An Accurate and Robust Indoor Localization System using Deep Learning and Passive Infrared Sensors

Kan Ngamakeur

A thesis submitted to Auckland University of Technology in fulfilment of the requirements for the degree of Doctor of Philosophy

Primary Supervisor: Dr. Sira Yonchareon

Secondary Supervisor: Assoc. Prof. Jian Yu

July 2022

School of Engineering, Computer and Mathematical Science

Abstract

Internet of Things (IoT) has evolved significantly over the past decade, enabling a wide range of applications for indoor environments in different fields such as smart building, healthcare, and energy management. A person's location is crucial information in providing their core services. Thus, indoor localization is a fundamental component of these applications. Different technologies are available for indoor localization. However, video camera technology and wearable devices are not always practical in all situations. A video camera may cause privacy issues. People may not cooperate to hold or wear the devices because they forget or feel uncomfortable. As a result, Passive Infrared (PIR) sensors are employed for indoor localization to mitigate these issues. The advantages of PIR sensors include their low power consumption, cost-effectiveness, and low electromagnetic interference. In addition, their analog or voltage outputs can provide fine-grained information regarding a person's location such as phase and amplitude. Thus, it is possible to leverage analog outputs to estimate a person's location. However, utilizing PIR analog outputs is not simple due to their ambiguity and a lack of PIR sensor specification.

This research focuses on addressing the challenges of indoor localization based on PIR analog signal including the design of a PIR sensor node, single person localization, and multi-person localization. Firstly, this thesis proposes a novel design of a PIR sensor node and a prototype is developed to collect data. Secondly, this thesis proposes a deep learning framework for a single person localization. The proposed CNN-LSTM model not only extracts features from PIR analog output automatically but also learn temporal dependencies between the extracted features. Lastly, the proposed localization framework is extended to support multi-person localization. This thesis proposes a channel separation method to generate inputs for each person. Then, deep CNN-LSTM estimates a location for each person and a mean bagging method is used to integrate multiple CNN-LSTM models for improving the accuracy of multi-person localization. A set of experiments are conducted, and the proposed localization methods can achieve good results for both single person and multi-person localization.

Keywords: Indoor localization, passive infrared, deep learning, PIR analog

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments, nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature:

Date: 13/July/2022

Co-authored Work

All co-authors in the following table have approved these chapters for inclusion in Kan Ngamakeur 's doctoral thesis.

Chapter	**Author %**
Chapter 2: A Survey on Device-Free Indoor Localization and Tracking in the Multi-Resident Environment Manuscript published in ACM Computing Survey Journal.	KN = 80 SY = 10 JY = 5 SR = 5
Chapter 3: A Passive Infrared Sensors System for Device-free Indoor Localization and Tracking Manuscript submitted to IEEE Sensors Journal.	KN = 85 SY = 10 JY = 5
Chapter 4: Deep CNN-LSTM Network for Indoor Location Estimation using Analog Signals of Passive Infrared Sensors Manuscript submitted to IEEE Internet of Things Journal.	KN = 85 SY = 10 JY = 5
Chapter 5: Multi-person Localization using Analog Signals of Passive Infrared Sensors Manuscript submitted to IEEE Internet of Things Journal.	KN = 80 SY = 10 JY = 5
Kan Ngamakeur (KN), Sira Yongchareon (SY), Saeed Ur Rehman (SR), Jian Yu (JY)	

We, the undersigned, hereby agree to the percentages of participation to the chapters identified above:

Sira Yongchareon Jian Yu Saeed Ur Rehman Kan Ngamakeur

Acknowledgement

I would like to express my sincere gratitude to all people who assisted me during the writing of this thesis. My deepest gratitude goes first and foremost to Dr. Sira Yongchareon, my supervisor, for his steady guidance and inspiration. Over the past years of my study, he has given me many useful suggestions for writing and improving each chapter of this thesis. Without his enthusiastic contribution and support, this thesis could not have reached its present form.

Second, I would like to express my gratitude to Associate Professor Jian Yu for his contribution in this thesis. For the past year, he has kindly offered his knowledgeable recommendations and help in improving my works in this thesis.

Third, I would like to show my gratitude to other co-authors including Prof. Quan Z Sheng and Dr. Saeed Ur Rehman for giving me useful comments and helping me revise my works.

Last but not least, I am so grateful to my fiancée, jutharat phienpattanavit, and her family for their emotional support and useful advice to motivate me to put through my stressful time.

Finally, my thanks would go to My father, Dr. Somsak Ngamakeur, and my beloved family for their great patience and encouragement. With their emotional support, it gave me great confidence to work tirelessly through all these years.

Kan Ngamakeur

13/July/2022

Chapter 1 Introduction

This chapter provides a comprehensive introduction to this thesis. There are four sections in total. The first section describes the background of device-free indoor localization and the motivation of PIR-based localization for multi-person. Research questions are presented and discussed in the second section. The third section highlights the contributions of this thesis. The fourth section presents the overall structure of this thesis.

1.1 Background and Motivation

Localization has gained interest from researchers because of a multitude of useful applications that can emerge from this field of study. For the past decades, Outdoor localization has been studied extensively and has reached its maturity, such as GPS technology. On contrary, Indoor localization is still in its infancy and has no such standard technology and protocol to realize this concept effectively and efficiently[1]. The rapid improvement of the Internet of Things (IoT) and sensor technology enables devices and sensors to communicate and exchange data with each other over the internet or other types of networks[2, 3]. This enhancement opens new possibilities for developing an indoor localization system that is robust and cost-effective due to the utilization of existing hardware and infrastructure.

Indoor localization can be organized into two major categories including device-based and device-free localization. For device-based localization, a person must carry a mobile or wearable device that sends its position from time to time. As a result, the location of a person can be referred to from the location of the device. Various mobile and wearable technologies can support device-based localization and range from expensive to cheap ones such as mobile phones, smartwatches, RFID bracelets, etc. Although device-based solutions obtain promising results in many applications, it may not be feasible in some cases[4]. A person does not want to carry or wear a mobile or wearable device because it is cumbersome and causes a person to feel intrusive[5]. Elderly and dementia patients are likely to forget to wear the devices[6, 7]. Thus, losing track of these people may endanger their lives. In addition, there is a risk of losing or damaging the devices[1, 8]. Some mobile and wearable devices need to have frequent battery maintenance. A person may need to remove them for charging the devices. Thus, the monitoring of a person needs to stop, and this could be a setback if constant monitoring is required. Device-free localization can locate people without the need to carry any device. A person only needs to interact with deployed sensors in a monitored area and sensor readings are analyzed to determine his location. It is a key component that empowers many applications in the field of ambient-assisted living (AAL) and smart building such as fall detection, remote health monitoring, building energy management, intrusion detection, etc. [3]

The passive technologies that can support device-free indoor localization are radiofrequency

(RF), infrared, acoustic and vibration, and electric field. The RF-based localization relies on signal phenomena that occur when a person obstructs a radio link between an RF transmitter and receiver, such as absorption, reflection, scattering, and diffraction to estimate a location of a person[9]. The major advantage of RF localization is that it can operate through a wall[10]. Received signal strength (RSS) measurement is a standard metric for most RF technologies such as RFID, Wi-Fi, and Zigbee/IEEE 802.15.4. Although RSS is quite easy to obtain, it cannot provide accurate results. To achieve higher accuracy, a dense deployment of RF nodes is required[11]. As a result, this increases hardware cost and energy consumption significantly. Unlike RSS, channel state information (CSI) achieves better results due to its fine-grained features, but it is not available in commodity WiFi hardware[12]. Thus, it restricts the use of pre-existing infrastructure. mmWave can localize a person with a distance error of 0.1 meters but an acquisition cost is remarkably high. Electric field-based or capacitive sensing localization requires a person to stay nearby or interact directly with capacitive sensing devices such as capacitive floor mats[13]. However, these devices may not be practical to deploy in a fully furnished area and have a high installation cost due to a dense deployment[14]. Acoustic and vibration localization employs microphone or seismic sensors to measure Time on Arrival (TOA) or Time difference of arrival (TDOA) between a person and sensors[15]. Then, triangulation methods are employed to estimate a person's position. The accuracy of acoustic and vibration can be affected by different floor structures, obstructions, and shoe types[16]. A video camera is not a part of the classification because it can violate the privacy rights of inhabitants.

Passive infrared (PIR) sensors are commonly used to detect a movement of a person for the purpose of security alarm, automatic lighting, etc. Typical PIR sensors consist of two pyroelectric elements, one positive and one negative. These elements can convert heat radiated from a source into electricity[17]. Thus, when a person moves across the FOV of a PIR sensor, it results in an output signal consisting of positive and negative pulses. Then, the output can further convert into a binary format or remain analog. Commonly, a PIR sensor is equipped with a Fresnel lens that focuses heat radiation on the sensing elements. Different designs of Fresnel lenses allow various adjustments to the number of detection zones, range, and angle of coverage. Thus, it is possible to leverage PIR sensors for indoor localization. Compared with RF technology, PIR sensors are much cheaper and consume less energy. More importantly, it is not affected by electromagnetic

(EM) interference. PIR sensors can provide more coverage distance than capacitive sensing devices and are easier to deploy. In addition, they can operate at any illumination condition, unlike video camera-based solutions. Thus, a PIR sensor is an appropriate option for deploying in an indoor environment such as a residential home.

This thesis categorizes PIR-based localization into two classes including binary-based and analog-based. Some binary-based solutions require a sophisticated design of a PIR sensor node[18-20]. As a result, their developed sensor nodes are large and bulky. Other binary works deploy a massive number of PIR sensors in a monitored area[21-23]. Thus, these complex settings are not feasible in real-life conditions. An analog output of the PIR sensor has gained more attraction from researchers recently because it can provide fine-grained features, unlike a binary output[24]. As a result, it is possible to leverage these features to provide an accurate localization while reducing the complexity of the sensor node and the number of deployed sensors. So far, only a small number of works have tackled analog-based localization. However, they just use the analog output to determine a coarse location[24, 25]. There is room for improvements of localizing with analog output of PIR sensors such as accuracy and multi-person handling, but there is a major issue that is a lack of detailed specifications of commercial off-the-shelf (COTS) PIR sensors. Therefore, it is quite hard to adopt standard localization algorithms such as Kalman filtering and Particle filtering. To be specific, there is not enough data to formulate an observation model.

Over the past years, deep learning approaches has been applied successfully in many applications such as image processing, speech recognition, fraud detection etc.[26] The major advantage of deep learning is that it can extract features automatically from raw data[27]. Thus, the above information issue can be resolved using deep learning. Despite its popularity and benefits, the adoption of deep learning for PIR-based localization is still relatively low[28, 29]. It opens up new opportunities to leverage the deep learning approach to improve the performance of PIR-based localization. Thus, this thesis aims to address various challenges in enabling accurate multi-person localization using PIR analog sensors in a deep learning framework.

1.2 Problems/Research Questions

Over the past years, device-free indoor localization has been realized and become an essential

part of applications in many areas such as healthcare, security, and smart building due to the rapid improvement of sensor technology. In addition, the recent advance in the deep learning approach has brought signification improvement to many areas of application such as image process and speech recognition. Similarly, device-free localization can leverage the power of the deep learning approach to improve the results of localizations. PIR sensor is one of the promising solutions for indoor localization due to its cost efficiency and robustness against EM interference and illumination condition. However, the device-free localization using PIR sensors is quite challenging. Thus, extensive studies need to be conducted to realize and evaluate the use of PIR sensors for both single and multiple-person localization. This thesis investigates three questions of PIR-based localization, which are listed below.

Research Question (1): What PIR sensor design and its placement achieve the best overall result for indoor localization?

PIR sensor has been employed in existing works to develop the device-free indoor localization system for the past years[24, 30-32]. However, there is no available public PIR datasets from these works. The lack of a PIR dataset is a major challenge for the progression of this thesis. Thus, a PIR sensor node needs to be developed. This requirement raises the question on what sensor design and sensor placement will be suitable for indoor localization. To answer this question, this thesis studies existing PIR sensor designs and their placements to identify their advantages and drawbacks. Then, a PIR sensor will be designed and developed. Finally, data collection will be performed, and various localization methods will be employed to evaluate the feasibility of the developed sensor node.

Research Question (2): By what means can a deep learning approach be used for PIR-based localization of a single person, and to what degree can the proposed approach achieve in the single person localization and what factors can affect its performance?

