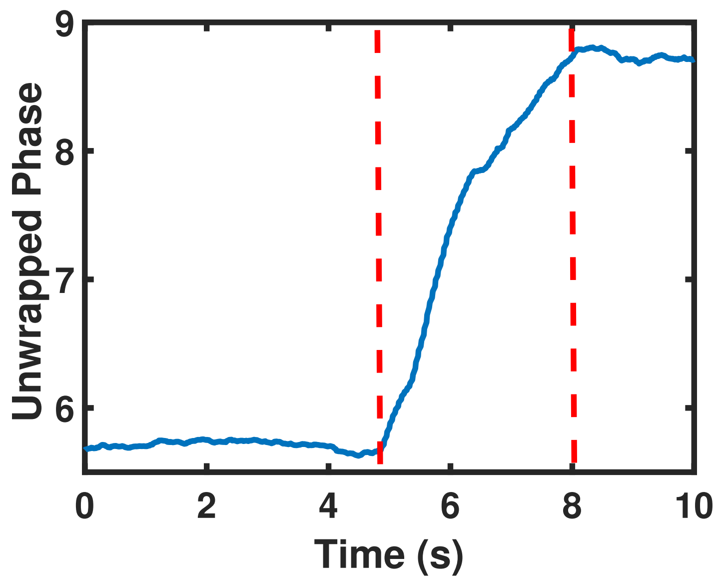


Figure 3-3: Sweep a finger from left to right on the surface of a tag.

We assume the human body can also absorb RF signal as the antenna does. When the finger touches different parts of the tag, θ_{TAG} will change correspondingly. The verification experiment proves our assumption. As shown in Fig.3-3, we collect phase data while sweeping a finger on the surface of the tag's antenna in 5s-8s. The blue line is the unwrapped phase value. It changes from less than 6 to nearly 9. And if the finger stops moving, the phase value does not change any more. This phenomenon is explained in the latest research [67]. The finger does change the impedance of the tag and futhermore changes the θ_{TAG} and θ finally. Thus, by monitoring the phase value of a tag, RF-Switch can control the volume.

3.3 Experiments and Results

Table 3.1: The performance of RF-Switch on different materials.

Materials	Wood	Plastic	Glass
On-off	50	50	43
Volume up	50	49	38
Volume down	50	50	34

To test the usability of RF-Switch, we implement it with the off-the-shelf devices at office scenario. We put the RFID antenna on the ceiling of the room to continuously monitor the phase of the passive RFID tag, which acts as a movable controller. And we ask a volunteer to interact with the tag, just as the same when he does with traditional controller on the wall. Then, we take a record of the actions that our RF-Switch have recognized and compare with the ground truth actions log. Also, to verify that the RF-Switch can work on wall built by different material. We chose three kinds of common material to represent the wall: wood, plastic, and glass. We ask the volunteer to repeat the action 50 times, respectively. If the user action is correctly recognized by RF-Switch, we count a right recognition. The overall performance is shown in Table 3.1.

From the Table 3.1, we can see that the proposed RF-Switch can recognize the on-off control and volume control with a relative good performance when it is put on the wood and plastic. However, when it is attached on the glass, the accuracy of recognition drops greatly. According to our analysis, we ascribe this to the weakness of the RFID tag we use. There are several kinds of tags produced by Alien. The tag we use in this work can suffer from the signal absorb with wood and plastic, however, it is not suitable with glass. If we chose another kind of tag that match with glass, we believe the performance will be as good as wood and plastic.

3.4 Summary

This paper presents a novel RF-Switch to replace the currently used switch. RF-Switch utilizes the characteristics of passive RFID tag to realize both on-off control and volume control for smart home platform. In the future, we will implement RF-Switch in our test house and attach it to different smart home platforms.

Chapter 4

High-level Activity Recognition and Prediction

4.1 Introduction

Over the last few years, the Internet of Things (IoT) has been greatly developed with the help of mobile computing, edge computing, and cloud computing. One of the most representative applications of IoT is smart home, which has good prospects in the future. In a typical smart home system, all the resources and devices can be controlled by the smart home platform. The long-term goal of smart home is to achieve automatic control according to both environment and inhabitant. Owing to the advances in sensing technology, it is not difficult to obtain the environment data such as illumination, temperature, and humidity. However, still no practical solution exists to recognize human activity at home to enable scenario-based smart services [68]. To enable the smart home platforms to know more about their host, human activity recognition (HAR) has become an urgent challenge for the researchers.

Unlike the online consumer activity, the activity of daily living (ADL) usually cannot produce any data for a computer system, thereby causing a gap between inhabitant and smart home system. To bridge this gap, the existing work has shown us a bright direction. The interaction between human and devices could be the channel to recognize the human activity. Wearable devices have already been widely used in recent years. Such devices equipped with different kinds of sensors and microprocessors are worn on the human body to monitor the state of humans [69]. Represented by Apple Watch, smart watches and smart wristbands can recognize simple activities such as waving hands and sitting still [63, 67, 70]. However, such activities do not contain any semantic meaning, and their recognition can be more appropriately called either gesture recognition or action recognition [71]. In this study, we define these kinds of activities as low-level activity. These activities cannot be used directly by the smart home system to provide scenario-based services. Another way to detect the interaction between human and devices is to attach sensors to objects used extensively by humans [38]. However, smart electronic sensors rely on batteries, thus the size of these smart sensors is not sufficiently small to be attached to all the devices, not to mention their price along with their maintenance cost. It is worth noting that passive radio frequency identification (RFID) tags seem to have the ability to replace such sensors. In addition, several works have been proposed to prove that passive RFID tags offer a good way to detect object usage [72, 73]. Based on these research works, our work achieves even further results of object–usage-based activity recognition.

In this study, we first deeply analyze the characteristics of human activity in home environments. This further clarified the goal of HAR in smart homes; the goal is to provide scenario knowledge to the smart home platform to achieve human-centered automatic service. Then, we propose an RF-ARP framework recognize and predict the human activity in smart homes, as depicted in Figure 4-1. Different from address-resolution protocol (ARP), which translates an IP address to MAC address, our RF-ARP translates a wireless radio frequency (RF) signal to the human activity. The proposed RF-ARP framework mainly contains three stages: recognition after the human activity, recognition in progress, and activity prediction in advance. In the first stage, we utilize passive RFID tags to detect the interaction between human and device and recognize a high-level activity by combining those low-level activities. In this stage, we can record the activities of the inhabitants. Furthermore, in the second stage, we weight the device based on term frequency–inverse document frequency (tf-idf) to ensure the significance of each device to each activity. In this way, we will not have to give the recognition result after the completion of the activity—while, in the third stage, we already have the activity log. Thus, we could use the long short-term memory (LSTM) network to model the ADL of the inhabitant. In this manner, the proposed framework will be able to predict the next activity that, perhaps, happens after the current activity. We finally test our framework using off-the-shelf equipment and an open-source database, followed by proving the effectiveness and efficiency of RF-ARP.

Building such a system involves several challenges. The first one is to achieve object detection without using wearable devices. The existing work usually uses a wearable RFID reader to detect the object usage according to the distance between the object and the hand of the human [72]—while, in our work, we use a fixed long-distance antenna to cover as much region as possible, leading to the invalidation of the previously reported method. Therefore, we propose a way to detect object usage by phase, which is a physical feature of the RFID signal. The second challenge is the recognition of concurrent activities. We introduce a task-oriented generative approach rather than a discriminative approach; therefore, the recognized result could be more than one activity. In addition, traditional machine learning methods rely on training data, causing the so-called "cold start" problem. However, our approach utilizes the prior knowledge to define an activity, thus the training data are not required in the first stage. Thus, the upper stages could be in motion after sufficient data have been produced in the first stage.

Compared with the existing work on activity recognition in the smart home, our framework has multiple advantages regarding different aspects. First, scalability is the most important strength of our approach, which is reflected in two aspects. The first aspect is that we allow the smart home platform to define all kinds of activities as it needs. This is great progress, as the activity recognized by the existing work may be not what the smart home platform requires. In addition, the sort of activities could not be changed in the previously conducted work on HAR. The other aspect is that our approach could work in different houses, even when the objects and devices are different from each other. However, the existing work cannot migrate the trained model to adjust in a new environment, limiting the model to only work in the laboratory. In addition, our activity prediction stage is going further than the current HAR in smart homes. This may largely promote fully automatic smart homes in the near future. The next strength is that our three-stage framework unwraps the task to recognize the high-level activity from wireless signal data. This brings huge flexibility because every stage can be optimized independently or even replaced by other algorithms. For example, we can substitute the LSTM model in the third stage with any other time-series data-mining algorithm because the labeled data provided by the first stage can be used to train different models. The last strength of our framework is that both the cost of RFID tags and the computational complexity are sufficiently low to implement our framework in the current houses without much effort.

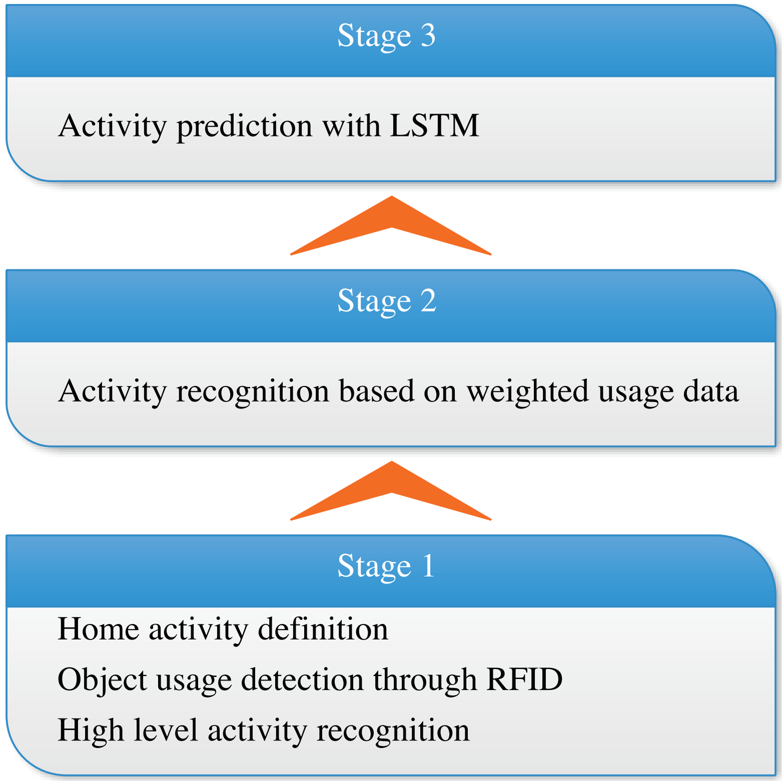


Figure 4-1: Three-stage framework to recognize and predict human activity in a smart home.

4.2 Off-line Activity Recognition

4.2.1 Definition of High-Level Activity

To ensure that the result of HAR is required by a smart home platform, the best solution is to grant the smart home platform the authority to define what it needs to know. Thus, we make a rule to enable such authority. The definition of an activity includes only the objects used in that activity. For example, television has a semantic meaning that greatly indicates the activity of "watching TV".

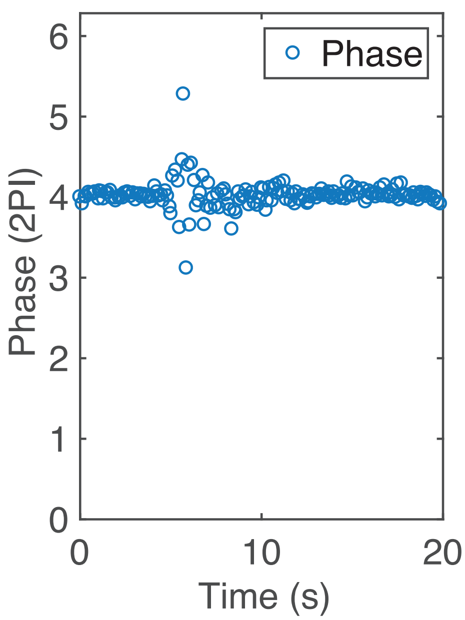
4.2.2 Object Usage Detection

In this section, we present a way to detect the object usage by using passive RFID tags and transform the physical-signal data to binary-state data.

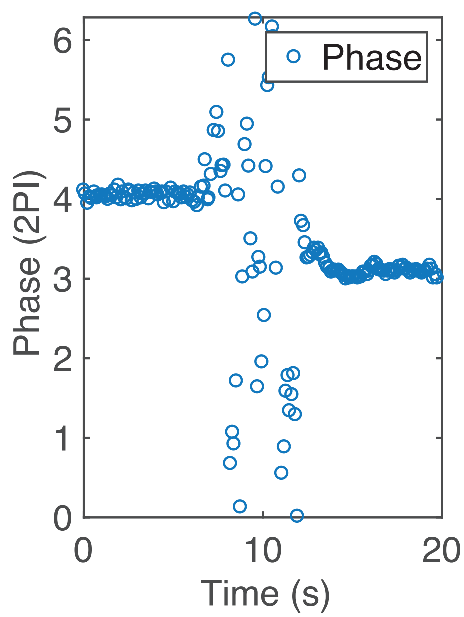
Passive RFID tags are small and sufficiently cheap to be widely used in applications such as logistics and warehousing [74, 75]. Moreover, researchers have found that RFID tags can be used to detect the object usage by attaching them on objects used every day [76]. Fazzzinga et al. even utilized RFID tags to trace the objects both online and offline [77] [78] [79]. This may considerably help in object–usage analysis. Several years before, Philipose et al. used a glove equipped with a near-field RFID reader to do such work by the state of readable and not of the tags [80]—while, in our work, we selected a long-distance antenna and ultra high frequency (UHF) reader to perform the same task. The reason is that a fixed long-distance antenna can cover a large area and scan all the tags in the area almost simultaneously. Furthermore, some objects are not used by hand, such as chairs and beds. In addition, we do not require the inhabitant to wear any devices, thereby reducing the inconvenience to the inhabitant.