Over the past years, the deep learning approach has been employed in many fields, such as image processing, speech recognition and natural language processing, etc. Compared with traditional machine learning, deep learning not only can extract features from raw data without the intervention of human experts, but it also delivers outstanding performance. However, the adoption of the deep learning approach is still in its infancy stage for the PIR-based localization.

Thus, the usage of the deep learning approach needs to be further explored. To answer this question, this thesis studies related works in similar area and other areas such as ECG signal analysis and speech recognition to identify a possible design of deep learning architecture to perform the localization of a single person using PIR analog outputs. Then, the chosen architecture of deep learning will be tested with a set of walking scenarios to determine the overall performance of the proposed architecture and factors that can affect its performance.

Research Question (3): By what means can the proposed deep learning approach be used for PIR-based localization of multi-person, and to what degree can it achieve in the multi-person localization and what factors can affect its performance?

Compared with a single person localization, the multi-person localization is quite challenging because ambiguous or unexpected information in a sensor reading is introduced by the presence of multiple persons. As a result, existing localization techniques perform poorly in this situation, and the deep learning approach is no exception. In addition, the traditional machine learning and deep learning approach require data collection to be performed for every person. Thus, this process is quite cumbersome and time-consuming. This issue restricts the adoption of the deep learning method in real life. In this thesis, the PIR-based multi-person localization will be investigated. The related works will be analyzed to identify a possible technique that can improve the performance of a deep learning method for multi-person localization. A set of multi-person walking scenarios will be conducted to evaluate the performance of the chosen method and the factors that affect its performance will be analyzed.

1.3 Contributions

This thesis proposes a PIR-based localization system that consists of a PIR sensor node and deep learning-based localization methods. The contribution of this thesis is listed below:

Contribution 1: A comprehensive literature review is presented for device-free indoor localization and tracking. First, preliminary knowledge of the device-free indoor localization and tracking is provided. Then, the taxonomy and classification of related works are presented.

A detailed analysis of localization and tracking techniques is presented as well as a comparison of related works. Finally, the research challenges and future directions are discussed. This

contribution is reported in Chapter 2.

Contribution 2: For the second contribution, existing PIR sensor designs, and their placements are analyzed to compare their benefits and drawbacks. Then, a novel design of ceiling-based PIR node is presented and the prototype of the proposed design is developed as a proof-of-concept. The developed sensor node is easy to install and provides good coverage. To validate the feasibility of the proposed sensor node, A set of experiments consisting of different walking scenarios are conducted and the standard localization techniques including deep learning and particle filter are employed to evaluate the outcomes of the experiments. The results show that our proposed sensor node can achieve similar or better results compared with the existing works in terms of accuracy and coverage distance. Finally, the PIR dataset is also provided in a public data repository for researchers to develop and evaluate their new localization techniques. This contribution is reported in Chapter 3.

Contribution 3: In this contribution, this thesis proposes a deep learning framework for single person indoor localization. The framework consists of two phases including offline and online phases. In the offline phase, raw PIR outputs are collected and used as a training dataset. The deep CNN-LSTM architecture is proposed to estimate the location of a person. The upper CNN level allows an automatic feature extraction while the lower LSTM level can learn the temporal relationship of extracted features. In the online phase. the trained CNN-LSTM model is used to estimate the location of a person. Unlike the offline training, the input signals are sent in a shorter sequence from time to time. To handle a shorter sequence appropriately, the hidden and cell states of the LSTM network are updated to maintain the time correlation of input signals in two consecutive time steps. Thus, the proposed model can provide an accurate localization result even though input signals of the next timestep are ambiguous. An extensive experiment is conducted to evaluate the proposed method. The results show that the mean distance error of 0.23m is achieved and 80% of distance errors are within 0.4 meters which is better than existing works[22, 32, 33]. This contribution is reported in Chapter 4.

Contribution 4: In this contribution, the proposed CNN-LSTM architecture from RQ 2 is extended to support indoor multi-person localization. Inspired by Yang, et al. [28], the two-step method is proposed to localize multiple persons. In the first step, the channel separation method can utilize an existing training dataset to separate raw PIR outputs generated by multiple persons

to generate individual inputs according to the number of persons. In the second step, each person's location is estimated by the CNN-LSTM model. The ambiguity and unexpected patterns from the presence of multiple persons are handled appropriately by integrating multiple CNN-LSTM models through the mean bagging technique[34]. A comprehensive set of experiments is conducted to evaluate the performance of the proposed method and to identify factors that affect multi-person localization. The results show that the proposed method can estimate the locations of two participants simultaneously with the mean distance error of 0.53 meters and 80% of distance errors are within 0.8 meters. The proposed method achieves similar or more accurate results than the existing works[32, 35, 36]. This contribution is reported in Chapter 5.

1.4 Thesis Structure

According to the research questions, this thesis is organized into 6 chapters and the details of each chapter are described in the following paragraphs:

Chapter 2 provides an overview of device-free indoor localization and tracking in the multi-resident environment. Section 2.2 introduces the fundamentals of device-free indoor localization and tracking. Section 2.3 describes the technologies that can support the localization and tracking in the multi-resident setting. Section 2.4 describes and discusses the localization and tracking techniques for the multi-occupancy setting. Section 2.5 describes the performance metrics. Section 2.6 discusses the open challenges and highlights the direction of future research. Finally, we conclude our literature review in Section 2.7.

Chapter 3 focuses on data collection for indoor localization and tracking using a PIR sensor node. Section 3.2 discusses our related works. Section 3.3 describes the data collection process and characteristics of our dataset. Section 3.4 presents a procedure and methods to evaluate and validate our dataset. Section 3.5 shows and discusses the results of the experiments. Section 3.6 describes the summary of our data collection.

Chapter 4 tackles PIR-based indoor localization for a single person using the use of deep learning models. Section 4.2 reviews the related works. Section 4.3 describes the problem formulation of indoor localization and tracking. The proposed methodology presented in Section 4.4 presents a proposed deep CNN-LSTM framework for PIR-based localization and tracking.

The experiment setup is described in Section 4.5. Section 4.6 reports and discusses the results of the experiments. Finally, Section 4.7 provides a conclusion of this chapter.

Chapter 5 focuses on multi-person localization using PIR analog signals. Section 5.2 discusses related works. Section 5.3 presents a proposed methodology for multi-person localization. Sections 5.4 describes the experiment setup including hardware and data collection. Section 5.5 discusses the results of multi-person experiments. Finally, the summary of chapter 5 is presented in Section 5.6.

Chapter 6 presents our conclusions. Sections 6.1 summarizes the contributions of this thesis. Section 6.2 discusses the limitations of the proposed methods in this thesis. Lastly, Section 6.3 highlights future works.

Chapter 2 A Survey on Device-Free Indoor Localization and Tracking in the Multi-Resident Environment

Indoor device-free localization and tracking can bring both convenience and privacy to users compared with traditional solutions such as camera-based surveillance and RFID tag-based tracking. Technologies such as WI-FI, wireless sensor, and infrared have been used to localize and track people living in care homes and office buildings. However, the presence of multiple residents introduces further challenges, such as the ambiguity in sensor measurements and target identity, to localization and tracking. In this chapter, we survey the latest development of device-free indoor localization and tracking in the multi-resident environment. We first present the fundamentals of device-free localization and tracking. Then we discuss and compare the technologies used in device-free indoor localization and tracking. After discussing the steps involved in multi-resident localization and tracking including target detection, target counting, target identification, localization and tracking, the techniques related to each step are classified and discussed in detail along with the performance metrics. Finally, we identify the research gap and point out future research directions. To the best of our knowledge, this survey is the most comprehensive work that covers a wide spectrum of the research area of device-free indoor localization and tracking.

Chapter 2 has been published by ACM computing survey Journal

Kan Ngamakeur, Sira Yongchareon, Jian Yu, and Saeed Ur Rehman. 2020. A Survey on Device-free Indoor Localization and Tracking in the Multi-resident Environment. ACM Comput. Surv. 53, 4, Article 71 (September 2020), 29 pages. DOI:https://doi.org/10.1145/3396302

2.1 Introduction

Indoor localization and tracking aims to estimate the current location of a person using a suitable device and predict this person's next locations from time to time in indoor areas such as residential homes and office buildings. This area of research has become crucial for applications in healthcare [37], energy management [13, 21], and security management [38]. For example, a monitored patient or elderly can receive prompt assistance in emergencies such as falling, parents can be alerted if their children enter potentially hazardous areas, rescuers can be informed with whereabouts of rescuees in a building to formulate appropriate rescue plans. Several indoor localization and tracking approaches and systems have been proposed in recent years. Some solutions require occupants to wear or carry localization and tracking devices such as radio tags and mobile phones [39-41]. However, it is not always convenient for occupants to carry tracking devices in residential homes or offices and there is always a chance that the devices are misplaced. [37, 42]. On the other hand, device-free localization and tracking solutions can overcome these drawbacks and there are various types of technologies that can be used such as radio frequency (RF) [43, 44], surveillance cameras [45], infrared [17, 46, 47], acoustic [48] and electric field [14].

Among the device-free approaches, the surveillance camera-based solutions have been used extensively [45, 46]. However, camera surveillance may not be appropriate for private spaces such as offices or homes. On the other hand, non-intrusive sensing technologies such as passive infrared and radiofrequency can sense the presence of a person without revealing his identity or appearance and thus are more appropriate for privacy-sensitive settings.

Among the existing surveys on device-free indoor localization, Palipana, et al. [9] provided a detailed review of device-free localization based on the RF technology. Liu, et al. [49] proposed a classification of the wireless localization and tracking techniques. Xiao, et al. [4] classified indoor localization from the device perspective. Different from these works, we also provide a comprehensive review on the technical aspect, which includes the methods and algorithms, of the selected device-free indoor (DFI) localization and tracking approaches with a special focus on approaches for the complex multi-resident environment. In recent works, Gu, et al. [50] focused only on techniques to improve the localization based on spatial information. Khelifi, et al. [51]

provided a general overview of localization systems and their applications in IoT environment. Whereas we provide an in-depth review of techniques for localizing and tracking multiple persons.

The main contributions of this work are as follows:

1) We provide the complete picture of DFI localization and tracking as well as the taxonomy and classification of related DFI localization and tracking techniques.
2) We present the detailed steps of the DFI localization and tracking process and discussed and evaluated the corresponding techniques including those for the multi-resident environment.
3) We identify the research gaps and potential future directions for DFI localization and tracking.

The rest of the paper is organized as follows. In Section 2.2, we introduce the fundamentals of device-free indoor localization and tracking. In Section 2.3, we review the technologies that can support the localization and tracking in the multi-resident setting. Section 2.4 discusses the localization and tracking techniques for the multi-occupancy setting. Section 2.5 discusses the performance metrics. In Section 2.6, we discuss the open challenges and highlight the direction of future research. Finally, we conclude this chapter in Section 2.7.

2.2 An Overview of Device-Free Indoor Multi-resident Localization and Tracking

In this section, we first introduce the concept of device-free indoor localization and tracking and the challenges related to its application in the multi-resident environment and then we present a taxonomy for device-free indoor location and tracking.

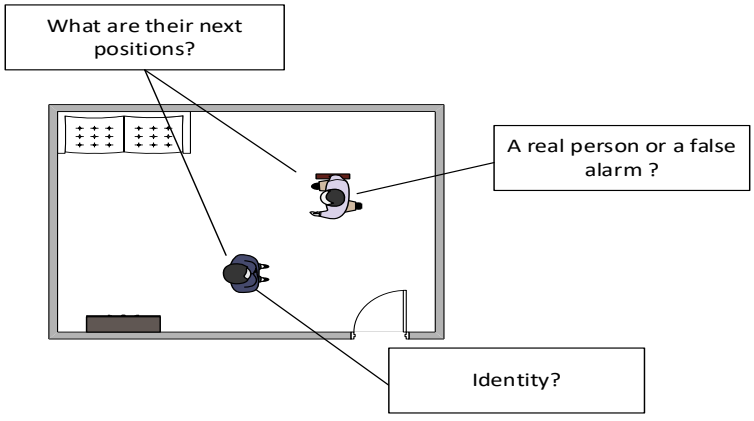


Figure 2-1 Device-free indoor localization and tracking for multi-resident

Some pioneer works such as [17], [52] and [43] have demonstrated the feasibility of non-camera-based indoor localization and tracking that does not require the occupants to carry tracking devices. Reseach in DFI localization and tracking has accumulated much interest in recent years. Since 2006, the number of yearly publications on this topic is steadily on the increase. Some early works focused on single-resident localization and tracking while recent works put attentions to the more realistic multi-resident scenarios. A multi-resident scenario introduces additional challenges to localization and tracking as shown in Figure 2-1. First, any person may enter a monitored area , move to anywhere in this area, and leave the area at any time. As a result, multiple residents may trigger multiple sensors at the same time. Also, the coexistence of multiple residents in the same monitored area can produce several complex localization and tracking situations such as one person walking away from a group, and two persons walking towards and passing each other. Such concurrent events make accurate localization and tracking difficult. Second, free-of-device means the identity information of residents is not available, which makes it difficult to maintain the tracks of a specific resident. A general solution/procedure for multi-resident localization and tracking is given in Figure 2-2, where various *technologies* and *techniques* are utilized, together with relevant *measurements* and *evaluation metrics*.

Figure 2-2 also gives a classification of techniques. It is worth noting that the rest of this chapter is organized based on this procedure.

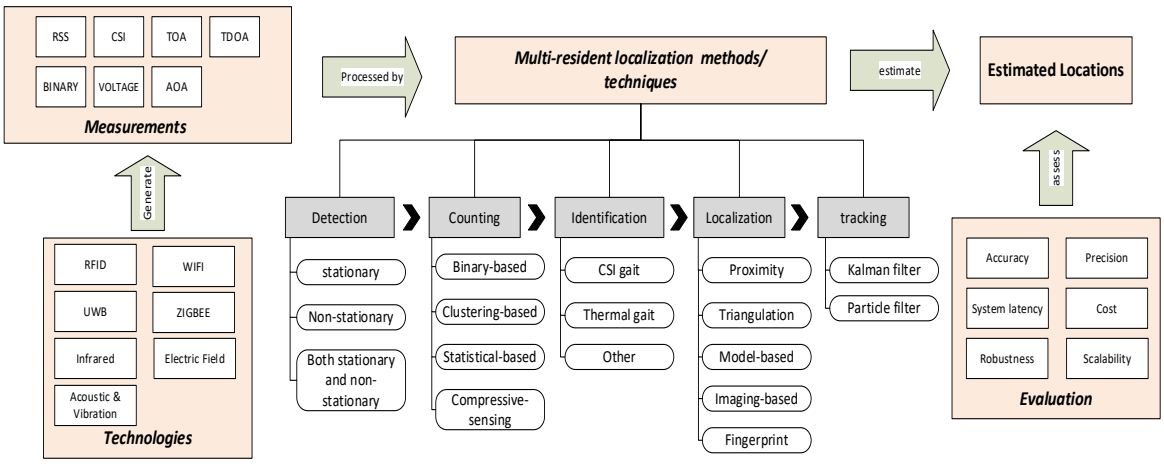


Figure 2-2 Technologies and techniques for multi-resident localization and tracking

To preserve the privacy of persons, passive sensing technologies are adopted including RFID [53], WI-FI [54], UWB [55], Zigbee [43], Infrared [56], Electric fields [14], and Acoustic and Vibration [57]. These technologies produce a set of measurements as shown in

Figure 2-2. The details of the measurements are not included here because they are well known and have been already discussed extensively. When deployed sensors detect changes in measurements caused by persons entering in a monitored area, the human detection step is performed to initiate localization and tracking. In this step, there are three classes for presence detection: *non-stationary, stationary, and both stationary and non-stationary*. Next, the number of persons is counted and maintained to ensure that there is no missing person. Human counting techniques are categorized into *binary-based, clustering-based, statistical-based, and compressive sensing-based counting*. Then, each person is identified and this information to keep track of each person. The types of identification approaches are classified as *CSI gait-based, Thermal gait-based,* and *other approaches.* Finally, the estimated locations of persons are determined by localization and tracking techniques. Localization techniques and tracking techniques can be applied together or interchangeably to estimate location of persons. The categories of localization technique are: *proximity, triangulation, model-based, imaging-based and RF fingerprinting-based technique*. For tracking techniques, there are two groups of technique including: *kalman filter* and *particle filer*.

2.3 Technologies

In this section, we describe technologies that can be used for DFI localization and tracking. These technologies include *radio frequency*, *infrared*, *acoustic and vibration*, and *electric field.* We also compare them to gauge their capacity in supporting localization in the multi-resident setting.

2.3.1 Radio Frequency

Radio Frequency (RF) [49] refers to the group of technologies that relies on radio waves or electromagnetic signals to transmit information. RF technologies that facilitate device-free indoor localization and tracking include RFID, Wi-FI, Wireless sensor network, and UWB.

2.3.1.1 Radio Frequency Identification (RFID)

RFID is a type of radiofrequency technology that uses RF-compatible integrated circuits to store and retrieve data sent through electromagnetic transmission [49]. The core elements of an RFID system include *RFID reader* and *RFID tags*. There are two types of RFID tags: *passive* and *active*. Passive tags are generally cheap, small and require no battery. Active tags need batteries, but it has better coverage than passive ones.

In the literature, Liu, et al. [58] proposed a fault-tolerant sequential pattern mining to extract frequency patterns from noisy data generated from an array of RFID tags. Ruan, et al. [59], [60] attached several passive RFID tags on the wall and divided monitor areas to small grids with each grid representing a user location. Ma, et al. [61] deployed two arrays of RFID tags orthogonally on walls to locate x and y axes of the target's location separately and then applied particle swarm optimization (PSO) to localize the target.

2.3.1.2 Wi-Fi

WI-FI is a mid-range RF technology that is commonly used to create a local area network [49]. Nowadays many buildings are installed with Wi-Fi infrastructure, which makes it possible to use existing infrastructure for localization and tracking. Moussa and Youssef [62] is the first to employed Wi-Fi technology and utilized received signal strength (RSS) measurement to realize device-free localization Unlike RSS which can only give an average value of the signal strength over an entire channel, Channel State Information (CSI) can provide detailed information of communication links. Thus, it is promising to use CSI measurements to localize multiple targets. Wang, et al. [63] used the CSI measurement to estimate fine-grained locations of multiple sparsely located persons. Adib and Katabi [10] described how Wi-Fi signals reflected off moving persons can be extracted to enable through-the-wall tracking.

2.3.1.3 Zigbee/IEEE 802.15.4

Zigbee/IEEE 802.15.4 is a low-data-rate, low-power communication protocol, which is suitable for creating wireless person area network for localizing and tracking a single or multiple persons in the long run [9]. MICA2 sensor nodes were used by Zhang, et al. [43] to form a grid

sensor array. RASS was introduced by Zhang, et al. [64] to divide a large area to small triangle cells with each of the cell formed by three TalosB nodes. Both works allowed multiple targets to be detected when they are located in different grids. In addition, the grid size can be adjusted to improve the accuracy of both solution but it may increase the cost of sensor deployment if the grid size is deceased Wilson and Patwari [65] proposed RTI (Radio Tomography Imaging) technique that can detect and locate multiple persons in real-time in an array of TelosB wireless nodes. While the above works required a dense deployment of the sensor nodes, Wang, et al. [66] leveraged compressive sensing theory, which can reconstruct a signal from a few measurements, to reduce the number of sensor nodes deployed on a monitored area. In this work, 24 MICAZ nodes were deployed to monitor an area of 144m^2.

2.3.1.4 UWB

Ultra-wide bandwidth (UWB) is a promising technology for DFI localization and tracking as it is not interfered by conventional narrowband and carrier wave transmission in the same frequency band, and can penetrate obstacles such as walls and furniture [67, 68]. In addition, the reflections from static objects in an environment and the reflections from a moving person can be distinguished due to the superior time resolution of UWB [69].

In the literature, Kilic, et al. [55] conducted an empirical study of UWB signal for device-free indoor localization and proposed a method for single person detection and distance estimation. Gulmezoglu, et al. [69] proposed a UWB system that can detect and track up to two moving persons using four UWB sensors that measure the travel time of signals between UWB sensors and human subjects. Liang, et al. [70] were able to detect and locate multiple stationary persons though a wall using a UWB radar system.

2.3.2 Infrared

Passive Infrared Sensor (PIR) detects heat radiated from a human body and use it to localize and track a person [71, 72]. Qi, et al. [18] developed two types of two-column infrared sensor modules with each column consisting of 4 infrared sensors using the Fresnel lens arrays. Type I module has three Fresnel lenses to form a FOV (Field of View) in each infrared sensor for multiple targets tracking. Whereas the FOV in type II module is modulated by pseudo-random coded

masks which allow various human motion attributes to be captured for target identification. Similarly, Yang, et al. [33] developed a custom PIR node consisting of six sensors that are arranged to form 12 overlapped detection zones around a sensor node to track multiple persons. The purpose of the arrangement is to estimate the detection angle of each PIR node. The intersections of these detection angles can be used to determine the locations of multiple persons. Tao, et al. [36] deployed an array of PIR sensors on the ceiling to avoid obstruction of sensors' FOV caused by furniture. Each PIR sensors has no overlapping with adjacent PIR sensors to reduce ambiguity and improve the tracking accuracy of multiple persons.

2.3.3 Acoustics and Vibration

Acoustic devices such as microphones can be used for estimating locations of persons [4]. In general, a set of microphones is used to measure the time that sound or vibration travels through a medium from a target to these devices. TDOA can be used to determine the location of persons. Hnat, et al. [48] presented the Doorjamb system that used an ultrasonic device to detect the presence of a person as well as to measure the height of a person. It can track and estimate which rooms are occupied and who were in these rooms. Chen, et al. [57] captured seismic signals that are caused by footsteps to locate a person using three geophones.

2.3.4 Electric Field

DFI localization and tracking in this area rely on changes in capacitive coupling on an electric field that is induced by a person entering a monitored area. Braun, et al. [73] and Valtonen, et al. [13] deployed capacitive floor mats in their testbeds. Pressure from a footstep of an occupant causes a capacity change in the sensors. Thus, the location of a person can be determined. Fu, et al. [74] used the capacitive floor mats but actual electric potential sensors were installed on a side of a monitor area for ease of maintenance. While both works require a person to interact with the sensors directly, Grosse-Puppendahl, et al. [14] proposed to use an electric potential sensor to detect electric potential changes in an environment caused by a human body.

Technology	Advantages	Limitations
RFID	-low cost, low power	-low coverage distance -signal collision -active tag and reader is expensive
Wi-Fi	-wide coverage distance -existing Wi-Fi infrastructure can be utilized	-high energy consumption
Zigbee	-low power consumption -low cost	- vulnerable to interference -short coverage distance
UWB	-high accuracy -good penetration -less interfere	- high cost -interference by metallic material
Infrared	-low cost, simple -low power	-limited coverage distance -false detection due to background temperature -low sensitivity to fine motion
Acoustic	-good accuracy	-susceptible to background noise -affected by multiple sources
Electric field	-high accuracy -low power, low cost (passive electric field)	-difficult to deploy -low coverage distance -susceptible to electromagnetic interference

Table 2-1 Advantages and disadvantages of each localization and tracking technology

2.3.5 Comparison between device-free technologies

The technologies for DFI localization and tracking have their strengths and weaknesses. The summary of the advantages and limitations of different technologies is shown in Table 2-1.

The main advantage of RF-based technologies over the other non-RF technologies such as infrared technology is that they are capable to penetrate obstacle such as walls, doors, etc. Thus, it is quite appropriate to apply in indoor areas which are cluttered with walls, doors and furniture such as residential home and office. The RFID technology is quite cheap and has low power

consumption, especially passive RFID tags but active RFID tag and RFID reader can be quite costly [49]. To cover a large area such as a hall, a warehouse, a large number of RFID tags must be deployed because RFID has a short coverage distance [4]. This can incur high deployment cost. Moreover, reading multiple tags at the same time can lead to signal collisions. The loss of data caused by the collision can affect localization and tracking accuracy. Wi-Fi technology can cover a long distance which makes it is a good candidate for localization in a large scale area. Implementing a Wi-Fi-based localization system can be fast and cheap because we can utilize an existing Wi-Fi infrastructure in a building. However, Wi-Fi-based device such as router and access point requires constant energy to maintain its operation and coverage. This need results in high energy consumption. Unlike Wi-Fi, Zigbee consumes less energy at the cost of a limited coverage distance. Although Zigbee-based transceiver and receiver are quite cheap, a dense deployment of transceivers and receivers can significantly increase the cost of localization and tracking system. Compared to RFID, Wi-Fi and Zigbee, UWB is less susceptible to interference of other signals and has good penetration. Thus, it achieves high accuracy in localization and tracking but may be quite expensive[68].