Several systems have been proposed for achieving object–usage detection via a UHF RFID reader [81, 82]. In these systems, the received-signal-strength indicator was merely used to distinguish the state of the tags; note that state is not a stable parameter. Our previous research explained that the RF phase of RFID tags can better reflect both the spatial attribute and the interaction between human and tagged objects [83].

Phase is a back-scattered RF-channel parameter, which can be continuously read by a UHF RFID reader. In our work, we use Impinj R420 (Seattle, USA) as the reader to obtain the phase value. According to our previous research, phase is sensitive to the interaction between human and RFID tags. As depicted in Figure 5.1, different interactions change the phase value in varying degrees. Figure 5.1a depicts the condition in which there is a human passing by the object with an RFID tag—while, in Figure 5.1b, the human walks to the object and picks it up, then puts it back to its position and walks away. This inspires us to use the dispersion of the phase data in a sliding window to distinguish among the interactions. We have then proved that only if the object is picked up will the tag state be revised as “1”, irrespective of how close the people are standing or passing by the object. Because phase is more sensitive to distance than interferences or reflections, even when several people are close to one another, their interactions with the objects could be detected correctly. The detailed analysis of this phenomenon is proposed in our previous work [83].



(a)



(b)

Figure 4-2: Interactions and the corresponding phase changes. (**a**) Passing by; (**b**) Picking up.

Apart from the above-mentioned interactions, people also interact with heavy furniture such as a bed and sofa without moving them. In spite of this, in these cases, the object usage can be detected by the phase. UHF RFID readers are able to scan the tags several times in a second. In our verification experiment, the average sampling rate for each tag is 12 times per second. Although the reader can keep receiving the back-scattered signal when the tag is interfered, it cannot see the tag while the tag is completely blocked. This enables us to use the following simple way to detect such interactions:

$$covered = if((t_0 - t_1) > T), \tag{4.1}$$

where t_0 represents the current timestamp, t_1 represents the timestamp of previous round scanning, and T is the threshold of the tag state. In this study, we set T as 1 s to ensure enough sufficient sensitivity to detect short-term interactions. For some specific objects such as bed, we can increase T to detect the right interaction. Note that, when the tag works as a switch, one more step is required to translate the interaction to the equip state. The relevant details were introduced in our previous work [84].

Thus far, we have introduced the way to detect object usage, and Table 4.1 presents the way to determine the object–usage state. When the tagged object is covered or picked up, it means the object is being used, and thus the usage state is set to "1". Otherwise, when the tagged object is interfered or still, it means the object is not being used; thus, the usage state is set to "0". As depicted in Figure 4-3, the matrix with a white background color is an example of object–usage array.

Table 4.1: Object-usage detection via interaction.

Usage	**Tag State**	**Interaction**	**Objects**
1	Covered	Sitting, lying, blocking	Chair, bed, sofa, switch, etc.
	Picked up	Picking up	Knife, toothbrush, chopsticks, etc.
0	Interfered	Passing by	All
	Still	Absence	All

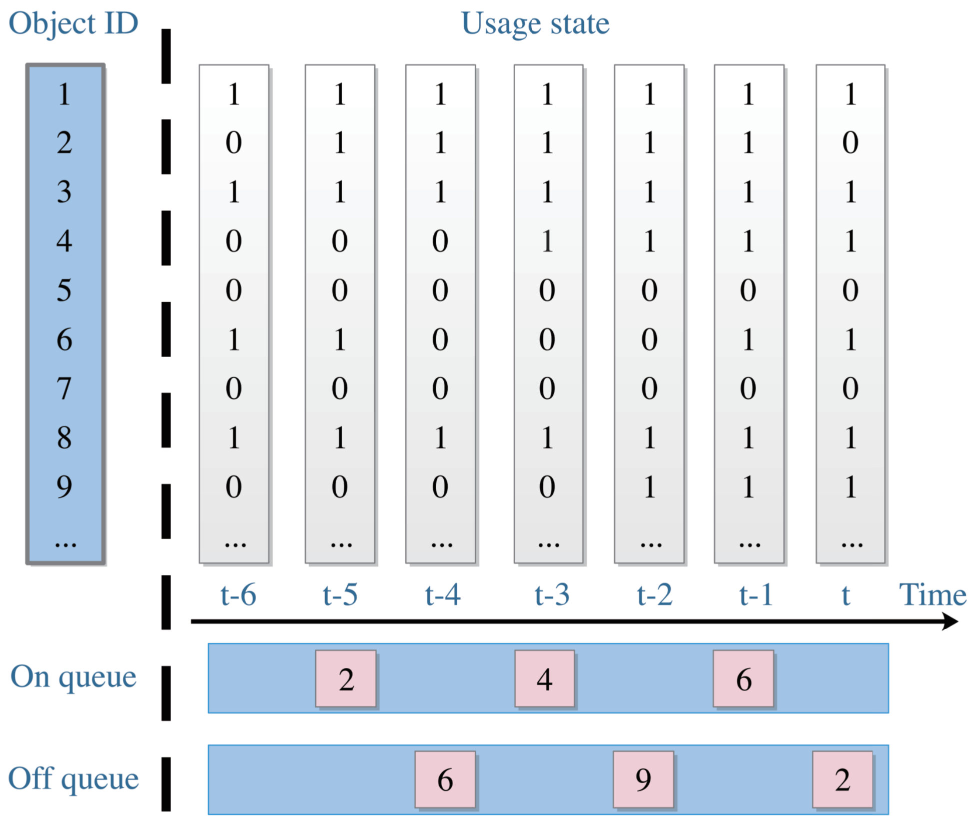


Figure 4-3: Object-usage state vector changes with time, and we can generate "On queue" and "Off queue", respectively.

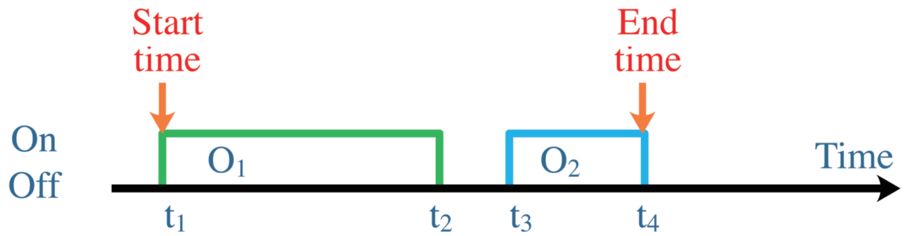
4.2.3 Activity Fusion and Segmentation

As depicted in Figure 4-3, after detecting the usage states of the objects, we built two queues to store the detected usage states. The "On queue" contains the object ID and the timestamp that the object starts using—while the "Off queue" contains the object ID and the timestamp that the object stops using. At this stage, the length of the queues is fixed as one day. It means that the task of this stage is to make a record of activities performed during the last one day.

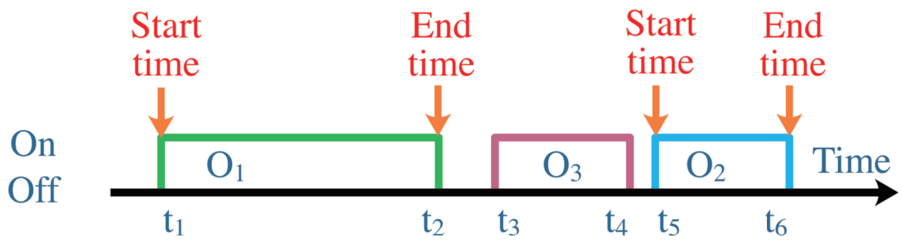
We now consider "watching TV" as an example, where O_2 represents the television. At $t-5$, the television is turned on, and it is turned off at t. Therefore, we can say that the activity of "watching TV" has been performed; in addition, the start time is $t-5$ and the end time is t. However, this is a merely single-object activity.

In general, a high-level activity tends to include more than one object. Therefore, we set the start time as when the first object is used, and the end time when the last object stops being used. However, we cannot obtain the correct end time because the activity may be interrupted by other activities. In this case, we are confused regarding the time at which the activity ends.

To overcome the aforementioned problem, we treat the object in the definition individually as a subset of the set that contains all the objects. In addition, two strategies are proposed to determine the start and end times, as depicted in Figure 4-4. In Figure 4-4a, there is no "On" action between t_1 and t_3, meaning that the activity is ongoing continuously, although the usage of O_1 ends before the usage of O_2 starts. Therefore, we can merge the two subsets to a new subset. The start time is still the timestamp of the first object in the "On queue", and the end time is set as the timestamp of the last object in the "Off queue". When the next object belonging to the same activity starts to be used, we check the interruption in the same way as above. If there is still no interruption, then we keep merging the subset to the former subset and reset the end time. However, when there is an interruption, we use the strategy depicted in Figure 4-4b. We can see that, between t_1 and t_5, O_3 has been used at t_3. Thus, in this case, the activity is interrupted by another activity. We, therefore, cut the relationship between the current subset and the former one that belongs to the same activity. The former subset has finished at t_2, so the end time is set to t_2. Moreover, the current subset becomes the initial subset, and the start time is t_5. However, the end time is t_6 and may be rewritten by the later subset too.



(a)



(b)

Figure 4-4: We propose two strategies to determine the start time and end time. O_1 and O_2 are objects that belong to one activity, and O_3 belongs to some other activity. (**a**) No interruption between two objects; (**b**) Interruption between two objects.

The reason why we chose these two strategies is because we wanted to recognize the activity as accurately as possible and reflect the real scene. Suppose, while cooking in the kitchen, that the inhabitant hears a phone call. After receiving the phone call, he resumes cooking. The former HAR approach can find it difficult to distinguish such a short-term activity from a continuous activity—while our strategies could record the ground truth in detail. Moreover, we allow each activity to create its own object set independently, so that the processes of HAR are also independent, even if an object may be involved in several processes. In this way, the multiple concurrent activities can all be recorded simultaneously.

4.3 On-line Activity Recognition

In the last section, we recognize the activity in an offline manner—while, in this section, we introduce an online approach that recognizes the activity in progress.

In the first stage of our framework, we define an activity by using the objects that are only used in that activity. Although it can record the log of activities that have been performed efficiently, the precision of the start and end times is not so satisfactory. In addition, the definition of activity has limitations. Many objects can not be included in the definition because they might be used in more than one activity. Thus, we introduce an approach to extend the definition of an activity, and, subsequently, recognize the activity based on the new definition.

In previously conducted research works, tf-idf was commonly used in both natural language processing and information retrieval [85, 86]. It is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus [87]. Owing to the ability to reflect the importance of element in frequency, we utilize the tf-idf to weight the objects according to the definition of activities. Before calculating the tf-idf value of each object, we have to generate the training data. In the first stage, the interval between two activities is large because of the very strict activity definition. Therefore, we extend the start time of one activity to the end time of the former one and extend the end time of the activity to the start time of the last one. The objects used to exist between two activities are involved in the definitions of the activities on both the sides of the usage. In this way, the definition of an activity is extended considerably.

Without loss of generality, we represent the set of objects O as

$$O = \{o_1, o_2, ..., o_n\},$$

where n denotes the number of objects. The set of activities A is denoted as

$$A = \{a_1, a_2, ..., a_m\},$$

where m denotes the number of activities. In addition, we define the set of activities in the process as $P \subseteq A$.

In the activity log data, we count the frequency of every objects o_i $(i \in n)$ that has been used in a specific activity a_j and note it as g_i^j. Note that that our only concern is whether the object has been used in one round of the activity, ignoring the number of times it has been used. If the object is used more than once in one round of activity, we still count one for one round. Moreover, to reduce the noise caused by the activity–definition extension, we use the follow equation as a high-pass filter:

$$g_i^j = \begin{cases} 0, & g_i^j < z * max(g_i^j), \\ g_i^j, & otherwise, \end{cases} \tag{4.2}$$

where z denotes a threshold set as 0.5 to control the filter in our work. The higher we set the value of z, the stricter the definition of an activity becomes. Particularly, if we set z as 1, the activity definition will be the same as that in the first stage.

The term frequency tf_i^j can be calculated using the follow equation:

$$tf_i^j = \frac{g_i^j}{\sum_{i=1}^{n} g_i^j}. \tag{4.3}$$

Moreover, the inverse document frequency idf_i^j can be calculated using Equations (4.4) and (4.5)

$$idf_i^j = log(\frac{m}{\sum_{j=1}^{m} f_{i,j}(T)}), \tag{4.4}$$

$$f_{i,j}(T) = \begin{cases} 0, & g_i^j = 0, \\ 1, & otherwise. \end{cases} \tag{4.5}$$

After obtaining tf_i^j and idf_i^j, the tf-idf_i^j then can be calculated as follows:

$$tf - idf_i^j = tf_i^j * idf_i^j. \tag{4.6}$$

Using the training data, we can finally generate a weight matrix that illustrates the importance of each object to different activities. Then, we utilize this matrix to realize online activity recognition.

Similar to the first stage, when a new object usage is detected, the object ID i ($i \in n$) is put into the "On queue". Subsequently, we start to check the weight matrix to obtain the maximum $tf-idf_i^j$ and the corresponding j. In other words, the weight matrix tells the most possible activity, as the object is most representative to that activity. Then, we need to check the set of activities in the process P to verify whether this activity is in set P or not. If it is in set P, we do not need to change anything and, thus, we keep waiting until the next object–usage–state change. However, if the activity is not in set P, we then add the activity ID in P and note the start time of this activity with the current timestamp.

When the usage of a new object is detected to have ended, we also need to check the weight matrix with the object ID i. We retrieve the activities whose $tf-idf_i^j$ are not 0, and then we withdraw all the relevant activities from the set P and set their end times, respectively. In addition, if the next activity is a part of the activities that have just been withdrawn, we merge them together as the first stage does, and then we set their end time to empty.

Note that the log of recognized activities is essential to determine the weight matrix. Moreover, to ensure the effectiveness of the proposed approach, the log data should be as large as possible. Furthermore, when a new activity or new device is involved, the weight matrix can to be updated automatically.

4.4 Activity Prediction with LSTM Model

In the last section, we introduced the way to recognize the activity in real time. However, in this section, we will go further to predict the activity that may happen later. In the second stage, we obtain the log of activity with the weight matrix. In addition, owing to our manner of segmenting the activities, some activities may end simultaneously. Note that, for a real-world situation, several activities may be going on at the same time. However, we believe that those activities will not start together. Therefore, we utilize the start times of the recognized activities as a sequence to represent the order of activities.

In this research, we treat human activity prediction as a time sequence prediction problem. We believe that the inhabitants perform different activities in a relatively fixed pattern. For example, according to the activity log, there is an inhabitant who always watches TV after having dinner. Therefore, if the inhabitant is detected to be having dinner currently, then his/her next activity is most probably watching TV. Such a problem with predicting next state based on the current state can be solved using the classical machine learning approach. Nevertheless, the next activity is related with not only the current activity but also the previous ones. Therefore, we introduce deep learning to solve this problem. The recurrent neural network (RNN) performs well on spatial temporal predicting problems, such as location prediction [88]. LSTM networks are a special kind of RNN, and they are proved to be more efficient than RNN [89]. LSTM networks can memorize both long and short-term knowledge, thereby tallying with the human mind.

As depicted in Figure 4-5, these are spread LSTM networks. $X_0 - X_t$ in this study represent the activity log, and $h_0 - h_t$ represents the prediction result, which is the next activity. We can see that, when the time stamp is t, the input of the model is the current activity X_t and the past knowledge remembered from $t-1$ to $t-n$. It means the model can predict the next by using not only the current activity but also several past activities. This just accords with our assumption that an activity does not happen randomly and that the motivation of the next activity is what the human has done currently. In this way, the prediction accuracy of LSTM is higher than that of the classical machine learning approach; as in the case of the former, more knowledge is considered to model the activity habits of inhabitants.

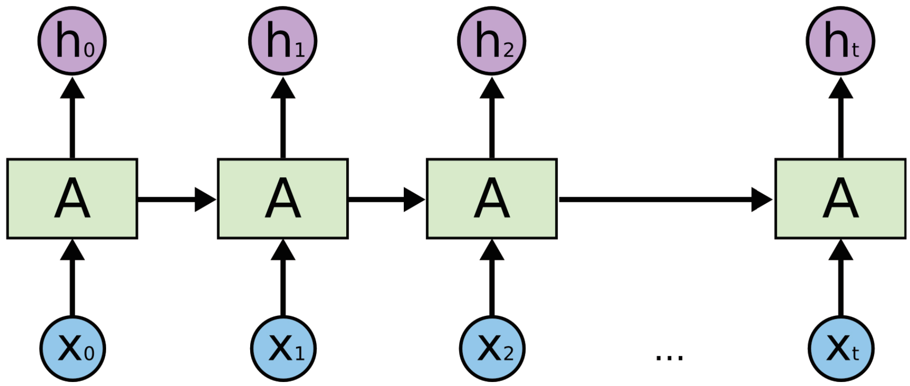


Figure 4-5: Activity sequence and recurrent neural network (RNN) model.

In addition, we also apply the method in the second stage to the process of prediction. Besides modeling the activity habit, we also utilize LSTM to model the object–usage habit. Therefore, the next object that might be used by the inhabitant can also be predicted using LSTM. Then, we find out the relevant activities associated with the object. Finally, we find the intersection of two prediction results to further improve the prediction performance. By knowing the current activity and the current objects in use, we can predict the next activity with a relatively high accuracy.

4.5 Experiments and Results

In this section, we show the performance of the proposed three-stage framework in recognizing and predicting the activity of the inhabitant. As we have unwrapped the HAR task into three stages, we test performance of all the three stages, respectively. To evaluate the activity recognition performance, we conducted an experiment on an open source dataset. The dataset generated by Ordonez [90] includes 10 ADLs performed by the inhabitant on a daily basis in their own house for 21 days. We selected this dataset because most sensors in this dataset can be replaced by RFID tags to represent the usage in a similar manner.

4.5.1 First Stage

In the first stage, the task is to record the activities that have been performed. There are two key factors that considerably govern the effectiveness of this stage. The first factor is whether the object usage can be detected correctly, and the second one is whether the activity can be labeled properly.

We attach two RFID tags to two commonly used objects: chair and toothbrush. Subsequently, we ask the volunteer to perform a specified interaction with the objects 50 times, respectively, and note the tag states and the corresponding object–usage states. Because both "interfered" and "still" represent no object usage, we treat them as one interaction. In Table 4.2, we set the following:

- TP represents that the usage is correctly detected;
- TN represents that the interference is correctly detected;
- FP represents that the interference is detected as a usage by mistake;
- FN represents that the usage is detected as an interference by mistake.

Table 4.2: Result of object–usage detection.

Objects	**TP**	**TN**	**FP**	**FN**
Chair	50	49	1	0
Toothbrush	49	47	3	1

Subsequently, the average object usage detection accuracy can be calculated as 97.5%. Moreover, the precision and recall are 96.1% and 99%, respectively. The usage–detection performance is sufficiently good to prove that RFID tags can be used to detect the object usage.

In the experiment, we first find out the objects that are the exclusive representatives for each activity. We then label the data with the activity ID by using those special objects. If the labeled time interval between the start time and the end time has an intersection with the ground truth label, we see it as an appropriate match. The activities that have representative objects can be recognized properly as long as the object usage is detected correctly. However, we find that very few activities can be recognized in this manner, as most of the activities do not have exclusive representative objects. Therefore, we fail to evaluate this part in the first stage, so we have to move to the second stage of our framework to recognize all kinds of activities.

4.5.2 Second Stage

In the second stage, we also utilize the Ordonez dataset to verify our method. The dataset is used for training and testing, respectively.

In the training part, we first classify the object–usage data according to their corresponding activity ID, and each activity contains several representative objects. Subsequently, we calculate the weight of objects in an activity to generate the weight matrix. In the testing part, we use the proposed approach to generate the activity log. After that, we compare the activity log with the ground truth activity log. If the label in ground truth log matches the recognized log, we take a count of TP for this labeled activity ID. However, if the ground truth activity a_p is recognized as another activity a_q, we take a count of $F_{p,q}$.

Table 4.3: Example of verification matrices of TP, FN and FP. TP represents the quantity of true positive; FN represents the quantity of false negative; and FP represents the quantity of false positive.

Activity ID	**1**	**2**	**3**	**FN**
1	TP_1	$F_{1,2}$	$F_{1,3}$	FN_1
2	$F_{2,1}$	TP_2	$F_{2,3}$	FN_2
3	$F_{3,1}$	$F_{3,2}$	TP_3	FN_3
FP	FP_1	FP_2	FP_3	-

Here, for easy understanding, we use the verification matrix to represent the recognition result in this stage, as presented in Table 4.3. As shown in the table, FN is the sum of the row elements apart from the TP in that row, and FP is the sum of the column elements apart from the TP in that column. FN represents the false negative to the activity, and FP represents the false positive to the activity. Then, the precision and recall can be calculated by Equations (4.7) and (4.8):

$$precision = \frac{1}{m}\sum_{j=1}^{m}\frac{TP_j}{TP_j + FP_j}, \tag{4.7}$$

$$recall = \frac{1}{m}\sum_{j=1}^{m}\frac{TP_j}{TP_j + FN_j}. \tag{4.8}$$

The confusion matrix of the recognized activities is presented by Table 4.4. The activity IDs 1–10 represent 10 activities: leaving, toileting, showering, sleeping, breakfast, dinner, lunch, snack, spare time/TV, and grooming. From the confusion matrix, we can see that the activities having representative objects, for example, activities such as toileting and showering, can be recognized accurately.

According to the calculation, the average precision of our framework in the second stage is 85.0%, and the average recall is 87.9%. In the experiment, we find that the primary element reasonable for false recognition is the temporal-sensitive activities. From Table 4.4, we can see that activities No. 6 and No. 7 are misidentified as each other three times, respectively. They represent "lunch" and "dinner", and they are the representative temporal-sensitive activities. In the dataset, activities "breakfast", "lunch" and "dinner" are treated as three different activities. Although they are different in the temporal space, the representative objects associated with these activities are similar. However, our proposed framework does not consider the temporal knowledge, thereby making it difficult to distinguish those activities. We believe that it is not fair to divide such temporal-sensitive activities merely according to the time because the inhabitants have different daily schedules—for example, if we set the lunch time as from 11:00 a.m. to 1:00 p.m. In addition, the inhabitant may just get up at 11:00 a.m. and have the meal. If we only judge according to the time, it should be "lunch". However, "breakfast" means the first meal of the day that causes a conflict. Therefore, to recognize such temporal-sensitive activities, we would like to combine the related activities in our future work.

Table 4.4: Confusion matrix of recognized activities.

Activity ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	37	0	0	0	0	0	0	0	0	2
2	0	91	0	0	0	0	0	0	0	0
3	0	0	11	0	0	0	0	0	0	0
4	0	0	0	28	0	0	0	0	0	0
5	0	0	0	0	22	0	0	0	1	0
6	0	0	0	0	0	11	3	0	0	0
7	0	0	0	0	0	3	13	1	0	0
8	0	0	0	0	0	1	1	45	0	0
9	1	0	0	0	0	0	0	2	98	1
10	2	0	0	0	0	0	0	0	0	92

We checked the original activity data, and found that "leaving" is recorded before the inhabitant leaves the house. This means that, when the inhabitant starts to prepare to leave, the interactions are treated as related to "leaving". In addition, "leaving" is often performed after "grooming". The segmentation between "leaving" and "grooming" is not noted precisely. Some objects are therefore mistaken as representative of "leaving". This leads to the fact that "leaving" is confused with "grooming", and it could be solved by adjusting the definition of activities.

4.5.3 Third Stage

In the process of prediction, we first need to adjust and clean the original data. We treat the same activity that repeats in a short time as a pseudo record and merge them together. Furthermore, we delete some false records that do not meet the common sense. Then, we normalize the data to use it to train the LSTM model. Unlike the second stage, in the third stage, the initial 70% data are utilized for training and the remaining 30% data are for testing.

In the experiment, we build a typical LSTM model on TensorFlow-GPU with Keras as the high-level API. The training epoch is set to 10,000 to ensure the model is well trained. The LSTM model contains four layers: one input layer, two hidden layers and one output layer. The loss is set as $ascategoricalcrossentropy$, and the optimizer as $adam$. The *timestep* and $neurons$ in the hidden layers are hyperparameters. We adjust the hyper parameters to ensure optimal performance of the model.

We find that the test accuracy reaches its optimal value when *timestep* equals 3. This means that the LSTM model can utilize the past three activities to predict the next activity and achieve the highest accuracy, which accords with our assumption. Furthermore, the accuracy begins to decrease after that, meaning that the pattern of activities can not be too large; otherwise, too much noise will be used.

We also compare our model with the classical Naive Bayes method, as depicted in Figure 4-6. Because the Naive Bayes method only uses the current one activity to predict the next activity, we can see that our solution achieves much higher accuracy than that of Naive Bayes. The top two prediction accuracy reaches 65.2%. Moreover, when we apply the method to the process of prediction in the second stage, the accuracy will be as high is 78.3%.