For non-RF technology, Infrared technology provides a cheap, simple and energy-efficient solution for localization and tracking[17, 46]. However, it has a limited range which makes it less appropriate for a large scale area[43]. Other drawbacks are that a false detection can often occur due to a background temperature and a fine motion is difficult to detect using infrared technology. An acoustic-based system can achieve good accuracy but background noise and multiple sources of sound and vibration can considerably affect its accuracy[75]. An electric field can provide accurate localization result due to a dense deployment of electric field sensors. However, the deployment process can be difficult and costly because the sensors usually are placed on a floor and furniture needs to be relocated during the installation[13, 73, 74]. In addition, electromagnetic interference can affect its accuracy[14].

2.4 Techniques

In this section, we discuss the techniques for implementing DFI localization and tracking. First, we discuss the human detection techniques in section 2.4.1. The techniques for human counting are presented in Section 2.4.2. The techniques for human identification are presented in

Section 2.4.3, and the indoor localization and tracking techniques are discussed in Section 2.4.4 and 2.4.5 respectively.

2.4.1 Human Detection

To track a person, the first task is to detect the presence of a person. When a person enters a monitored area, this person's body interferes with signals transmitted from sensors and we can analyse sensor readings to infer such an event. Mainly there are three classes for presence detection: *non-stationary*, *stationary*, and *both stationary and non-stationary* human detection.

2.4.1.1 Non-stationary human detection

This type of detection aims to detect a person from her movements. In general, measurements are stable in a monitored area without any person. When persons enter or move in a monitored area, the movements cause changes in measurements. Thus, the difference of measurements between with and without movements can be used to determine the presence of persons.

Zappi, et al. [30] deployed three PIR sensors in three different orientations allowing the detection of multiple targets walking side-by-side in a small hallway. In [76] and [25], the authors detected not only the presence of persons and their directions but also estimated their coarse locations. However, an increase in the number of persons affected the detection accuracy. Zou, et al. [77] used two commodity WI-FI routers to detect a person moving past these routers by comparing the similarity between static CSI and occupied CSI measurements in the time domain. Compared to infrared and ultrasonic, WI-FI can provide better coverage and is suitable for a large hallway or entrance.

The detection at a hallway or entrance may be suitable for certain types of areas that only entrance and exit information are needed, such as museums and libraries. But it may not be suitable for an office or home that detection must be performed for an entire area. Moussa and Youssef [62] proposed an RSS fingerprinting technique that compares measurements collected during the offline phase to current measurements in the online phase as the alternative to their previous threshold-based methods [52]. While RSS fingerprinting may not perform well in the real-world environment as changes in the environment can affect the detection accuracy, Kosba, et al. [78] addressed this problem by updating the offline fingerprints according to changes in the

environment. Thus, this work is more robust and accurate than the previous one. Zhang, et al. [43] modelled RSS changing behaviours in WSNs consisting of ZIGBEE-based sensors to detect the movement of a person. However, detection latency increases with node density. RASS [64] adopted the same model while utilized multi-communication channels and proposed triangle topology for a sensor deployment to reduce the latency significantly. Because RSS measurements are susceptible to severe multipath effect in an indoor environment, Xiao, et al. [79] leveraged CSI measurement of WI-FI devices, which is more stable as a fingerprint as it detected the shift of CSI feature patterns to indicate the presence of a moving person in the monitored area. Similarly, Xiao, et al. [80] used burst patterns in CSI amplitude measurement to detect a moving person but their detection method was based on a threshold value. Gong, et al. [81] proposed an adaptive scheme that can automatically adjust threshold values according to multipath propagation in different monitored areas. Thus, this approach can facilitate threshold calibration of CSI measurement. Instead of using the CSI amplitude as the above WI-FI works, Xin, et al. [82] detected a moving person based on the phase difference of CSI amplitude in multiple receiving antennas but a slow-moving person was difficult to detect in this works as the difference could be too small. The above works based on CSI amplitude could detect a fast-moving person accurately but they failed to detect a slow-moving person. Compared to CSI amplitude, CSI phase is more sensitive to slow movements of a person. Thus, both amplitude and phase of CSI measurement were employed in [83] and [84] to detect a person with fast movement and slow movement.

2.4.1.2 Stationary human detection

Compared to non-stationary detection, stationary human detection is challenging as a stationary person induces only small changes or does not produce any change in measurements at all [85, 86]. PIR technology is almost impossible to detect a stationary person because PIR technology relies on movements of a person to cause the temperature difference in an environment [23]. Whereas, the RF technologies such as UWB and WI-FI is promising as it is sensitive to any changes in a monitored area.

The influence of a human presence on RF links such as shadowing and reflection can be used to detect a stationary person. Palipana, et al. [87] modelled non-linearities in CSI amplitude under

the influence of a person and used it to detect a static person. However, they could only accurately detect a person stands in a line of sight between a transceiver and receiver. Omnidirectional Passive Human Detection (Omni-PHD) was introduced in [88] and [89] to detect a human in four directions. Kilic, et al. [67] relied on the delay of reflected signals to determine a presence of a person in a UWB network. Yuan, et al. [85] employed a variance of time of flight (TOF) to detect a static person through a wall.

To detect a stationary person, vital signs of a human such as breathing can be used as an intrinsic indicator. An early work could detect a person by capturing chest motions caused by breathing but it utilized an expensive radar infrastructure [90]. Alternatively, Patwari, et al. [91] detected a breathing person using sinusoidal variation in RSS measurements but it required a dense sensor deployment. Whereas Kaltiokallio, et al. [92] utilized only a pair of TX and RX to determine the presence of a person but required a dedicated setup. WI-FI hardware was also employed to detect a person from their breathing. Using a similar setup as [92], Liu, et al. [93] could detect a person breathing using CSI measurements. While the above works detected the presence of a person in a very small area, Wu, et al. [94] could detect a person from her breathing in a much larger area using some WI-FI routers. UWB technology can also be used to detect a station person from vital signs. Liang and Deng [95] extracted a vital jiggle signal, i.e. chest or heart movement, from transmitted signal between the UWB radar and a person to detect a person through a wall. Liang, et al. [70] utilized multiple-input and multiple-output (MIMO) UWB radar to detect multiple stationary persons through the wall simultaneously.

2.4.1.3 Both Stationary and non-stationary target detection

In some application such as remote health monitoring and intrusion detection, a person may or may not move but it is required to continually detect the person. Therefore, we are required to detect both moving and stationary person.

Braun, et al. [73] were able to detect both stationary and moving-person using capacitive floor mats but the deployment of the floor mats may not be practical if a room was already occupied by furniture. Zhao, et al. [86] used the histogram difference of RSS measurements to detect both stationary and moving person but it required a dense deployment of RF nodes. Zhou, et al. [96] leveraged CSI fingerprint and Support vector machine (SVM) to detect a person

regardless of their motion state. Han, et al. [97] constructed a frequency-based fingerprint from CSI measurement. In addition to the detection of a person, the state of a person, i.e. stationary or moving, can be distinguished. Fang, et al. [98] employed a deep learning approach that analysed both CSI amplitude and phase information to detect the presence of a person as well as her basic activities such as walking and sleeping. Domenico, et al. [99] used features of the mean Doppler spectrum estimated from CSI measurements to enable through-the-wall presence detection for both stationary and moving persons. The above works offer end-to-end process to detect both moving and stationary person. Whereas, the following works require two separate methods to detect a moving and stationary person. Wu, et al. [94] employed both CSI amplitude and phase information to enable the detection of a moving person with dynamic speed and the detection of a stationary person through breathing. Adib, et al. [100] adopted Frequency Modulated Carrier Waves (FMCW) to detect a moving person using the changes in Time of Flight (ToF). To overcome the issue that a stationary person could not be detected caused by the elimination of reflections of all static objects including a stationary person, the authors also use breathing as an indicator for detecting stationary persons.

2.4.2 Human counting

Target counting is interesting to a wide range of applications. For example, energy consumption can be managed efficiently if we can accurately count the number of persons in an indoor area such as home and office. The counting techniques can be classified into *binary-based*, *clustering-based*, *statistical-based*, and *compressive sensing-based* counting.

2.4.2.1 Binary-based techniques

In this technique, binary sensors such as PIR sensor are installed in a monitored area. A sensor generates an output of '1' if one or more persons stay within its detection area. Otherwise, a '0' is generated when no one is in the area. Based on this fact, snapshot-based counting has been employed to count the number of persons in [101-103]. The snapshot technique considers states of all binary sensors, i.e. on or off, at a particular time as a snapshot and use this information to count the number of people. Song and Wang [102] proposed a dynamic colouring technique (DEC) to estimate the number of persons by creating and maintaining a set of target distribution scenarios

of different colours. For each scenario, shrinking and expanding methods were introduced to determine feasible regions of persons based on sensor reading before and after an event. This work was further improved in [101]: instead of utilizing only a single snapshot, a sequence of snapshots that contains both spatial and temporal dependencies was used to improve the estimation accuracy. Wahl, et al. [104], [105] relied on occupants' paths extracted from PIR sensors that are distributed systematically in an office area to count the number of persons and require fewer sensors than the above works. Due to the binary measurement, the technique is not computational expensive but false detections, which often occur for binary sensors, can degrade the accuracy of the person counting.

2.4.2.2 Clustering-based techniques

Clustering-based techniques are focused on identifying multiple non-overlapping clusters, each of which may contain one or more targets. The objective is to cluster measurements generated by different persons so the number of clusters can infer the number of persons in a monitor area. Zhang and Ni [11] deployed a grid of wireless nodes on the ceiling to monitor persons passively. In the multiple person setting, clusters of communication links are formed according to the number of persons. However, it can only provide accurate counting when a person is not close to each other. ACE [106] employed hierarchical agglomerative clustering to group candidate locations and counting the number of persons without any prior knowledge of the number of persons. Similarly, attenuation peaks occurring in RF imaging-based localization or Radio tomographic imaging (RTI) localization can be counted to indicate the number of persons [107]. However, it may not be accurate as some peaks may be noise. Thus, clustering techniques were also employed to remove noise and improve the accuracy of the counting. Nannuru, et al. [108] employed K-mean clustering to track the number of persons in their RTI system because the number of persons is known in advance. Wang, et al. [107] employed soft clustering techniques to determine the number of persons without prior knowledge. Similarily, Zamzami, et al. [109] used soft clustering to group data generated from different types of sensors e.g. PIR, temperature, and humidity to estimate the number of occupants. However, the clustering approach may provide an inaccurate counting because two persons staying close to each other may result in only one cluster due to similarity of measurements.

2.4.2.3 Statistical-based techniques

Statistical-based techniques rely on statistical models to estimate the number of persons. The probability distribution function can be used to represent the distribution of multiple persons. Raykov, et al. [110] used a single analogue PIR sensor to count the number of persons and modelled the PIR data from different meetings with different Laplace distributions. Then, regression techniques are employed to determine the number of occupants. Wu, et al. [111] represented a distribution of multiple persons in a wireless sensor network using a specific probability distribution that can be estimated using the regression techniques. Then, the number of persons is estimated by a maximum likelihood estimation technique. Depatla, et al. [112] modelled the probability distribution of the received signal amplitude as a function of the total number of occupants. Then, it was compared to the one obtained from experiments using the Kullback-Leibler divergence as a metric that shows how much two probability distribution differ from one another. Then, the argument that minimizes this metric is taken as the estimated number of occupants. Statistical metrics such as mean, variance, etc is also adopted in a number of works for occupancy estimation. Adib and Katabi [10] used spatial variance to indicate the number of persons. Thus, the spatial variance increases correspond to the number of persons. Domenico, et al. [113] used mean and variance of CSI measurements as features for the linear discriminant classifier to estimate the number of targets. Xi, et al. [114] formulated a stable monotonic function that describes the relationship between the number of occupants and variance-based CSI features. Then, a new metric based on the percentage of nonzero elements in a dilated CSI matrix was proposed to improve the accuracy of the occupancy counting.