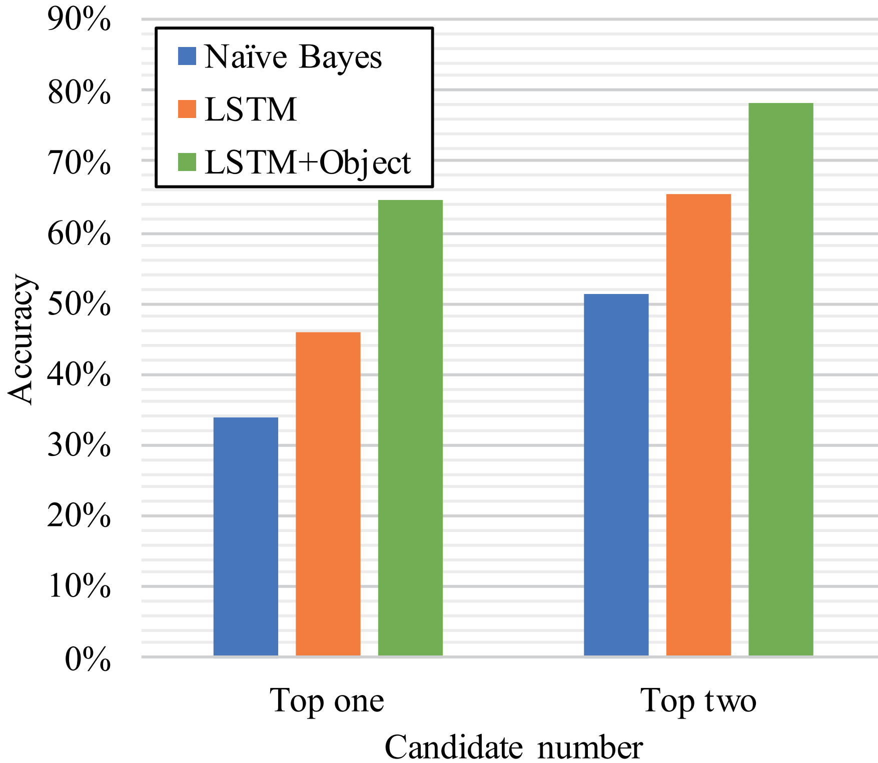


Figure 4-6: Accuracy of Naive Bayes and long short-term memory (LSTM) solution.

4.6 Summary

In this study, we presented RF-ARP, which is a three-stage framework to deal with the issue of HAR in smart homes. According to the object–usage, our framework can infer the high-level activities and further predict the next possible activity. Without any requirement to the inhabitant, the proposed framework can be widely promoted to a different house at a relative low cost and using less energy. The framework is evaluated on an open source dataset of ADL. The recognition precision can reach 85.0% and the prediction accuracy is 78.3% in the condition of two inhabitants. Compared with the existing work, our framework performs better.

However, there are still some limitations to our current framework. The framework finds it difficult to handle the activities having temporal semantic meaning. This is because our activity recognition method is mainly based on the spatial knowledge. In the future, we would like to move forward a single step to lead the temporal knowledge into our framework. Moreover, we would also like to implement our RF-ARP framework in a living environment to further verify its ability to work in the situation of different and multiple inhabitants, as the dataset used in this study is based on a single-user environment only. In addition, a standard activity definition library needs to be built to reduce the trouble of "cold start".

Chapter 5

Reading Activity Recognition in Smart Library

5.1 Introduction

Today, the Internet of Things (IoT) has created changes in various industries. With the help of perception, the core of the IoT, an increasing number of services have become intelligent and efficient. For example, personalized recommendations have been applied to improve consumer services. In most cases, only knowledge from online interaction is used, which is not sufficient. However, more offline activities such as reading in the library can be utilized to promote this kind of application, especially with the development of a large-scale RFID system[91]. Reading activity can be inferred by monitoring the state of books, providing more accurate book recommendations to the reader and assisting the library in book management.

The concept of a smart library had been put forward several years ago. Research has been conducted both in industry and academia. Min proposes his conceptual model of a next generation library information service, which is an all-encompassing infrastructure utilizing cloud computing[92]. Toivanen introduces a book-finding system based on RFID, consisting of a portable handheld reader and customized tags[93]. Markakis designs an RFID-enabled library management system using a fixed low specific absorption rate antenna[94]. Shangguan uses spatial-temporal phase profiling collected by a mobile antenna to detect the misplacement of books in the library[95]. Liu proposes an RF-Scanner which is a robot with an RFID antenna that can localize the book and detect the book if it is lying down[96]. Among these studies, the Min study pays more attention to information communication[92], the Toivanen and Markakis efforts only help the reader find a specific book[93][94], and the Shangguan and Liu proposals lay special emphasis on book misplacement[95][96]. None of these studies involve recognizing reader activity.

Our long-term goal is to develop a context-aware system that can automatically recognize reader activity in real time and provide personalized services to the reader according to the specific activity in the smart library. Several technologies can be utilized to build such a system, e.g., RFID, computer vision, and Wi-Fi. In this study, we focus on reading activity recognition based on the phase profiling of passive RFID. RFID phase is a physical attribute of the return signal from a passive RFID tag. It is collected by the RFID reader in every query of a tag. The detail of the RFID phase is introduced in Section **??**. We analyze the feasibility for detecting activity in the library using RFID phase in theory. Then, we find that the variation of RFID phase can reflect the interaction between the reader and the book. We can determine the state of the book through the variation of phase in a sliding window. So, we process the phase signal of tagged books and distinguish different activities with thresholds. Furthermore, to evaluate the effectiveness of our approach, we implement the approach on a bookshelf in which there are 50 books in total, and the COTS reader and UHF tags are used to obtain the phase.

There are several challenges in such a system. The first is that the external size and behavioral characteristics are different for each reader. This makes it impossible to train the model for a specific reader. To address this challenge, we use different features to describe the phase profiling rather than use it directly. The second challenge is that the system should operate in real time. We put the antenna on the top of each bookshelf and kept scanning the tags on the books continuously. The computational complexity is low enough to guarantee real-time operation. The third challenge is versatility. According to [96], book misplacement detection should be considered. Our approach can not only detect the activity of the reader but also has the ability to detect the misplaced book, without any other equipment.

There are three main advantages to our approach over the prevalent waveform-based activity recognition, such as described in [63]. First, only when the distance between the reader and the bookshelf is short enough is the phase waveform identifiable while the reader walks before a book. According to the verification experiment, our approach still works even for a distance greater than 80 cm. Next, the computation complexity of our approach is significantly much lower, because our approach does not have to generate a particular waveform for each activity. Last, but not least, people may pick up a book in different ways, causing totally different phase waveforms.

5.2 Preliminaries and Inspiration

The basic idea of our approach is to recognize activity based on the amplitude of the variation of RFID phase. This section provides a detailed analysis of factors that can affect phase value. We suggest that the phase of a tag is determined by three aspects: distance, angle, and multi-path effect.

5.2.1 RF Phase Values

In a typical RFID system, the reader transmits a continuous-wave signal to the tags and then receives a backscattered signal from the tags. The phase value of the RF signal describes the offset between the transmitted and received signal, which depends on the round-trip ($2d$) and hardware-specific factors[66]. Parameter d is the distance between the antenna and tag. For a standard phase-distance model, the hardware-specific factors include three parts: θ_T, θ_R, and θ_{TAG}, which represents the transmission circuit, receiver circuit, and tag reflection characteristic of the reader.

Furthermore, the RF phase is a periodic function ranging between 0 and 2π. Thus, the phase value θ of a tag measured by the reader can be expressed as:

$$\theta = mod(2\pi\frac{2d}{\lambda} + \theta_T + \theta_R + \theta_{TAG}, 2\pi) \tag{5.1}$$

where λ is the signal wavelength[97]. The phase value is a common attribute that is supported by most COTS RFID readers, e.g., ImpinJ R420[98]. Note that the hardware-specific factors are constant since the devices are produced. Phase measurement contains random errors in real-world situations, and different tags have different levels of random errors following a Gaussian distribution, with a standard deviation of 0.1 radians. As denoted in Eq.5.1, the phase is linear to the distance between the antenna and tag, ignoring the random error.

For example, ImpinJ R420 can be set to work at 920$\sim$926 MHz with 16 channels. Thus, λ is approximately 320 mm, and θ will repeat if the distance is larger than 160 mm. In particular, when the θ is near 0 and keep decreasing, θ jumps from 0 to 2π, and vice versa. This is called the "phase jump".

Eq.5.1 only considers distance, ignoring the phase deflection caused by the rotation of tags. For most phase-based application, Eq.5.1 works well. However, the rotation of devices has to be considered in our situation to distinguish different activities, which will be discussed in the next subsection.

5.2.2 Rotation of Tags

The phase of a tag is affected by the relative orientation to the reader antenna. This observation is mentioned in the Wei study, published in 2016[99]. To reinvestigate this phenomenon, we conduct an experiment as illustrated in Fig.5-1.

As depicted in Fig.5-1, we place a slim tag 1 m in front of a reader antenna. The center of the antenna and tag is aligned to ensure that the distance between the antenna and tag does not vary. Then, the tag is rotated along each axis. Moreover, the tag is placed facing the antenna on the both front and side to investigate if the

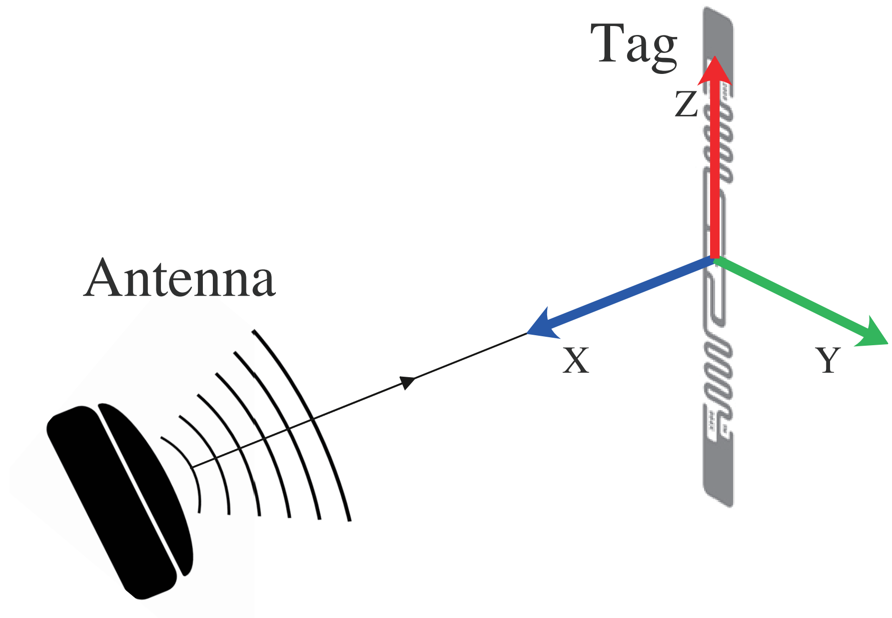
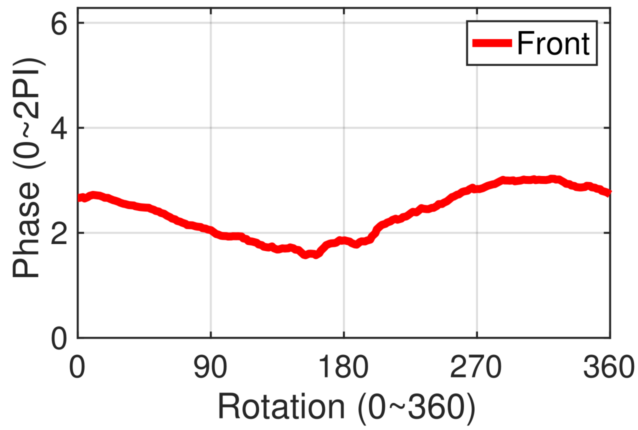


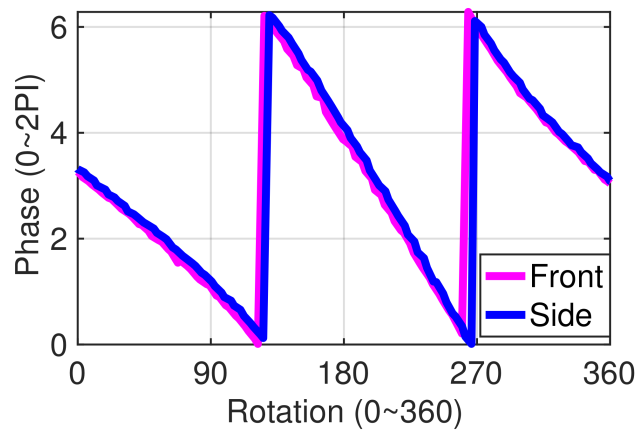
Figure 5-1: Measuring the phase of a tag rotating along each axis.

posture of tag influences the phase or not.

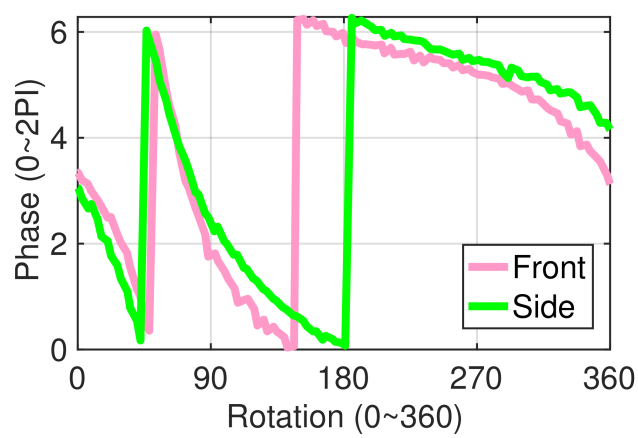
The result is shown in Fig.5-2. In Fig.5-2(a), there is only one line, because, when the tag is rotated along the Z-axis, the posture of the tag does not matter. We can see that the phase does not change much while rotating along the Z-axis. However, it does shifts periodically when it is rotated along the X-axis and Y-axis as it is shown in Fig.5-2(b) and Fig.5-2(c). The phase changes linearly with rotation along the X-axis and nonlinearly with rotation along the Y-axis, and the phase jumps twice per period. Wei[99] explains this phenomenon with the polarity of RFID antennas. The tag's antenna is usually designed as a dipole, which is linear-polarized along the tag body. However, most RFID readers' antennas are circular-polarized. They comprise two perpendicular dipoles, fed with a signal of a 90° phase difference to ensure that they can read tags from a variety of angles. Thus, when the tag rotates, the phase of the signal received will change. Moreover, the signal traverses a round-trip, so the corresponding phase shift, measured by the reader, doubles. This means that the phase will change by π when rotating around both the X-axis and Y-axis by 90°. Furthermore, we find that whether the tag is facing the antenna on the front or side,



(a) Rotate the tag along Z-axis



(b) Rotate the tag along X-axis



(c) Rotate the tag along Y-axis

Figure 5-2: Measured phase of a tag rotating along three axes.