2.4.2.4 Compressive-sensing-based counting

In the Compressive Sensing (CS) theory, a monitored area is divided into N grids, which are represented as a signal vector of N elements and the number of persons in the monitored area is equal to K. The corresponding n^{th} element of the signal vector is marked as '1' if there is a person in the n^{th} grid and '0' otherwise. The number of persons is usually smaller than the number of grids, which indicates that the signal vector has a sparse-nature. Thus, we can estimate the number of persons by recovering the signal vector from a vector of length M containing a small number

of measurements instead of using measurements from all N grids. To be specific, given the measurement vector and a sensing matrix phase, which is obtained at the offline phase and contains the difference of measurement caused by a person for each of the N grids, a signal vector with a minimum possible number of nonzero entries can be computed. The advantages of CS theory is that the estimation of a location and the number of persons is in general accurate. however, this technique also suffers from the complexity of sparse signal recovery.

Zhang, et al. [115] was the first to prove the applicability of CS theory for counting problem and proposed a variant of matching pursuit algorithm so-called greedy matching pursuit, which does not require prior knowledge such as the number of persons and was more efficient than other approaches such as Orthogonal Matching Pursuit (OMP) and sparsity adaptive matching pursuit (SAMP). Wang, et al. [66] applied the same algorithm to solve the problem of counting and localizing persons in a large-scale area and proposed a systematic approach to select an appropriate grid size to achieve higher accuracy. Wang, et al. [116] proposed E-HIPA system that leverages the advantage of CS theory to reduce the number of sensor nodes. Thus, their work is energy efficient. In addition, the proposed adaptive orthogonal matching pursuit (AOMP) algorithm is more efficient than OMP and SAMP. The above works assume that a person stays inside a grid only. Therefore, a sparsifying dictionary can be constructed corresponding to the grid and the representation coefficient encrypts the number and positions of multiple persons. However, the violation of this assumption can affect the accuracy of compressive sensing due to the dictionary mismatch between the assumed and actual ones. Sun, et al. [117] dynamically adjust the grid to alleviate or even eliminate dictionary mismatch. Guo, et al. [118] proposed a more accurate sparse approximation model to reduce errors created by off-grid and incorporated incomplete prior information to improve the counting and localization accuracy for off-grid persons. Both off-grid works achieved a higher accuracy but their complexity is signification higher due to their iterative algorithm compared to the on-grid ones.

2.4.3 Human Identification

Target identification is one of the important tasks for localization and tracking of multiple persons as it allows us to distinguish between different persons correctly, which improves the accuracy of localization and tracking. We categorize approaches for target identification based on

types of collected data or signal. As a result, there are three categories including *CSI gait-based*, *Thermal gait-based*, and other approaches.

2.4.3.1 CSI gait-based target identification

Channel State information of WI-FI devices contains useful information and can be used to identify a person. The CSI gait approaches focus on extracting a step cycle of a person and use this information to identify a person. Wiwho [119] was the first work that realized person identification using wireless signals. To identify a person, step cycle and walking behaviours were extracted from CSI amplitude using a peak-valley detection algorithm and frequency domain analysis respectively. Similarly, Zhang, et al. [120] employed CSI amplitude but extracted only walking behaviour from the CSI data. Instead of CSI amplitude, Hong, et al. [121] used subcarrier-amplitude frequency (SAF), which is the frequency on values of each subcarrier and amplitude pair, as a feature and achieved a better result than [119] and [120]. In [122] and [123], CSI data was divided into segments according to walking steps. Thus, per-step features can be extracted and utilized together with walking behaviours to improve identification accuracy as well as to recognize strangers. Although these works can achieve a good identification result, they perform identification for the single-person setting only. In addition, persons need to walk in a straight line without stopping or changing direction, otherwise, the identification may fail.

2.4.3.2 Thermal gait-based target identification

Infrared light emitted from human bodies can be detected by a PIR sensor equipped with a Fresnel lens. The lens divides a detection area of a PIR sensor into different zones. Thus, body parts of a person generating distinct sensor readings can be used to identify this person. Qi, et al. [18] developed a double-column sensor module that each column has four binary PIR sensors with FOV masking and also proposed a fusion scheme to improve the identification accuracy, especially for a single person. An analogue output of the PIR sensor was later employed as it provides more features than the binary one. Yun and Lee [124] utilized time-domain features such as amplitude and passage duration as a person's profile and used Support Vector Machine (SVM) to achieve a promising result. While the previously mentioned works utilized only features from the time domain, Xiong, et al. [31] extracted features for both time and frequency domain. Thus,

they can achieve better results for multiple moving persons. Xiong, et al. [125] utilized frequency domain features extracted by Fourier transform, short-time Fourier transform and wavelet packet transform and fused them to identify a person walking past an array of different height PIR sensors. This work provided a good identification result as each part of a human body was captured by individual sensors. Yan, et al. [126] conducted an empirical study on the factors that affect human identification using PIR sensors; then, a mathematical model is derived to map the relationship between two factors and their influence on identification accuracy. One challenge of the thermal gait approach is that hardware cost can be high because it relies on a large number of PIR sensors to capture different parts of a human body. Changes in ambient temperature or person clothing can also affect the identification performance. These issues may prohibit the use of this approach in real-life condition.

2.4.3.3 Other approaches

Instead of utilizing gait information, some existing works rely on other types of information to identify a person. Tao, et al. [36] distinguished each person based on object interaction and behaviours such as private desk information, room entering/leaving information, path, and speed. Hsu, et al. [127] leveraged RF reflections to capture users-object interaction as an image. Then, the convolution neural network (CNN) was adopted to identify persons. However, these works rely heavily on prior knowledge such as object locations and map layout. Alterations of this information can affect the identification rate considerably. The physical properties of a person can also be used to diffferentiate between a person. Hnat, et al. [48] measured the height of a person using ultrasonic sensors and used this information to identify a person. However, identifying persons with the same height is an issue in this work. Zhao, et al. [128] leveraged millimeter wave radar to capture the body shape of each person and used this information sequentially so both body shape information and movement behaviours were taken into account.

2.4.4 Localization

In this section, we describe the fundamental techniques for indoor localization including *proximity*, *triangulation*, *model-based*, *imaging-based* and *RF fingerprinting-based* techniques.

2.4.4.1 Proximity-based localization

Proximity techniques provide symbolic relative location information. Usually, it relies upon a dense grid of sensor nodes, each having a well-known position. When a single sensor node detects a person, the person is considered to be co-located with it as shown in Figure 2-3. For example, the person P1 activates the sensor node N2. Thus, the location of the person P1 is considered to be the same as the location of a sensor node N2. This method is in general easy to implement using different types of technologies. In particular, device-free indoor localization systems using infrared [129] and radio frequency identification (RFID) [130] are often based on this method. If a person *i* are in the range of sensor *j,* the measurement is equal to '1', otherwise, it is equal to '0'. A threshold value is used to determine whether a person is in range or out of range. Hosokawa, et al. [131] relied on an array of PIR sensors attached on the ceiling to achieve accurate results. These works relied on the assumption that persons are likely to maintain their trajectories and desks information to solve the ambiguity caused by multiple people. Zhang, et al. [130] used RFID tags array having passive tags and active tags to locate multiple persons as well as to determine their frequency trajectories. However, this approach relied on the assumption that persons do not change their direction abruptly and do not standstill. Similarly, Liu, et al. [58] aimed to determine the frequency trajectories of persons from an array of active RFID while the deployment cost can be expensive due to the use of active RFID tags. Overall, the proximity technique is easy to implement while the localization result is coarse. This technique is suitable for applications that do not need high localization accuracy. To achieve an accurate result, high density of sensor nodes is required, which might not be practical and economical in a real-life situation.

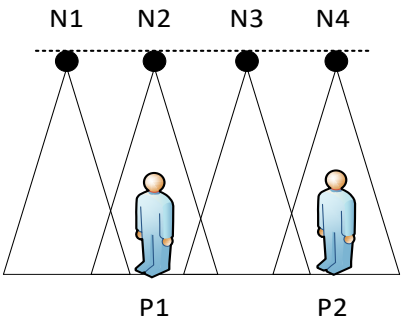


Figure 2-3 Illustration of Proximity-based techniques

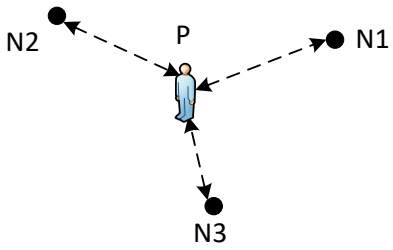


Figure 2-4 Localization based on lateration Techniques

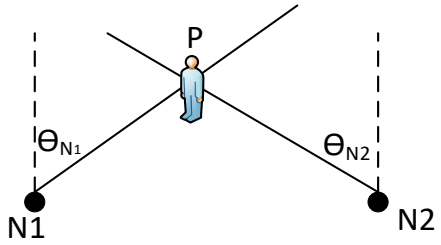


Figure 2-5 Localization based on Angulation techniques

2.4.4.2 Triangulation

A person's location can be estimated using the geometric properties of a triangle in the triangulation-based techniques. There are two categories: *Lateration* and *Angulation*.

As shown in Figure 2-4, in the lateration technique, a person P is located between the reference points N1, N2, and N3. For each reference point, we need to estimate the distance of the person and a reference point. Once the distances between the person and the three reference points are known, we can estimate the location of person P. Common measurements for lateration are time of arrival (TOA) and time difference of arrival (TDOA). The distance can be obtained by multiplying a radio signal velocity and travel time or by calculating the attenuation of radio signal strength [49]. Angulation (AOA estimation) techniques draw direction lines by measuring an angle of arrival (AOA) from each node to a person. Then, the intersection of several pairs of direction lines can be used to estimate the location of a person [132]. In Figure 2-5, two angles θ_{N1} and θ_{N2}, are measured to identify the direction of the person P from nodes N1, N2 respectively. Then, the two lines are drawn from two nodes. The intersection of the two lines is considered to be the location of person P. The AOA technique is not computational expensive as it only needs a weighted least square linear system solver [133]. Wang, et al. [7] leveraged multipath problem to measure AOA and this information to localize a person. Li, et al. [134] used Dynamic Music algorithm to identify AOA of a signal from a moving person. Adib and Katabi [10] proposed the through-wall localization technique that extracts AOA of the RF signals reflected from a moving person to localize a person from static objects such as furniture and wall. One of the limitations of these two works is that they cannot detect the non-moving person. For TOA-based solutions, Adib, et al. [100] relied on multi-shift FMCW(Frequency Modulated Carrier Waves) to extract reflections from multiple persons. However, it required non-standard

hardware. In addition, a coverage range is restricted to 10 meters. Thus, it may not be suitable for large-scale deployment. Triangulation technique can achieve good accuracy but indoor environments such as wall and furniture pose issues such as multipath which can affect its accuracy. In addition, some specific hardware may be required to overcome the negative effects of indoor environments.

2.4.4.3 Model-based localization

Model-based localization techniques can reduce or even avoid offline training efforts. A presence of a person in the Fresnel zone can affect transmitted signals such as absorption, diffraction and reflection. For example, as shown in Figure 2-6, transmitted signals from a TX node can travel to an RX node through the lateral side of a person body due to the diffraction. Thus, there are diffraction paths. Provided that the distances from the diffraction points to the main link are known, the distance from a diffraction point of each path to a sensor can be estimated. Hence, we can model these effects mathematically and use them to determine the location of a person between TX and RX. Shadowing models aim to characterize RF attenuation happens when a person obstructs RF links in a monitored area. There are different shadowing models. An excess path delay model proposed in [53, 135] characterizes the influence of a human presence on communication between an RFID reader and passive RFID tags. Then, Maximum likelihood and linear Least Squares methods were used to localize a person. Based on the same model, particle filtering and Kalman filtering are employed to locate a person [44].