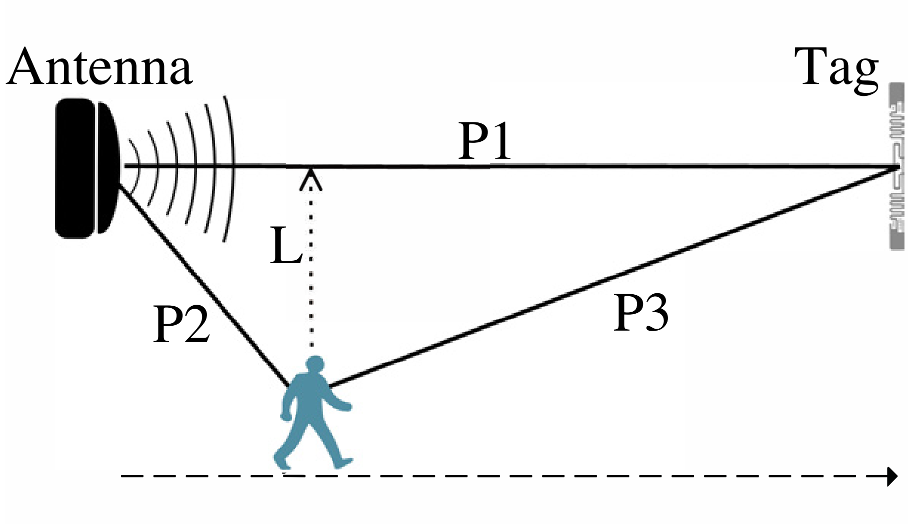


Figure 5-3: Measuring the phase under multi-path effect.

the phase changes in the same pattern.

5.2.3 Multi-path Effect

The RFID signal is a kind of radio wave with specific frequency. So, it undergoes the multi-path effect as other radio waves do[100]. The wireless signal is radiated to all directions rather than in a line. While some signals go directly to the destination, other signals are reflected by surroundings and then proceed to the destination. Thus, different paths are generated. The multi-path effect is caused by the difference in the distance of these paths. The multi-path effect influences the phase according to Eq.5.1.

To better understand the multi-path effect, we monitor the phase under the multi-path effect in a real-world situation as described in Fig.5-3. A tag is placed 2 m in front of the antenna. $P1$ is the line-of-sight(LoS) path. A man walks along a straight line, parallel with $P1$, and the gap between $P1$ and the man is L. Thus, the signal reflected by the man will generate a new path, $P2$-$P3$. The man repeats walking along the line, changing L from 10 cm to 110 cm. The phase measured by the reader is shown in Fig.5-4. From this figure, we can see that even when the distance between

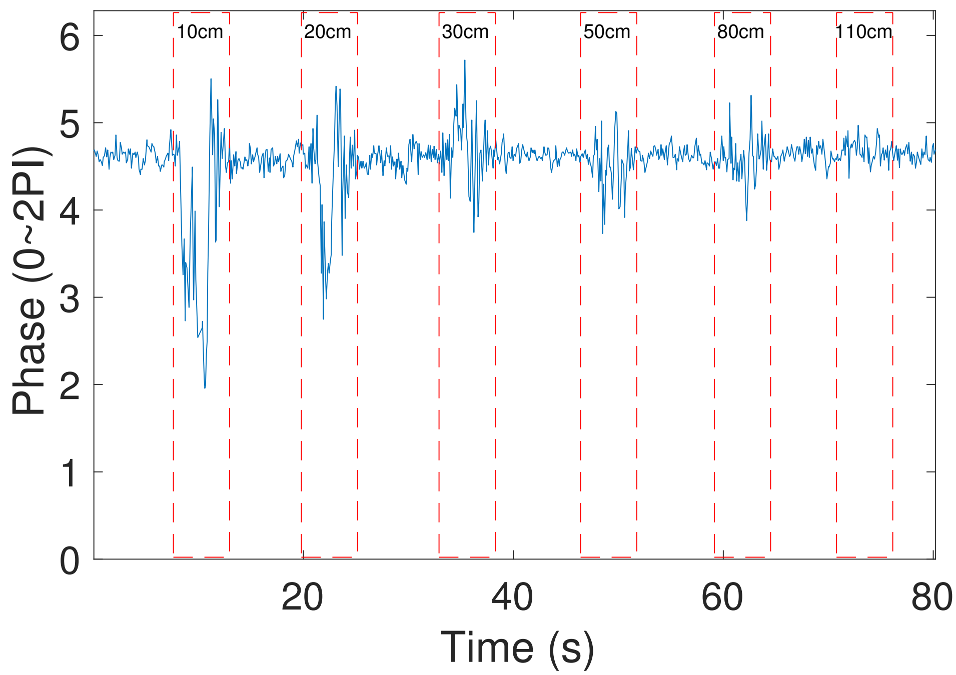


Figure 5-4: The phase changes while L is set from 10 cm to 110 cm. The phase in the red frame is affected by the multi-path effect.

reader and book is 80cm, phase data can still reflect this motion. This range is large enough to cover the entire bookshelf from top to bottom.

From Fig.5-4, we can see that the rangeability of the phase becomes decreasingly smaller with the growth of L. When L reaches 110 cm, it is hard to distinguish between the multi-path effect and random error. There are two reasons for this phenomenon. First, when L increases, the propagation fading F increases accordingly based on wireless communication principles[101]. F is calculated as follows:

$$F = 10log(\frac{\lambda^2}{16\pi^2 d^2}) \tag{5.2}$$

where d is the length of the path between the transmitter and receiver. The propagation fading may make the signal transmitted from the transmitter along $P2$-$P3$ invisible to the receiver sometimes. Thus, when L increases, there is no specific wave shape since L is larger than 30 cm.

Another reason is that the multi-path effect occurs when the path is within the Fresnel Zone[101]. A Fresnel zone is an ellipsoid whose foci are the transmitter and the receiver. Radius r of the circular cross section of the Fresnel Zone is given by:

$$r_n = \sqrt{(\frac{d}{2} + \frac{n\lambda}{4})^2 - (\frac{d}{2})^2} \tag{5.3}$$

where d is the distance between the transmitter and receiver and λ is the wavelength. Particularly, n is the Fresnel Zone number which is commonly smaller than 12. That is why when L is larger than r_{12}, the multi-path effect does not occur. It is clear that the length of the path within a Fresnel Zone becomes shorter as L increases. It is noteworthy that if the line along which a man walks is parallel to the LoS, the line will not stride across more than one Fresnel Zone. This means that the range of phase variation caused by the multi-path will never exceed 2π.

Inspiration: Previously, we analyze three aspects that can bring about changes to the phase value. Specifically, activities in the library imply these aspects more or less. This inspires us to infer the activity by monitoring the transformation of the phase value. Although there are some applications based on RFID phase applied to areas, such as object use detection[102][103], gesture recognition[63], and indoor localization[66], this study is the first to involve activity recognition in the library according to the intensity of the variation of phase. Details of our system are expounded in the next section.

5.3 System Design

In this section, we present the details of our approach to detect the activity based on RFID in the library. Our approach utilizes the RFID reader antenna placed on the top of the bookshelf to monitor the phase value of tags attached to books. We first provide an analysis of the relation between activity and phase transformation. Then, a method based on the phase value distribution is proposed to infer the activity as

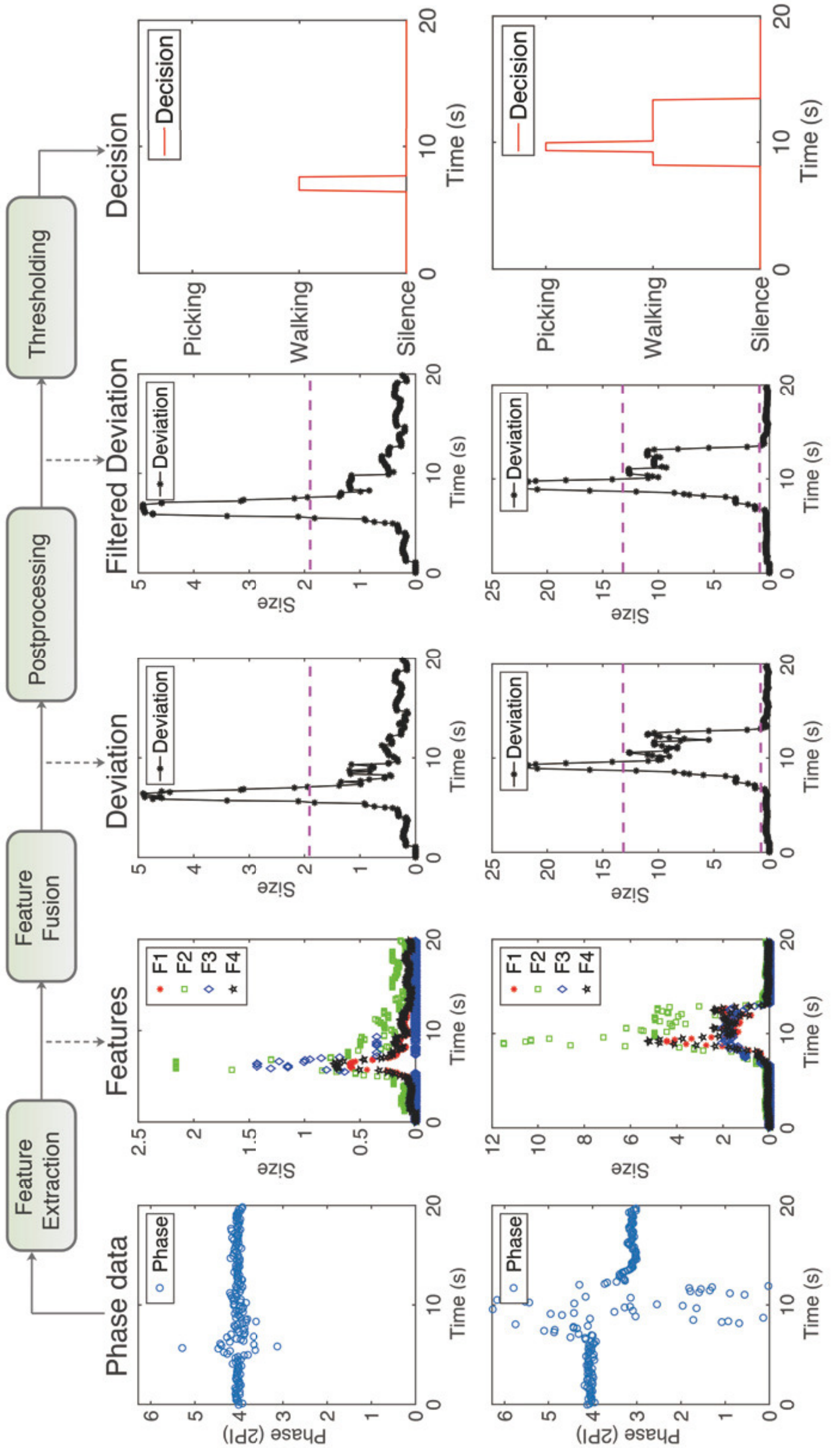


Figure 5-5: The architecture of activity recognition approach. F1, F2, F3 and F4 in the second picture represent four features respectively. The upper line is an example of walking before a book and the under line is an example of picking up a book.

shown in Fig.5-5. We first collect phase data of every tag by an RFID reader and describe the dispersion of the phase data in a sliding window with four features. After feature fusion and postprocessing, we utilize thresholding to differentiate the states of each book and then recognize the corresponding activity of the reader.

5.3.1 Activity Definition

In this study, we consider the reader in the library who wants to choose books at the bookshelf and then leaves to read as the target. Generally speaking, such a behavior in the library follows a common process. First, the reader walks to the area where there are books he has interest in. Next, the reader will pick up books one by one and read for a while to find out the ones that he prefers to read through. Then, he will take the books to the reading area. After reading, he replaces the books to where they were. This process actually contains four activities: walking before a book, picking up books, taking books away and putting books back. Because the RFID antenna has a limited examination area, it is easy to detect whether one book is taken away or put back through the tag visibility. Thus, the challenge is how to distinguish the first two activities from each other. Of course, there is also the static status, which means that no one comes into the examination area.

According to the analysis in Section 5.2, it is clear that walking before a book only causes the multi-path effect. Meanwhile, picking up a book will change both the tag angle and the distance between the tag and antenna. When the reader walks before the bookshelf, the book on the bookshelf is labeled "Walking", no matter if the reader is walking with books or without books. The book picked up by the reader is continuously labeled "Picking" until it is put back on the bookshelf. If no one is moving before the bookshelf, all the books are labeled "Silence". Thus, the system is able to recognize the reader activity through the book state.

Furthermore, the reader sometimes returns the book to the wrong place. In this study, we propose a nearly zero-cost way to detect the misplacement of books.

Figure 5-6: The white antenna is placed at a 45° angle to the ground to ensure it can cover as much region as possible, because of its fan-shaped radiation regions.

5.3.2 Data Collection

We carry out this approach in one bookshelf on which there are 50 books in total, as depicted in Fig.5-6. The books are placed closely to each other in one row, to test the robustness of the system under the mutual interference of tags. We collect RFID data using an ImpinJ R420 reader with only one antenna, setting the reader to max-through power and dual-target search mode[104]. The slim tags are attached to the back of each book. The reader keeps querying the state of each tag within the examination area. The RFID reader is set to only query the tags whose ID exist in the predefined list, so that the other books do not impact the specific bookshelf. In our situation, the mean sampling rate for each tag is about 12 times per second. In Fig.5-5, blue circles represent the phase data of one tag, which fluctuate with different activities.

Actually, the attributes of the RFID signal include not only the phase but also the received signal strength (RSS) and Doppler shift. However, our pretest result shows that the Doppler shift is not accurate enough for our purposes, which was also mentioned by [105]. Moreover, it is difficult to combine RSS with different activities in theory. Thus, we just use the phase to achieve our goals.

Nevertheless, the pretest highlights a bug while measuring the phase. As shown in Fig.5-7, the measured phase happens to generate a deviation of π randomly, which is called the "half-wave effect". This will produce some fake phase values. Although some filters can overcome this, they are not portable to an online system. Thus, we constructed a regularized equation to solve this issue:

$$p_r = (p_o + mod(p_o + \pi, 2\pi)) - \pi \tag{5.4}$$

In the above equation, p_o is the original measured phase value, and p_r is the regularized phase value. Regularization first generates the mean value of the true phase value and the fake phase value. Then, it considers the values from 0.5π $\sim 1.5\pi$ to $0 \sim 2\pi$. Regularization does not need to distinguish if the phase is fake or not. Note, regularization does not affect the tendency of the phase, even though it changes all of the phase values. In this study, only the dispersion of phase value is important. Thus, regularization will double the dispersion, which is good for our approach.

5.3.3 Phase Dispersion of Activity

Other than the phase waveform-based activity recognition approach[63], we recognize the activity based on the phase dispersion within a sliding window. As depicted in Fig.5-5, "Picking" causes much more violent dispersion than "Walking". To quantify this phenomenon, we first extract the features to describe the phase dispersion. Then, we use the weighted summation method to fuse the features into a deviation index. After postprocessing, we can finally generate the recognition decision.

The traditional way to recognize the activity using RFID is based on the specific phase waveform caused by relative gesture. There are several reasons why the use of the waveform-method is not appropriate in the library. First, the distance between

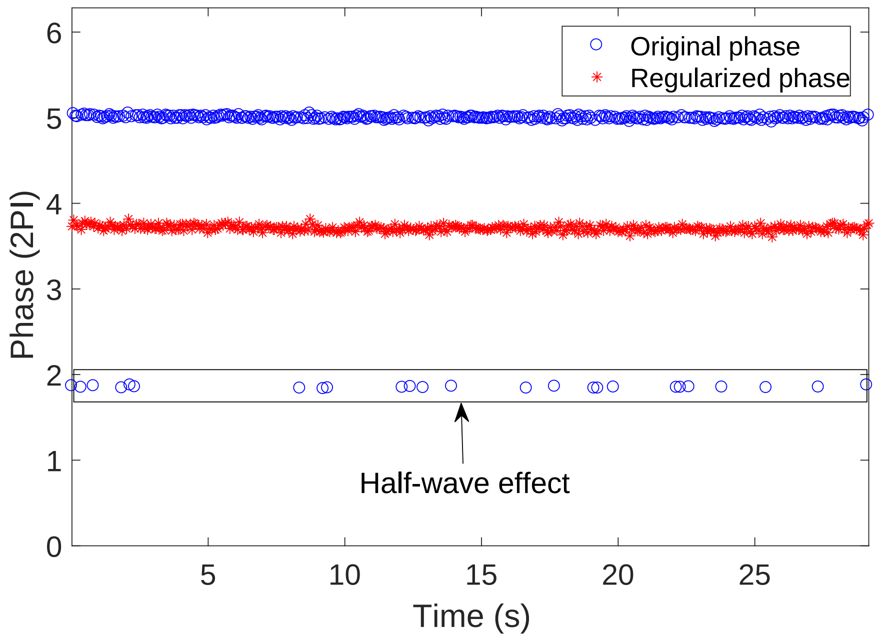


Figure 5-7: Half-wave effect and the regularization. The blue circle is the original phase value and the red star is the phase after regularization.

the reader and book has to be short enough that it can generate a stable waveform when the reader walks before a bookshelf. As shown in Fig.5-4, the distance has to be less than 20 cm, which is impractical in a real scenario. Also, people may pick up a book in different ways, along with the fact that the stature of each reader is different. This makes it impossible to produce a congruent waveform when the reader picks up a book. Nevertheless, our approach will not suffer from the above problems.

There is the challenge of calculating the amplitude of the variation of phase. As introduced in Section **??**, the phase value is a periodic function, called the wrapped phase, which causes the traditional functions not to calculate accurately. So, we have to unwrap the phase before computing the deviation. Here, we utilize a common way to solve this problem called the One-Dimensional Phase Unwrapping method[106]. This method is efficient and has a low computational complexity. It assumes that the difference between two adjacent sampling points over π is negligible. This is feasible since the phase is measured constantly, and the sampling rate is high enough.

The traditional indexes to weigh the dispersion include the standard deviation, range, entropy, and coefficient of variation. In this study, we mainly use the first three features to describe the phase dispersion. Furthermore, the arithmetical average deviation of phase values in a sliding window can also be used as a feature. For a phase dataset (X) in a sliding window, the intensity of the phase variation can be described by the following indexes.

Phase Standard Deviation

In statistics, the standard deviation is used to quantify the amount of variation or dispersion of a set of data values. Generally, the standard deviation $\sigma(X)$ is calculated by the following equations:

$$\mu = \frac{1}{N}(x_1 + \cdots + x_N) \tag{5.5}$$

$$\sigma(X) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{5.6}$$

Phase Range

The range is used to describe the scope of random variables. After phase unwrapping, the range of phase can be calculated directly by the difference between the maximum and the minimum.

$$R(X) = max(X) - min(X) \tag{5.7}$$

Range is easy to understand and calculate. However, it is sensitive to noise. Thus, other indexes are necessary.

Phase Entropy

Entropy can be used as an expression of the randomness of a system. In our approach, we use entropy H to describe the phase value distribution as below.

$$H(X) = -\sum_{i=1}^{n} p(x_i)logp(x_i) \tag{5.8}$$

It is known that the phase value varies between 0 and 2π. We divide the range into n equal portions, and $p(x_i)$ are the quantities of the phase value that are within a specific portion. In this study, we set n equals to 10 and e as the base, serving as the default parameters. According to information theory, the larger the phase entropy is, the more evenly the phase values are distributed.

Phase Average Deviation

$$\delta(X) = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|x_i - x_j|}{\frac{n*(n-1)}{2}} \tag{5.9}$$

The phase average deviation $\delta(X)$ is the mean value of the differences between any two phase values. This index can reflect the stability of the phase values in a sliding window.

Feature Fusion

Feature fusion merges all the features together to generate a robust index for recognition.

$$M(X) = exp(c_1\sigma(X) + c_2R(X) + c_3H(X) + c_4\delta(X)) \tag{5.10}$$

Finally, the comprehensive deviation M can be calculated by Eq.5.10, where c_j is the weight of features. In Fig.5, we can see that four features describe the dispersion in different degrees. With feature fusion, the mixed set of indexes has the dramatic ability to discriminate the book state.

5.3.4 Thresholding

In this study, we first need to detect if there is an activity. Then, we further recognize if the activity is picking up a book or not. If a book is not picked up, the activity must be walking before a book. To ensure the system can achieve good instantaneity, we choose thresholding rather than a complex machine learning technique.

On the grounds of the analysis in Section **??**, if the surroundings are steady, the phase value will not change. However, the phase measurement contains random errors in real-world situations, causing the phase to follow the Gauss distribution instead of an accurate value. So, we have to first obtain the phase dispersion in a static condition. We assume that the feature that describes the phase distribution also follows the Gauss distribution. Thus, we can use the upper limit of each tag phase feature distribution function to detect if there is an activity. If the value of the feature is greater than the upper limit, it is assumed that there is an activity.

Another threshold is needed to distinguish between walking before a book and picking up a book. This threshold is generated by a large number of experiments. As discussed in the previous section, picking up a book will change both the angle of tag and the distance between tag and antenna. Therefore, the phase dispersion of picking up a book is definitely larger than walking before a book. Therefore, we let several people who are different in stature walk before the bookshelf many times and collect the phase data at the same time. Then, we find the maximum phase dispersion with this phase data. Hence, the maximum value can be used as a threshold. However, we find that this does not work through experiment. The reason is that when the reader picks up a book, it will influence the neighboring books. Thus, the deviation index of these books will increase. So, we change the strategy. The volunteer is asked to pick up books successively from the first book to the last several times. Meanwhile, we record the phase values of all the books, and their deviation indexes are computed. Then, we count the maximum deviation index for all the books except the book picked up. Finally, we repeat a few times to generate the final maximum for every book and use this as the threshold.

In Fig.5-5, the purple broken lines represent the threshold to distinguish the activities. Note that in the implementation, we set two parameters according to the pre-experiment for the two thresholds to generate a better result. The final experiment shows that the thresholds can be used universally.

5.3.5 Postprocessing

Using a threshold can distinguish different activities; however, it is quite essential to achieve a higher recognition postprocessing rate. We find that the deviation index calculated by Eq.5.10 is not a smooth function. Sudden drops and rises will cause incorrect recognition decisions. We process the deviation index with the following steps:

Smoothing

We set two sliding windows w_1 and w_2, for each deviation index point. When a deviation index is generated, we check the deviation indexes in w_1. We use the maximum in w_1 instead of the current deviation index. This can filter out the sudden drops. Later, we compare the index with the thresholds to generate a pre-decision. Then, when a recognition pre-decision is generated, we check the pre-decisions in w_2. If they are equal, the pre-decision is true; otherwise, it is a sudden rising noise.

Merging Adjacent Instances

If a book is picked up several times, while the reader is standing right in front of the bookshelf, we treat the activity as one time. This may happen when the reader is trying to compare two books with each other.

5.4 Experiment and Evaluation

To evaluate the proposed approach, we carry out some experiments. The performance is judged by recognition accuracy, as shown below:

$$P_a = \frac{\sum_{i=1}^{N} A_{ti}}{N * m} \tag{5.11}$$

$$P_b = \frac{A_{ti}}{m} \tag{5.12}$$

where N is the quantity of books which is 50 in this paper, A_{ti} is the amount of correct recognitions of book i, and m is the frequency of test activities. Eq.5.11 represents the overall accuracy and Eq.5.12 represents the partial accuracy. Note that, if there is no one walking before the bookshelf, the system will stay silent. After 1 hour, there was no false positive. Thus, our approach works very well to detect the existence of the reader.

In the first part, we test the performance of detecting "Walking". We ask five volunteers who differ from each other in both height and weight to walk past the bookshelf from left to right and in the opposite direction 25 times. Then, the accuracy is calculated with Eq.5.11, as shown in Table 5.1. In this part, we only use Eq.5.11, because all the books are in the same situation for the activity "Walking".

Table 5.1: Performance of detecting walking past.

Volunteers	P1	P2	P3	P4	P5
Accuracy	97.28%	95.92%	96.76%	96.88%	95.84%

From this table, we can see that our system performs well for detecting "Walking". Moreover, this recognition is only achieved for an individual book, separately. If we mearge the results for the adjacent books together, the accuracy may be even higher.

In the other part, we test the performance of detecting "Picking". We ask a volunteer to walk to the bookshelf, randomly pick up a book, put it back, then walk away. This is repeated 50 times. The two results in Table 5.2 reflect the partial and the overall accuracy. For the former one, we only consider the books that have been picked up. The partial accuracy is calculated by Eq.5.12, where m means the number of test activities which equals to 50. A_{ti} represents the successful recognition of book i. For the latter one, we consider all the books in the calculation with Eq.5.11. For the chosen book, "Picking" is the right recognition. Meanwhile, for the other books, "Walking" is the right recognition. That is to say, this overall accuracy can be used to measure the global performance of the proposed system.

Table 5.2: Performance of detecting picking up.

Result	Partial	Overall
Accuracy	96%	92.2%

The overall accuracy is 92.2%, which is lower than the first part of the experiment. This is because when a book is picked up, it will affect the neighboring books. In future research, we will try to avoid this error.