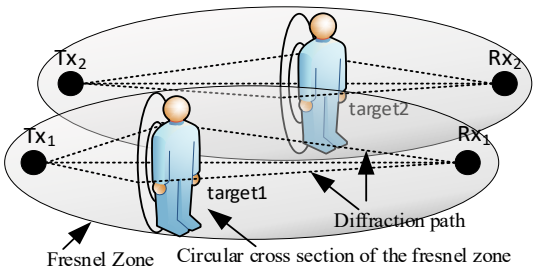


Figure 2-6 Illustration of a Fresnel zone concept to model the effect of human locating between transceiver and receiver

An elliptical model can describe the RSS variation caused by a person obstruct a radio link. In this model, the RSS variation of RF links is inversely proportional to the square root of the length of the radio link. This model has been used in a number of papers such as [66], [136]. Instead of RSS variation, Wang, et al. [137] proposed a shadowing model based on differential RSS that do not require a reference measurement to be collected and more stable in a dynamic environment. While other models focused only on RF attenuation, Exponential-Rayleigh model (ERM) [138] not only considered RF attenuation but also took the target-induced multipath interference in the account. Thus, the accuracy of localization was improved. Compared to ERM, a Saddle surface model (SaS)[139] took the distance to the transmitter and receiver in consideration. Hence, it could better describe the nature of a shadowing effect. In diffraction models, RF signals cannot penetrate a human body but are diffracted by a side of a human body to travel to a receiver. Wang, et al. [140] modelled a person as a cylinder and employed diffraction theory to characterise RSS changes accurately. This model has been employed in [141, 142] and [143]. While other works focused on RSS based model, a power fading model [63] is used to describe the effect of human on a CSI measurement. Propagation fading, diffraction fading, and person absorption can influence the power fading between to a pair of transceiver and receiver.Then, a gradient descent (GD) and genetic algorithms are employed to estimate locations of persons. Although This appraoch can reduce a burden of fingerprint collection, model parameters need to be tuned for every new location.

2.4.4.4 Imaging-based localization

Radio tomography Imaging is a transmission-based imaging technique that relies on the signal strength from various links. The RTI system is centred on estimating an image vector that characterizes the amount of the signal attenuation caused by an object in a monitored area. Thus, the location of a person can be estimated as a location of the occurred attenuation in the monitored area as shown in Figure 2-7. RTI technique was first proposed by Wilson and Patwari [65]. It has been employed in many works such as [144], and [117]. RTI technique was further improved to provide a through-the-wall localization. The motion-induced variance of received signal strength was exploited to locate a moving person behind a wall in [145].

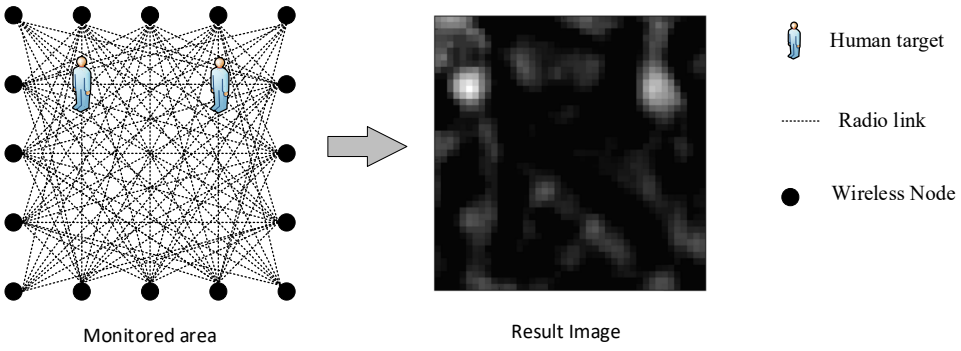


Figure 2-7 Illustration of radio tomography image

However, locating stationary persons cannot be performed properly because signal variation slowly diminishes resulting in persons varnishing from the image. Whereas. Kaltiokallio, et al. [146] employed the fade level which is a function of RSS to determine an RF link quality and was able to locate a stationary person behind a wall. In reference [147], the inverse area elliptical model (IAEM) was proposed to enable RTI technique for a complex structured building. Multiple persons can be simultaneously detected in RTI network due to high density of RF node. Based on the fact, a number of works were proposed to handle multiple persons: Nannuru, et al. [108] proposed a magnitude model to localize up to 3 persons. However, the number of persons is fixed and particle-based approach can incur high computational usage when the number of persons is increased. Bocca, et al. [148] employed the fade level and adopted machine vision methods to detect up to 4 targets in real-time. In addition, intersecting trajectories was handled in this work.

Compared to the above works, Wang, et al. [107] proposed variational Bayesian Gaussian mixture model (VB-GMM) algorithm to locate multiple persons accurately without any knowledge of locations and the number of people. However, the RTI technique was quite computationally expensive, incurs high deployment cost, and create a high power consumption due to a high density of RF sensors. Some works focused on reducing the computational expense such as [149] and [150]. To reduce the number of RF nodes, there was an attempt to incorporate the compressive sensing theory into the RTI framework including [151], [152], and [153].

2.4.4.5 RF Fingerprinting-based localization

RF fingerprinting-based localization is a technique for estimating the location of a person by mapping the current measurements with the pre-collected fingerprints as shown in Figure 2-8. In general, it consists of two phases: the offline phase and the online phase[154]. In the offline phase, radio signals are observed when a person is at know position in a target environment. Then, the observed signals are collected as fingerprint databases. In the Online phase, the current radio signal at an unknown location is compared with those stored in the databases and the closest match is returned as the estimated user location. There are fingerprint-based localization techniques such as probabilistic methods [79, 155, 156], support vector machine (SVM) [96], and neural networks [54, 157]. For RSS based fingerprint, Seifeldin, et al. [155] developed a Nuzzer system in a large scale indoor setting with existing Wi-Fi infrastructure. However, localizing multiple persons can be difficult for Nuzzer. Xu, et al. [158] localized multiple persons using a fingerprint collected from a single person. Thus, this can reduce time and effort in fingerprint collection considerably. Sabek, et al. [106] used a cross-calibration technique and an energy minimization framework to reduce the calibration effort in the offline phase. In addition, the clustering technique was used to remove outliers and achieved better localization accuracy in multiple people setting than Nuzzer and SCPL. Ruan, et al. [59] leveraged human-object interaction to improve the localization accuracy of the fingerprint method. CSI-based fingerprinting exploits the amplitude and/or phase of each subcarrier of the channel. Thus, CSI fingerprinting is more robust and more accurate than the RSS one. Pilot [79] was the first to employ CSI measurement and proved that CSI is more superior that RSS.

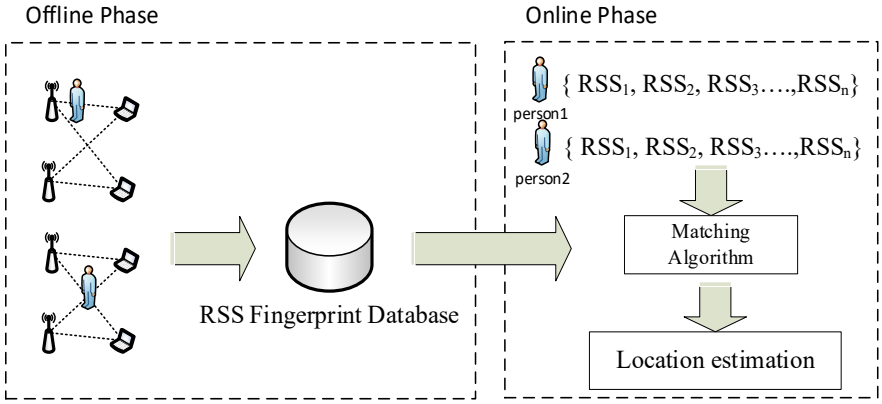


Figure 2-8 Illustration of Fingerprinting localization techniques

Then, CSI based fingerprint localization became popular and has been adopted by many recent papers such as [159] and [96]. Due to rich features of CSI measurement, CSI-based fingerprint location not only can locate a person but also can estimate a person's basic activities or poses such as standing, walking, and squat. Wang, et al. [159] employed wavelet features which are more informative than both time and frequency domain features to perform localization and activity recognition simultaneously but the suitable features must be manually explored and tested. As a result, it becomes labour intensive and time-consuming. However, This problem was addressed by deep learning approach which can automatically extract and learn discriminative features [54].

To sum it up, the collection of a fingerprint database is the main issue for this approach, especially a large scale area because it requires a long time and intense human efforts. The fingerprint collection can be alleviated by crowdsourcing in the device-based localization but it is not possible in device-free setting. Some efforts has been made to transfer a fingerprint from one location to another location [160]. Nevertheless, the trasferring methods can reduce the efforts to some extent and cannot obtain accurate results.

2.4.5 Tracking

In this section, we review the tracking techniques that are commonly used in indoor environments. There are two main categories of tracking techniques: Kalman filtering and particle filtering. In the rest of this section. We present the foundation of these two techniques and how they are used for multi-resident tracking in the existing solutions.

2.4.5.1 Kalman Filtering

Kalman filtering can use a set of noisy measurements observed over time to provide an accurate location estimation than a single measurement [161]. In general, Kalman technique assumes that a system model is linear and a posterior probability distribution is gaussian. It has two phases that are performed recursively including prediction and correction. The prediction phase predicts the state for the next time step. In the correction phase, the predicted state is corrected using the new measurement. Main advantages of Kalman filtering are that it is computationally efficient and relatively simple to implement [162]. For these reasons, the Kalman

technique has been adopted by a number of papers to track a single person in the device-free localization and tracking system [44, 46, 145, 163]. In multiple person setting, the tracking problem can be complex nonlinear and non-Gaussian conditions [137]. However, existing works have shown that kalman filtering can handle this tracking problem [18, 35, 148, 164]. Essentially, they divided the multiple people tracking into a set of single person tracking problems using data association methods. Then, each person is tracked using Kalman filtering independently. According to the above works, this is proof that Kalman filtering can provide an accurate location estimation for both single and multiple persons. However, these reviewed works assumed that a person movement is linear. Thus, a location estimation can be inaccurate when a person changes their walking speed or direction abruptly. Even though there are variations of Kalman filtering that can handle non-linear movements such as extended Kalman filter and unscented Kalman filter, most reviewed papers still prefer to use a traditional one due to its efficiency and accuracy.

2.4.5.2 Particle Filtering

Compared to linear problems, estimation of the posterior distribution of nonlinear problems is more difficult. Particle filtering uses a set of particles to represent the estimated posterior probability density. Thus, it is suitable for nonlinear systems and can handle very large stage space [161]. Similar to Kalman filtering, it operates in two phases iteratively. In the prediction phase, particle filtering uses the principle of importance sampling to generate a set of samples. In the correction phase, weights of particles are assigned proportionally to the observation likelihood. However, particle filtering can suffer from a degeneracy problem that low weighted particles are neglected or disappear after a few iterations [165]. As a result, particles are lack of diversity so this problem decreases the performance of Particle filtering dramatically. To overcome this issue, the effective sample size measurement and the resampling step are introduced to maintain the effective number of particles to maintain its accuracy. Many works have employed Particle filter successfully to track a single person because it can handle non-linear motions accurately [139, 165, 166]. Thus, both stationary and moving-person can be tracked. Furthermore, many works also showed that particle filter can track multiple persons, which are highly non-linear [108, 138, 167, 168]. Despite its high accuracy in the location estimation, the efficiency of particle filtering depends on the number of particles and a dimension of a state. The particle filter can perform

poorly for multiple person tracking which has a high state dimension and requires a large number of particles[162]. Although this problem can be addressed using one particle filter for each person[169], we should be careful when we applying the particle filter.

2.5 Performance Evaluation

In this section, we first describe the commonly-used metrics for evaluating the performance of DFI localization and tracking systems then we compare some existing works based on these metrics. The metrics described in this section include: 1) Accuracy, 2) Precision, 3) Real-time complexity, 4) Cost, 5) Robustness, and 6) Scalability.

2.5.1 Accuracy

Accuracy is the major concern for localization and tracking of multiple persons. In general, to measure accuracy, estimations are compared with ground truths. In our review, we discuss the accuracy for three tasks including localization, identification, and counting.

There are two widely-used accuracy metrics for localization and tracking: the *mean distance error* and the *location error rate*. The former metric is used to measure the accuracy for two-dimensional location. The mean distance error is calculated by averaging the Euclidean distance between the predicted coordinates and the ground truth coordinates. The latter is mainly adopted for discrete location estimation. It is calculated by dividing the number of correct locations with a total number of locations. For Target Identification, the commonly used metric for evaluating target identification is identification rate and it can be calculated by dividing the correction identification with the total number of ground truth. In addition, a confusion matrix can be used to analyse the identification results. To measure the accuracy of person counting, the estimated number of targets is compared to the actual number of targets. The accuracy of target counting then can be calculated using counting percentage which is a ratio between the estimated count and the actual number of persons.