5.5 Misplaced Book Detection

As mentioned above, the misplacement of books in the library causes difficulty in book management. How to detect misplaced books is a challenge for the future smart library. Note that books in the library are strictly ordered by their IDs to ensure that readers can find a specific book easily. The approach proposed by [95] utilizes the relative localization of RFID tags to locate the misplaced books. However, the approach needs to sweep all the books repeatedly, which results in a high workload in a large library. Our proposed work can greatly reduce the area that needs to be swept, because if none of the books in the area is picked up, there will be no misplacement. Furthermore, we present a zero-cost approach to detect the misplacement remotely,

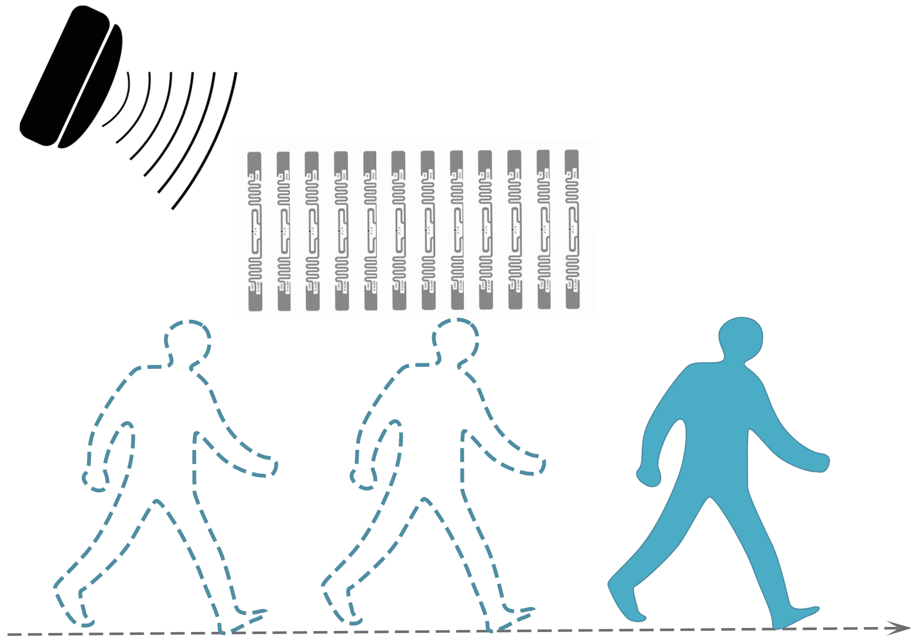


Figure 5-8: The sequence of time slot that caused by reader's movement reflects the order of books.

which can also work in real time.

We believe that when a reader walks before a bookshelf, it will cause phase variation successively to the books in one row. Through monitoring the phase deviation of each tag on the books, we can generate the order of the books. The basic idea is that once a reader walks before a bookshelf, our system can check the order of the books one time. If this order does not match the original order of the books, the misplaced book can be detected easily. In this way, the system can check the misplacement in real time, without any extra devices or extra work of the staff.

As shown in Fig.5-8, our activity recognition model can obtain the time slot that the reader walks before a book. We use the timestamp corresponding to maximal deviation point within the time slot to mark the order of the book. After the reader walks before a row of books, there will be a sequence on the timeline. This sequence represents the order of books in a row. However, the result of the verification experiments shows that the sequence does not necessarily correspond to the order of books. The distance n between the misplaced book and its original position influences the performance of our proposed approach. Thus, a trade-off between accuracy and recall is performed by redefining the book misplacement. If n is less than limitation N, we do not treat it as a misplacement. In this study, we set N equal to 10, which is acceptable in the actual scene. To further reduce false detection, we label it a true misplacement only if the book is reported as a misplacement three times.

In this way, our approach can monitor the entire area in real time. The main advantage of this approach is that the reader is not hindered, which allows the library to be open continuously. On the contrary, the approach relies on the reader without any additional requirements. Theoretically, the more readers there are in the library, the more efficiently our approach works. Although it is a coarse-grained detection approach, it also can be used in cooperation with other robot-based methods to reduce the cost both in time and energy.

5.6 Summary

In this study, we analyze the feasibility of detecting activity in the library using RFID phase, in theory. Then, we find that the variation of RFID phase can reflect the interaction between the reader and the books. Subsequently, we process the phase signal of tagged books and distinguish the different activities with thresholds. Finally, we implement our approach to evaluate the performance and overall accuracy, which reaches 92.2%. Also, we present a way to detect the book misplacement that works in passive mode.

Our proposed system can detect a reader's existence with very high accuracy and in real time. This makes reader location tracing possible, which is of great importance in a smart library. Any indoor localization-based method may be introduced into the library, thanks to the high accuracy of localization with books. Furthermore, the system can navigate the reader to the desired book with the shortest path. Moreover, the library can adjust the book position based on the statistical result of reader trajectories.

In the future, we will focus on the personalized book recommendation system. This will bring benefit to both the library and reader. The library can acquire the demand of a specific book and put the popular one in a prominent position to attract more readers. The system can also help the reader to find what he might be interested in, not only based on the borrow record but also on the pick record from him and other readers. Similar personalized recommendation systems have been put forward in other areas. In this vein, the Han study on in-store shopping recommendations is worth learning. Han proposes CBID[103], which is a customer behavior identification system. With CBID, a store can discover a popular item and its relation to a customer's movement pattern. This is what we want to achieve in a future smart library.

Chapter 6

Conclusion

6.1 Discussion and Conclusion

To recognize the human activity will greatly improve the ability of intelligent indoor IoT systems. Smart IoT systems represented by smart home will provide human-oriented service in different scenarios in the near future. While how to make the computer systems to understand human mind and activity still face a lot of challenges. Even though inner state of human can already be monitored in different appropriate ways. High level human activity is so complex that can hardly be comprehended by computer, because of the diverse semantics in variety of activities. This builds the gap between human and computer, naturally. And the characteristic of human activity, include concurrency, multiplicity, complexity, diversity, and randomness, make great trouble for computer systems to recognize the human activity. Besides, the hierarchical human activities covering fine-grain and coarse-grain activities, make it difficult to define the activity in a suitable way.

Inspired by the existing work, this dissertation proposed a novel interaction-based HAR approach. The approach can be scalable enough to be applied in different indoor scenarios. In this dissertation, we chose smart home and smart library as the main scenarios. The dissertation makes a deep analysis on indoor human activity, and finally determines to utilize the interaction between human and objects to infer the human activity. Because only in this way, the proposed approach can be scalable, flexible, and avoid of "cold start" problem which appears commonly in data-driven approaches. The proposed approach utilizes passive RFID technology to recognize the interactions between human and the objects in daily life. And with the help of machine learning and deep learning, the proposed approach can combine the recognized low level activities to infer the high level activities. So that the proposed approach can be applied to recognize all grain activities including both fine-grain and coarse-grain activities.

In the smart home scenario, the passive RFID technology is used to monitor the usage of daily life first, and to record the activity log accordingly. Thanks to the tf-idf weight, which we treat as the possibility of one action belongs to one specific high level activity, the proposed approach can recognize the human activity in real time. What is more, we also utilize the LSTM to model the habit of the inhabitant, which represents the pattern of the activity in daily life. With the modeled pattern, the proposed approach can predict the next possible activity that the inhabitant might perform. The whole framework can be seen as a smart agent that can understand the human activity and keep eyes on the inhabitant without any privacy problems. And the framework can cooperate with existing smart home platforms to enable them provide more personalized and intelligent service to human.

In the smart library scenario, the passive RFID tags on the back of the books are utilized to monitor the reading activity in the library. The proposed approach does not require the reader to carry any extra devices, while monitoring which book have been picked up to take a look. Knowing this knowledge, more sensitive personalized book recommendation can become true. Since the system can aware of what the reader might have interest in according to the books that has been picked up for trial reading. Moreover, the proposed approach can help the book management locate the misplaced book with the unawareness help from the reader. In this way, much effort on book management can be saved since locating the misplaced book is a tough task for librarians in the past.

6.2 Directions and Future Works

This dissertation proposed an interaction-based HAR for indoor scenarios using passive RFID technology. In the future, the proposed approach can be applied to more different scenarios, such as office, hospital, gymnasium. And for different scenarios, the activities of human are different. So the approach need to be adjusted to adapt to those scenarios. For example, what the boss wants to know about the employee might be the working state of them. While, in the hospital, the state of patient might be helpful to doctor to determine the treatment. Thus the proposed approach is promising to show the ability in more scenarios.

Also, the proposed approach can be seen as a smart agent in the smart indoor systems. Even though the approach can understand the human activity, how to format the recognized knowledge is another challenge. Only when the knowledge of human activity is formalized, the upper systems can make use of such knowledge without trouble. And formalized knowledge will make the approach universal to different systems.

About the human activity, there is still more we need to do. The proposed approach can be used to recognized and predict the human activity. While we can hardly know if the activity is performed correctly. This issue is very urgent and waiting to be solved in the area of fitness. With the development of society, sharing economy becomes a new trend all over the world. Among them, the shared gym favored by the urban employee has achieved success in some big cities. The users are allowed to do exercise in a small closed space alone without any disturbing from others. The good privacy protection will make such shared gym more popular in the future. However, there is also a risk to allow users to exercise by themselves. The incorrect posture may lead to huge damage to human body. Thus, to monitor the action precisely rather than just to know what the action is, becomes a topic that needs more research. Thus, we would like to explore new solution with wireless sensing in the future works.

Bibliography

[1] K. Ashton *et al.*, "That 'internet of things' thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.

[2] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[3] L. Zhou, D. Wu, J. Chen, and Z. Dong, "When computation hugs intelligence: Content-aware data processing for industrial iot," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1657–1666, 2017.

[4] P. Daugherty, P. Banerjee, W. Negm, and A. E. Alter, "Driving unconventional growth through the industrial internet of things," *accenture technology*, 2015.

[5] C. Yang, R. Sui, W. S. Lee, and Q Zhang, "Precision agriculture in large-scale mechanized farming," *Precision Agriculture Technology for Crop Farming*, pp. 177–212, 2015.

[6] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25–36, 2014.

[7] H. Farhangi, "The path of the smart grid," *IEEE power and energy magazine*, vol. 8, no. 1, pp. 18–28, 2009.

[8] *Apple homekit*, https://developer.apple.com/homekit/.

[9] *Samsung smartthings*, https://www.smartthings.com.

[10] *Google nest*, https://nest.com.

[11] K. G. Davis and S. E. Kotowski, "Prevalence of musculoskeletal disorders for nurses in hospitals, long-term care facilities, and home health care: A comprehensive review," *Human factors*, vol. 57, no. 5, pp. 754–792, 2015.

[12] M. Asadullah and A. Raza, "An overview of home automation systems," in *2016 2nd International Conference on Robotics and Artificial Intelligence (ICRAI)*, IEEE, 2016, pp. 27–31.

[13] *Apple watch*, https://www.apple.com/apple-watch-series-5/.

[14] *Panasonic ew-nk63*, https://panasonic.jp/calorimeter/p-db/EW-NK63.html.

[15] Y. Liu, L. Nie, L. Liu, and D. S. Rosenblum, "From action to activity: Sensor-based activity recognition," *Neurocomputing*, vol. 181, pp. 108–115, 2016.

[16] A. H. Maslow, "A theory of human motivation.," *Psychological review*, vol. 50, no. 4, p. 370, 1943.

[17] S. Saguna, A. Zaslavsky, and D. Chakraborty, "Complex activity recognition using context-driven activity theory and activity signatures," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 20, no. 6, p. 32, 2013.

[18] J. C. Augusto, H. Nakashima, and H. Aghajan, "Ambient intelligence and smart environments: A state of the art," in *Handbook of ambient intelligence and smart environments*, Springer, 2010, pp. 3–31.

[19] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[20] A. Veeraraghavan, A. R. Chowdhury, and R. Chellappa, "Role of shape and kinematics in human movement analysis," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, vol. 1, 2004, pp. I–I.

[21] G. Xu *et al.*, "Viewpoint insensitive action recognition using envelop shape," in *Asian Conference on Computer Vision*, Springer, 2007, pp. 477–486.

[22] X. Lu, Q. Liu, and S. Oe, "Recognizing non-rigid human actions using joints tracking in space-time," in *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.*, IEEE, vol. 1, 2004, pp. 620–624.

[23] E. Shechtman and M. Irani, "Space-time behavior-based correlation-or-how to tell if two underlying motion fields are similar without computing them?" *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 11, pp. 2045–2056, 2007.

[24] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM international conference on Multimedia*, ACM, 2007, pp. 357–360.

[25] W.-L. Lu and J. J. Little, "Simultaneous tracking and action recognition using the pca-hog descriptor," in *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, IEEE, 2006, pp. 6–6.

[26] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, Ieee, 2011, pp. 1297–1304.

[27] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3d points," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, IEEE, 2010, pp. 9–14.

[28] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning actionlet ensemble for 3d human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 5, pp. 914–927, 2013.

[29] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A review on human activity recognition using vision-based method," *Journal of healthcare engineering*, vol. 2017, 2017.

[30] M. C. Mozer, "The neural network house: An environment hat adapts to its inhabitants," in *Proc. AAAI Spring Symp. Intelligent Environments*, vol. 58, 1998.

[31] S.-W. Lee and K. Mase, "Activity and location recognition using wearable sensors," *IEEE pervasive computing*, vol. 1, no. 3, pp. 24–32, 2002.

[32] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," in *International conference on pervasive computing*, Springer, 2004, pp. 18–32.

[33] H. Harms, O. Amft, G. Tröster, and D. Roggen, "Smash: A distributed sensing and processing garment for the classification of upper body postures," in *Proceedings of the ICST 3rd international conference on Body area networks*, ICST (Institute for Computer Sciences, Social-Informatics and . . ., 2008, p. 22.

[34] T Finni, M. Hu, P Kettunen, T Vilavuo, and S Cheng, "Measurement of emg activity with textile electrodes embedded into clothing," *Physiological measurement*, vol. 28, no. 11, p. 1405, 2007.

[35] C. D. Metcalf, S. Collie, A. Cranny, G. Hallett, C. James, J. Adams, P. Chappell, N. White, and J. Burridge, "Fabric-based strain sensors for measuring movement in wearable telemonitoring applications," 2009.

[36] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones.," in *Esann*, 2013.

[37] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua science and technology*, vol. 19, no. 3, pp. 235–249, 2014.

[38] E. M. Tapia, S. S. Intille, L. Lopez, and K. Larson, "The design of a portable kit of wireless sensors for naturalistic data collection," in *International Conference on Pervasive Computing*, Springer, 2006, pp. 117–134.

[39] D. H. Wilson and C. Atkeson, "Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors," in *International Conference on Pervasive Computing*, Springer, 2005, pp. 62–79.

[40] M. Srivastava, R. Muntz, and M. Potkonjak, "Smart kindergarten: Sensor-based wireless networks for smart developmental problem-solving environments," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, ACM, 2001, pp. 132–138.

[41] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel, "Inferring activities from interactions with objects," *IEEE pervasive computing*, no. 4, pp. 50–57, 2004.

[42] K. P. Fishkin, M. Philipose, and A. Rea, "Hands-on rfid: Wireless wearables for detecting use of objects," in *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, IEEE, 2005, pp. 38–41.

[43] D. J. Patterson, D. Fox, H. Kautz, and M. Philipose, "Fine-grained activity recognition by aggregating abstract object usage," in *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, IEEE, 2005, pp. 44–51.

[44] M. R. Hodges and M. E. Pollack, "An 'object-use fingerprint': The use of electronic sensors for human identification," in *International Conference on Ubiquitous Computing*, Springer, 2007, pp. 289–303.

[45] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "Epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition," in *2009 IEEE International Conference on Pervasive Computing and Communications*, IEEE, 2009, pp. 1–9.

[46] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Aaai*, vol. 5, 2005, pp. 1541–1546.

[47] U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, "Location and activity recognition using ewatch: A wearable sensor platform," in *Ambient Intelligence in Everyday Life*, Springer, 2006, pp. 86–102.

[48] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *IEEE pervasive computing*, vol. 9, no. 1, pp. 48–53, 2009.

[49] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden markov models for complex action recognition," in *CVPR*, vol. 97, 1997, p. 994.

[50] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *International conference on pervasive computing*, Springer, 2004, pp. 1–17.

[51] F. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[52] M. S. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori, "A hierarchical deep temporal model for group activity recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1971–1980.

[53] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson, "Mining models of human activities from the web," in *Proceedings of the 13th international conference on World Wide Web*, ACM, 2004, pp. 573–582.

[54] D. Wyatt, M. Philipose, and T. Choudhury, "Unsupervised activity recognition using automatically mined common sense," in *AAAI*, vol. 5, 2005, pp. 21–27.

[55] H. A. Kautz, "A formal theory of plan recognition and its implementation," *Reasoning about plans*, pp. 69–125, 1991.

[56] L. Chen, C. D. Nugent, M. Mulvenna, D. Finlay, X. Hong, and M. Poland, "A logical framework for behaviour reasoning and assistance in a smart home," *International Journal of Assistive Robotics and Mechatronics*, vol. 9, no. 4, pp. 20–34, 2008.

[57] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 961–974, 2011.

[58] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, ACM, 2013, pp. 27–38.

[59] F. Adib and D. Katabi, *See through walls with WiFi!* 4. ACM, 2013, vol. 43.

[60] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14)*, 2014, pp. 317–329.

[61] M. Vacher, B. Lecouteux, J. S. Romero, M. Ajili, F. Portet, and S. Rossato, "Speech and speaker recognition for home automation: Preliminary results," in *Speech Technology and Human-Computer Dialogue (SpeD), 2015 International Conference on*, IEEE, 2015, pp. 1–10.

[62] A. Jalal, M. Z. Uddin, and T.-S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, 2012.

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

[64] S. A. Ahson and M. Ilyas, *RFID handbook: applications, technology, security, and privacy.* CRC press, 2017.

[65] H. Li, C. Ye, and A. P. Sample, "Idsense: A human object interaction detection system based on passive uhf rfid," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2015, pp. 2555–2564.

[66] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, ACM, 2014, pp. 237–248.

[67] S. Pradhan, E. Chai, K. Sundaresan, L. Qiu, M. A. Khojastepour, and S. Rangarajan, "Rio: A pervasive rfid-based touch gesture interface," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, ACM, 2017, pp. 261–274.

[68] S. Feng, P. Setoodeh, and S. Haykin, "Smart home: Cognitive interactive people-centric internet of things," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 34–39, 2017.

[69] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31 314–31 338, 2015.

[70] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia*, Acm, 2015, pp. 1307–1310.

[71] M. Ziaeefard and R. Bergevin, "Semantic human activity recognition: A literature review," *Pattern Recognition*, vol. 48, no. 8, pp. 2329–2345, 2015.

[72] T. Gu, S. Chen, X. Tao, and J. Lu, "An unsupervised approach to activity recognition and segmentation based on object-use fingerprints," *Data & Knowledge Engineering*, vol. 69, no. 6, pp. 533–544, 2010.

[73] P. Palmes, H. K. Pung, T. Gu, W. Xue, and S. Chen, “Object relevance weight pattern mining for activity recognition and segmentation,” *Pervasive and Mobile Computing*, vol. 6, no. 1, pp. 43–57, 2010.

[74] X. Liu, J. Cao, Y. Yang, W. Qu, X. Zhao, K. Li, and D. Yao, “Fast rfid sensory data collection: Trade-off between computation and communication costs,” *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1179–1191, 2019.

[75] X. Liu, X. Xie, X. Zhao, K. Wang, K. Li, A. X. Liu, S. Guo, and J. Wu, “Fast identification of blocked rfid tags,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 9, pp. 2041–2054, 2018.

[76] X. Liu, X. Xie, S. Wang, J. Liu, D. Yao, J. Cao, and K. Li, “Efficient range queries for large-scale sensor-augmented rfid systems,” *IEEE/ACM Transactions on Networking*, 2019 in press.

[77] B. Fazzinga, S. Flesca, F. Furfaro, and F. Parisi, “Offline of rfid trajectory data,” in *Proceedings of the 26th International Conference on Scientific and Statistical Database Management*, ACM, 2014, p. 5.

[78] ——, “Cleaning trajectory data of rfid-monitored objects through conditioning under integrity constraints,” in *EDBT*, 2014, pp. 379–390.

[79] ——, “Exploiting integrity constraints for cleaning trajectories of rfid-monitored objects,” *ACM Transactions on Database Systems (TODS)*, vol. 41, no. 4, p. 24, 2016.

[80] M. Philipose, “Large-scale human activity recognition using ultra-dense sensing,” *The Bridge, National Academy of Engineering*, vol. 35, no. 4, 2005.

[81] X. Li, D. Yao, X. Pan, J. Johannaman, J. Yang, R. Webman, A. Sarcevic, I. Marsic, and R. S. Burd, “Activity recognition for medical teamwork based on passive rfid,” in *2016 IEEE International Conference on RFID (RFID)*, IEEE, 2016, pp. 1–9.

[82] S. A. Vora, W. M. Mongan, E. K. Anday, K. R. Dandekar, G. Dion, A. K. Fontecchio, and T. P. Kurzweg, "On implementing an unconventional infant vital signs monitor with passive rfid tags," in *2017 IEEE International Conference on RFID (RFID)*, IEEE, 2017, pp. 47–53.

[83] Y. Du, Y. Lim, and Y. Tan, "Activity recognition using rfid phase profiling in smart library," *IEICE Transactions on Information and Systems*, vol. 102, no. 4, pp. 768–776, 2019.

[84] Y. Du, Y. Tan, and Y. Lim, "Rf-switch: A novel wireless controller in smart home," in *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, IEEE, 2018, pp. 1–2.

[85] Y. Du, Z. Li, M. Stojmenovic, W. Qu, and H. Qi, "A low overhead progressive transmission for visual descriptor based on image saliency.," *Journal of Multiple-Valued Logic & Soft Computing*, vol. 25, no. 1, 2015.

[86] Y. Du, Z. Li, W. Qu, S. Miao, S. Wang, and H. Qi, "Mvss: Mobile visual search based on saliency," in *2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing*, IEEE, 2013, pp. 922–928.

[87] A. Rajaraman and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press, 2011.

[88] Q. Liu, S. Wu, L. Wang, and T. Tan, "Predicting the next location: A recurrent model with spatial and temporal contexts.," in *AAAI*, 2016, pp. 194–200.

[89] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[90] F. Ordóñez, P. de Toledo, A. Sanchis, *et al.*, "Activity recognition using hybrid generative/discriminative models on home environments using binary sensors," *Sensors*, vol. 13, no. 5, pp. 5460–5477, 2013.

[91] X. Liu, X. Xie, K. Li, B. Xiao, J. Wu, H. Qi, and D. Lu, "Fast tracking the population of key tags in large-scale anonymous rfid systems," *IEEE/ACM Transactions on Networking (TON)*, vol. 25, no. 1, pp. 278–291, 2017.

[92] B.-W. Min, "Next-generation library information service-'smart library'," *Environment*, vol. 9, no. 11, 2012.

[93] L. Toivanen, M. Heino, A. Oksman, A. Vienamo, J. Holopainen, and V. Viikari, "Rfid-based book finder [education corner]," *IEEE Antennas and Propagation Magazine*, vol. 58, no. 3, pp. 72–80, 2016.

[94] I. Markakis, T. Samaras, A. C. Polycarpou, and J. N. Sahalos, "An rfid-enabled library management system using low-sar smart bookshelves," in *Electromagnetics in Advanced Applications (ICEAA), 2013 International Conference on*, IEEE, 2013, pp. 227–230.

[95] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu, "Relative localization of rfid tags using spatial-temporal phase profiling.," in *NSDI*, 2015, pp. 251–263.

[96] J. Liu, F. Zhu, Y. Wang, X. Wang, Q. Pan, and L. Chen, "Rf-scanner: Shelf scanning with robot-assisted rfid systems," in *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*, IEEE, 2017, pp. 1–9.

[97] *Speedway revolution reader-low level user data support*, Revision 3.0, ImpinJ, Inc, 2013.

[98] *Impinj*, https://www.impinj.com/.

[99] T. Wei and X. Zhang, "Gyro in the air: Tracking 3d orientation of batteryless internet-of-things," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, ACM, 2016, pp. 55–68.

[100] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "Lifs: Low human-effort, device-free localization with fine-grained subcarrier

information," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, ACM, 2016, pp. 243–256.

[101] A. F. Molisch, *Wireless Communications.* USA: Wiley, 2012.

[102] S. Parlak, I. Marsic, A. Sarcevic, W. U. Bajwa, L. J. Waterhouse, and R. S. Burd, "Passive rfid for object and use detection during trauma resuscitation," *IEEE Transactions on Mobile Computing*, vol. 15, no. 4, pp. 924–937, 2016.

[103] J. Han, H. Ding, C. Qian, W. Xi, Z. Wang, Z. Jiang, L. Shangguan, and J. Zhao, "Cbid: A customer behavior identification system using passive tags," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2885–2898, 2016.

[104] *Speedway revolution quick start guide*, Revision 2.3, ImpinJ, Inc, 2013.

[105] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, and R. S. Burd, "Deep learning for rfid-based activity recognition," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, ACM, 2016, pp. 164–175.

[106] K. Itoh, "Analysis of the phase unwrapping algorithm," *Applied optics*, vol. 21, no. 14, pp. 2470–2470, 1982.

Publications

Journal

[1] Y. Du, Y. Lim, and Y. Tan, "A Novel Human Activity Recognition and Prediction in Smart Home Based on Interaction," *Sensors*, Vol. 19(20), pp. 4474 - 4490, Oct. 2019

[2] Y. Du, Y. Lim, and Y. Tan, "Activity Recognition Using RFID Phase Profiling in Smart Library," *IEICE Transaction on Information and Systems*, Vol. 4(102), pp. 768 - 776, Apr. 2019

International Conference

[3] Y. Du, Y. Lim, and Y. Tan, "RF-ARP: RFID-based Activity Recognition and Prediction in Smart Home," *25th IEEE International Conference on Parallel and Distributed Systems (ICPADS 2019)*, Tianjin, China, 2019, pp. xx - xx

[4] Y. Du, Y. Lim, and Y. Tan, "Activity Prediction Using LSTM in Smart Home," *8th IEEE Global Conference on Consumer Electronics (GCCE 2019)*, Osaka, Japan, 2019, pp. xx - xx

[5] Y. Du, Y. Lim, and Y. Tan, "Reading Activity Recognition in Smart RF-Library," *7th IEEE Global Conference on Consumer Electronics (GCCE 2018)*, Nara, Japan, 2018, pp. 447 - 450

[6] Y. Du, Y. Tan, and Y. Lim, "RF-Switch: A Novel Wireless Controller in Smart Home," *5th IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW 2018)*, Taichung, Taiwan, 2018, pp. 1 - 2

Domestic Conference

[7] Y. Du, Y. Lim, and Y. Tan, "Activity Prediction Using LSTM in Smart Home," *IEICE General Conference 2019*, Tokyo, Japan, BS-4-40, Mar. 2019

[8] Y. Du, V. C. Pham, Y. Tan, and Y. Lim, "A Solution of Activity Recognition in Smart Home Using Passive RFID Tags," *IEICE General Conference 2018*, Tokyo, Japan, AS-1-3, Mar. 2018