2.5.2 Precision

DFI localization and tracking systems not only need to provide location estimations accurately but also need to perform consistently, i.e. they need to produce less variation in their

results over a series of trials. Precision is a metric that determines the consistency of location estimation. It is normally used in conjunction with accuracy. Cumulative probability functions (CDF) of a distance error can be used for measuring the precision of DFI localization and tracking systems. It gives the distance error between distributions of the estimated location and the actual location. For example, if the CDF of the first system is 90% within 2.3 m and the CDF of the second one is 50% within 2.3m, then, it is favourable to select the first system as it is more consistent than the second one.

2.5.3 System Latency

System latency measures how fast a DFI localization and tracking system process measurements and locate a person in real-time. Based on our review, most works report the system latency by measuring the total time that is required to compute estimations in units of second or millisecond.

2.5.4 Cost

The cost can affect the adoption of a proposed solution. The cost usually comprises three parts: hardware cost, energy cost, and human effort cost. Ideal solutions should have reasonable hardware cost. In addition, the solutions should require fewer human efforts and time to install and calibrate. The power consumption of the solution should be low so that it can reduce energy cost. A cost-efficient solution can be achieved by analysing and understanding the needs of the localization and tracking system, e.g. area size, resolution of localization and tracking result, and the number of occupants.

2.5.5 Robustness

Robustness means a DFI localization and tracking system performs properly even when there are irregular or incomplete measurements caused by sensor breakdown, severe environment, or different deployment environments. Robustness can be evaluated by conducting experiments in testbeds that have various geographical layouts and furniture. For example, a large hall may experience a less multipath problem than a cluttered office room. Then, the accuracy of various testbeds is compared. Mean distance error and CDF of a distance error can be used as metrics in

a comparison of these results. In addition, a sensitivity analysis, which evaluates each output to determine its contribution towards an output uncertainty, could be used to measure the robustness.

2.5.6 Scalability

Scalability means a system can scale to different size of areas. Scalability requires that the performance of a system should not decrease dramatically in a large deployment area. Normally, the scalability is evaluated in relation to the accuracy. Experiments in the reviewed papers are usually conducted on a few testbeds that have different area size. Then, the results of these experiments are compared to determine how accurate their systems are when the size of a monitor area is varied. The scalability can be determined in term of the density of sensor nodes. It refers to the number of sensor nodes per unit area e.g. node/m^2. Scalability can not only be calculated using the density of sensor nodes but also be calculated using the density of persons.

2.5.7 Performance comparison

Based on the above-discussed metrics, we evaluate and compare some existing works for DFI localization and tracking as shown in Table 2-2. we have selected the state-of-the-art works from their respective technology. We extracted the actual values from their works and then presented them in the table.

Table 2-2 Comparison of Existing DFI localization and tracking solutions

Technology	Measurement	Reference	Technique	Accuracy	Precision	Latency	Robustness	Scalability	Deployment Cost
ZIGBEE	RSS	[158]	FP	≈1.3m	N/A	≈0.9s	-office -open area	-150 m^2, -400 m^2	-22sensors -20sensors
	RSS	[148]	IM	≈0.45-0.55m	80% less than 0.6m	≈0.015 s	-open area -apartment -office	-70m^2 -58m^2 -67m^2	-30sensors -33sensors -32sensors
	RSS	[64]	MD	≈1m	70% less than 1m	≈0.26s	-open area	-400m^2	-23sensors
	RS	[170]	MD	≈0.8-	80%	≈2s	-open area	-108m^2	-16sensors

	S			1m	less than 2m			-300m^2	-63sensors
RFID	RSS	[130]	PX	99%-75%	N/A	≈30s	-open area	N/A	-5 readers -81 tags
	RSS	[7]	TR	≈0.17 m	90% less than 0.17m	≈0.5s	-lab -library -open area	-108 m^2 -70 m^2 -75 m^2	Each area -4readers -21tags
WI-FI	CSI	[63]	MD	≈0.5m	80% less than 0.7m	≈0.065 s	-home -classroom -library	-150 m^2 -70 m^2 -108 m^2	Each area -4aps -7mps
	RSS	[106]	FP	≈1.3m	80% less than 3m	≈0.002 s	-apartment -office	-114 m^2 -130 m^2	Each area -2aps -3mps
	RSS	[155]	FP	≈2m	75% less than 9.3m	≈0.003 s	-office	-750 m^2 -130 m^2	-3aps,2mps -3aps,3mps
INFRARED	AOA	[168]	TR	≈0.5m	N/A	≈19s	-open area	-100 m^2	-9 nodes
	BIN	[36]	PX	≈0.3-0.4m	90% less than 2m	N/A	-office	-128 m^2	-43 sensors
	AOA	[18]	TR	≈0.5m	N/A	N/A	-open area	-81 m^2	-4 nodes
UWB	AOA	[69]	TR	≈0.25 m	N/A	≈1s	-office	-9 m^2	-4 sensors
	TOA	[67]	TR	≈0.12-1.8m	N/A	N/A	-open area	-15 m^2	-4 sensors

TR = Triangulation IM = Imaging FP = Fingerprint MD = Model PX=Proximity

From the table, we can see that Zigbee/IEEE 802.15.4 and Wi-Fi are adopted by a majority of the reviewed papers including the ones shown in the table due to their reasonable cost and availability as off-the-shelf products. Whereas UWB is more expensive and less used. Other technologies, such as RFID and PIR, are also adopted for small-area settings due to their short coverage distance. RSS measurement is popular among RF technology as it is relatively easy to implement. Similarily, Triangulation-based measurements such as AOA and TOA are popular as they can be measured using different technologies. Whereas CSI measurement is limited to Wi-

Fi technology only and Binary measurement is less popular as the information it can provide is limited. The localization techniques such as fingerprint, model-based and triangulation are applicable for most indoor localization and tracking scenario. Whereas, Imaging-based technique requires specific setup, which is more appropriate for the through-the-wall scenario than general localization and tracking scenarios. The accuracy of these works is reasonable with respect to their techniques and settings. They can locate a person with an error of less than 1 meter. The best result can be achieved with less than 20 cm error but such approaches usually require the dense deployment of sensor nodes or specific equipment. Most works in the table can provide real-time localization and tracking with latency lower than 1 second. However, they only conducted experiments using a small number of people. It is questionable to see how well they can handle a large number of people.

The robustness of these systems is tested on different testbed as shown in the robustness column in Table 2-2. Most of them can achieve 70-90% of the acceptable accuracy but they only experiment in a controlled manner, which makes it hard to interpret their performance in a real-life situation. The scalability of each work is tested with the area size between 70-150 m^2. Only a few works in the table tackle a larger area but we observe that all of them localize only the small number of persons, i.e. 3-5 persons. This indicates that the current state-of-the-art works are not tested in a large crowd setting. For the cost of deployment, Zigbee, RFID, and PIR use a similar number of sensors to cover a monitored area. However, Zigbee and RFID have the advantage of being RF technology and thus are more appropriate to deploy in a cluttered area than PIR. Although Wi-Fi is moderately more expensive than the above technologies, it can cover a large area with fewer devices, for example, an area of 750m^2 can use only five Wi-Fi devices.

2.6 Challenges and Future Trends

In this section, we first discuss the open challenges of DFI localization and tracking in a multi-resident environment and then we describe the future trends in DFI localization and tracking.

2.6.1 Open challenges

2.6.1.1 Complex scenarios in multi-person localization and tracking

In general, persons need to stay in different areas or zones or to locate sparsely to facilitate localization and tracking tasks. However, DFI localization and tracking systems often operate in a small space such as an office room or a living room which introduces extra challenges: first, the measurement of one person can be interfered by that of another person [59] which affects the accuracy of location estimation. Second, an occlusion occurs when persons are tightly close to each other [23, 148] resulting in the loss of targets. Although assumptions can be made such as treating multiple persons as a single one [43], assuming that the trajectories of persons remain unchanged, or assuming that persons are likely to return to specific locations such as desks [164, 171], such assumptions may not be practical in real-life situations. For example, two persons may meet and then leave in different directions. How to effectively and efficiently deal with the above-mentioned issues raised by complex scenarios remains an open challenge to researchers.

2.6.1.2 Variety and identity of targets

DFI localization and tracking systems are usually calibrated and trained with a single target. However, each person has individual physical properties such as weight, height, and width. These variations may lead to less accurate results as varieties of targets induce irregular measurements [172]. One of the solutions is to train the system using all possible targets. However, it is expensive and time-consuming. There is a work that used a transfer method to handle a diverse category of targets but it still requires prior knowledge [172]. When a new target that does not belong to any category, it leads to inaccurate localization and tracking outcomes. Thus, the problem of a variety of targets is still not fully solved. To identify a person, each work used a certain type of localization and tracking technologies such as PIR [173], WI-FI [122], and Electric field [14]. Different features have been extracted from measurements of chosen technologies to create profiles of persons. Some works used raw features such as an amplitude of signals[124], while the others extracted gait features [120-122]. However, different persons may have similar physical properties or behaviours so similar measurements and features are generated. Thus, it becomes problematic to identify resembling persons. In addition, a person may change their

outfits such as cloths and shoes daily so it may cause some variations in measurements collected by the system and it leads to inaccurate identification [14]. Thus, it remains open challenge to effectively improve the accuracy of these two issues.

2.6.1.3 Scalability to the number of persons

Most existing works on DFI localization and tracking conducted experiments using different numbers of participants and the number of persons is known [35, 100, 164, 174]. However, the number of persons may be unknown and vary from time to time in real-life situations. This uncertainty can affect the performance of the system [175]. In real-world applications, the system may need to handle a large crowd in a large-scale area such as museums, shopping malls, and conference halls. Thus, the development of DFI localization and tracking approaches that can cope with a large crowd is still challenging[9].

2.6.1.4 Changes in an environment

Changes in the environment are one of the factors that affect the performance of DFI localization and tracking systems. With the elapse of time, the performance of the system might degrade continuously [9]. For example, relocation of furniture can affect RF signals in some RF-based DFI localization and tracking systems. The systems need to be recalibrated and re-trained to keep up with all changes occurred in an environment. Thus, a huge amount of human effort is needed. Changes in an environment can happen for both single and multi-person localization and tracking. Some preliminary works addressed the problem using the less sensitive measurements such as CSI [63] and Differential RSS [137]. In addition, most existing works conducted experiments to evaluate their proposed solution for changes for a short period. However, these cannot ensure the performance of the systems in the long run. Online learning techniques which are incremental techniques that are capable of lifelong learning on a device with limited resources in real-time may be a promising solution to this problem [176]. So far, we have found some works that have proved the feasibility and effectiveness of the online sequential extreame learning (OS-ELM) technique for a device-based indoor localization [177, 178]. Thus, how to effectively handle dynamic changes in the long run remains a challenging issue and needs to be further investigated.

2.6.2 Future Trends

2.6.2.1 Knowledge transfer

DFI localization and tracking systems are calibrated and tested in a specific environment and setting. Whenever DFI localization and tracking systems are deployed in a new location, the system becomes less accurate as the previous calibration does not take into account changes in layouts and surroundings such as area size, furniture, and walls. As a result, we need to collect new data and recalibrate the system entirely. To reduce the human effort and time for collecting radio maps or fingerprints, one can transfer and reuse existing fingerprints from a current site to a new site. A transfer algorithm will map the radio maps from an original location to a new location based on the similarity of environments. In addition, knowledge about persons that are collected in the current location could be transferred. This could be particularly useful for the multi-occupancy setting. Some preliminary works attempted to realize this idea such as [172, 179].

2.6.2.2 Self-adaptive localization and tracking system

Changes in an environment such as furniture relocation and room size adjustment can affect the accuracy of location estimation. Human efforts are required to recalibrate parameters and recollect fingerprint databases. To tackle this problem, recent work on WI-FI technology can adapt from a small area to a bigger area by mapping CSI distribution of references grids from a previous location to a new location and using reference grids to estimate CSI distribution for the rest of the new area [160]. As a result, it reduces time and human efforts for collecting the fingerprint. Furthermore, online learning techniques such as OS-ELM that have been employed in device-based solutions [177, 178]. We anticipate that online learning will be adopted in device-free solutions too because these techniques are particularly useful for high dynamic or large-scale areas that are inconvenient to collect new data such as shopping malls. Therefore, we expect to see more DFI localization and tracking works that can adapt to cope with changes in the environment automatically in the future.

2.6.2.3 Deep learning

Deep learning has been employed extensively in many fields including speech recognition and computer vision over the past years. Deep learning is capable of learning the right features and making an appropriate prediction without domain expert or human intervention. It has been recently adopted in a field of DFI localization and tracking [54, 157]. The results show that deep learning can learn discriminative features automatically from RF signals. We believe that deep learning may potentially be useful for device-free localization in the multi-person setting. For example, deep learning technique can be used to identify the most appropriate features and improve the accuracy of multi-resident localization and tracking.

2.6.2.4 Extending localization and tracking to support activity recognition

Recently, techniques of DFI localization and tracking have been extended to recognize simple gestures. The recent works show that recognizing simple activities, such as standing, walking and sitting are feasible [145, 157, 159, 180, 181]. In general, complex activities are composed of simple activities. Thus, we expect further improved techniques will appear so that complex activity can be recognized such as cleaning, cooking.

2.7 Conclusion

Over the past year, device-free indoor localization and tracking are gaining popularity. Although much research effort has been carried out in this area, the presence of multi-residence introduces several new challenges. In this chapter, we have presented an in-depth review of DFI localization and tracking with a focus on the multi-resident environment. We have comprehensively introduced technologies and techniques for device-free localization and tracking. The components of multi-resident localization and tracking were discussed including human presence detection, target counting, target identification, localization, and tracking. For each component, we provided a classification of its related techniques as well as the details of these techniques. We have also discussed the performance metrics for device-free indoor localization and tracking. Finally, we identified the open challenges and future trends. We concluded thereby that our survey can benefit researchers in knowing the latest development in DFI localization and

tracking as well as understanding the pros and cons of the state-of-the-art approaches and systems to develop novel approaches that outperform existing ones.

Chapter 3 A Passive Infrared Sensors System for Device-free Indoor Localization and Tracking

Location estimation or localization is one of the key components in IoT applications such as remote health monitoring and smart homes. Due to having low cost, low energy consumption, and good accuracy, technology based on passive infrared sensors is an attractive option for location estimation. In this chapter, we propose a PIR-based localization system and a dataset collected from our sensor system consisting of commercial-of-the-shelf (COTS) sensors without any modification. Our dataset includes profile data that have over 1,000 samples of different walking directions and test data consisting of multiple scenarios with a sequence length of over 2,000 timesteps. In addition, the dataset contains 36 classes in total. To evaluate our system and dataset, we select and compare different deep learning methods such as CNN, RNN, and CNN-RNN. The results prove the applicability and feasibility of our system and illustrate the viability of deep learning methods for PIR-based localization and tracking. The best methods can achieve an accuracy of 85%, a kappa score of 0.84, and an f1 score of 0.77. We also show that our dataset can be converted for the coordinate estimation so that deep learning methods and the particle filter approaches can be applied to estimate coordinates. As a result, the best performer achieves a distance error of 0.25 meters.

Chapter 3 has been submitted to Journal of Pervasive and Mobile computing

3.1 Introduction

ocation is one of the key information for many IoT applications such as remote health monitoring, smart building, energy management, and security management[3, 6]. It is so important that location information must be precise and processed promptly to ensure that critical incidents such as medical and security alerts can be handled efficiently without any delay [6, 37]. Over the past decade, various surveillance technologies have been developed to provide location estimation in an indoor environment. Video camera offers sequential images of persons but can be used only in the public area due to privacy concern. Wearable and mobile devices, such as smartwatches, smartphones, etc., can be employed to estimate a person's location but they are considered intrusive and uncomfortable because a person is required to wear or hold such device all the time[8]. On the other hand, passive sensing technology such as passive infrared sensors, WiFi, Zigbee, RFIDs, etc. not only can overcome these drawbacks of the intrusive monitoring devices but also provide accurate localization and tracking. WiFi can cover a large area and is not obstructed by a wall, but it incurs a high energy cost to maintain its operations. mmWave can give a very accurate result with a distance error of less than a centimeter but hardware cost is very expensive, and it has high energy consumption. Zigbee also demand a large amount of energy due to a dense deployment of transmitter node and receiver node.

Passive infrared technology such as passive infrared (PIR) sensors and thermopile array sensors require very low energy in a unit of μW and are not affected by electromagnetic interference [35]. In addition, a coverage area can be controlled according to user requirements. Although these kinds of sensors have a low to moderate coverage distance and are susceptible to changes in their surrounding environment, the accuracy of most infrared sensor systems is still within one meter thus making the infrared technology quite attractive.

Most of the existing PIR-based systems utilized binary outputs for analyzing human movements and locations, however, they rely on a sophisticated sensors node or a huge number of sensors, which are not practical for a real-world setting [1, 18, 22]. Some other works leveraged characteristics of the PIR's analog outputs to reduce the number of sensors that are used for location estimation [32, 182, 183]. Analog outputs of a PIR factor in the distance

between a person and the sensor and the person's movement speed direction [24]. A few analog PIR-based location estimation techniques have been proposed, but none of them has made their datasets available to the public. As a result, new researchers cannot reproduce the research and cannot benchmark their proposed method with the others.

In this chapter, we propose a new PIR system and dataset to address the above shortcomings. The dataset contains profile data that has over 1,000 walking samples in different directions and test data that contains both simple and complex scenarios and has more than 2,000 time steps for each scenario. These data have been collected using our developed sensor node consisting of commercial-of-the-shelf (COTS) PIR sensors in a controlled environment without any specific modification to ensure the correctness and unbiasedness of collected data.

Deep learning methods have been introduced over the past years and achieved promising results for localization and tracking based on RF signals [12, 157, 184]. In this chapter, apart from a traditional tracking method such as Particle filters, we also apply some well-known deep learning algorithms to evaluate our proposed dataset. With well-trained deep learning models, we can achieve an average of 77% accuracy and maximum accuracy of 85%. On the other hand, we show that both deep learning methods and particle filters can achieve accurate localization with an averaged distance error of 0.4 meters and a minimum distance error of 0.25 meters.

We believe that our work will facilitate and benefit researchers that are interested in PIR-based localization and tracking by providing our ready-to-use dataset with ground truths to explore new challenges and develop new techniques while being able to evaluate and compare their techniques with existing techniques.

- We propose a new PIR-based localization system that is low-cost and easy to install.
- We perform a data collection to create a comprehensive PIR dataset based on analog signals and provide our dataset publicly for comparing and testing new localization and tracking techniques.
- We empirically evaluate our dataset based on different deep learning architectures and particle filters to show the applicability of the dataset.

The rest of this chapter is organized as follows. Section 3.2 discusses our related works. Section 0 describes the data collection process and characteristics of our dataset. Section 3.4 discusses the procedure and candidate methods that we use to evaluate and validate our dataset. The results

of the experiments are discussed in Section 3.5. Finally, the summary of our work is presented in Section 3.6.

3.2 Related works

In this section, we review existing works for device-free location estimation for both non-infrared-based technology and infrared-based technology.

3.2.1 Non-Infrared technology

Radiofrequency (RF) technology such as Zigbee/IEEE 802.15.4, WiFi, RFID, UWB, and mmWave can be employed as a promising solution. The main advantage of RF technology is that it can cover a long distance and can penetrate a wall [9]. Zigbee/IEEE 802.15.4-based, and RFID-based solutions require a dense deployment of transmitter nodes and receiver nodes, which makes these solutions impractical in a real-life conditions [116]. The common localization techniques for both Zigbee and RFID technology are the imaging method [65, 144, 145], proximity method [130], and model-based method [43, 44, 135, 185]. WiFi-based localization systems are in the mainstream because we can reuse existing WiFi routers to perform localization or location estimation and take advantage of their large coverage. Thus, equipment and installation costs are significantly reduced [52]. Recently, many researchers promoted the use of WiFi channel state information (CSI) because it can provide more information and features than radio signal strength (RSS) [34, 79, 186]. However, CSI is accessible to some specific routers [12]. Thus, The WiFi CSI-based solutions may be less favorable because new hardware must be invested. The main drawback of WiFi-based solutions is that they need to have a constant supply of energy to maintain their coverage [8]. As a result, they incur a high energy cost. Fingerprinting methods [106, 158, 187] are commonly used in WiFi-based solutions. Ultra-wide bandwidth (UWB) is not interfered with by a conventional narrowband and carrier wave transmission in the same frequency band and has superior time resolution, which allows a static person and a moving person to be differentiated [69]. Thus, UWB is also suitable for locating a person. mmWave-based localization takes advantage of mmWave's high-frequency radio wave and directional beamforming to provide very accurate location estimation [188, 189]. However, mmWave devices are very expensive, difficult to procure, and have a high energy usage. Both

UWB and mmWave solutions usually use the triangulation-based technique to localize a person [68, 188]. Besides RF technology, capacitive floor mats can be employed for location estimation. This type of technology can achieve accuracy of location estimation within 1 meter, but an installation of floor mats can be difficult if an area is already fully furnished [73]. Acoustic-based localization can estimate a location of a person accurately by detecting a footstep and measuring angle of arrival (AOA) between a footstep and a microphone [75, 190] but it may be susceptible to background noise. Visible light-based localization utilizes light sources and light sensors. When a person is located within the vicinity of light sensors, these light sensors can measure changes in light level or illuminance passively. As a result, we can infer a person's location by analysing measurements from these light sensors [191]. One advantage of VL based localization is that we can leverage the existing lighting infrastructure. In addition, it is immune to RF interference. However, it cannot operate without a light source and visible light can be obstructed by walls.

3.2.2 Infrared technology

Passive infrared technology can be employed to detect a movement of a person or an object. In general, passive infrared devices such as passive infrared (PIR) sensors and thermopile sensors are commonly used for this task. The basic principle of both types of sensors is that a movement of a person causes changes in temperature on the sensing elements of both sensors and an electric current is generated due to the thermal-to-electric conversion of the sensing element's material. Then, the electric current is converted to voltage or binary as the final output. By utilizing voltage outputs (a.k.a. analog outputs) and binary output, it is possible to adopt both PIR and thermopile sensors for location estimation. Compared with other technologies, both passive infrared sensors consume much less energy, usually in a unit of μW, have a low price, and are simple to employ. Although the infrared sensors are less accurate than some RF technologies such mmWave, they still can achieve the distance error within a sub-meter unit. For these reasons, PIR and thermopile sensors are also viable options for location estimation in an indoor environment such as home and office buildings. There are a substantial number of papers that proposed passive infrared-based location estimation. We classify these papers according to their sensor deployment including *floor-mounted*, *wall-mounted*, and *ceiling-mounted*.

For the floor-mounted category, Hao, et al. [46] developed infrared sensor modules consisting of 8 PIR sensors to detect a movement in 360 degrees. FOVs of 8 sensors are overlapped to form 16 detection regions and a visibility coding scheme is used to associate these regions to the detection of sensors. In their experiment, 4 sensor nodes were placed at each side of a 9m x 9m monitored area. Particle filtering and Kalman filtering are employed to track a person, which achieved an approximate distance error of 1 meter. Yang, et al. [33] used a similar approach and developed a PIR node consisting of six sensors that are arranged to form 12 overlapped detection zones around a sensor node. 9 sensor nodes are deployed uniformly on a 10m x 10m area and they obtained an average distance error of 0.5 meters. Narayana, et al. [24] analyzed characteristics of PIR analog output that are relating to location estimation and develop a sensor tower consisting of 4 sensors for height classification and location estimation. The distance between a person and the sensor tower can be estimated by adjusting variable gains for different PIR to create various detection regions. They placed two sensor towers perpendicular to each other to monitor an area of 8m x 8m and achieve a distance error of 0.3 meters.

For the wall-mounted category, Kemper and Hauschildt [192] utilized 4 thermopile sensors and deployed 4 of them at all corners of a 4.9m x 6.2m area. Probability Hypothesis Density (PHD) filter was implemented to perform location estimation and obtained a mean distance error of 0.25 meters. Mukhopadhyay, et al. [182] adopted the peak-to-peak amplitude measurement from [24] and proposed two models based on hyperbolic function and Piecewise linear function to express a relationship between the peak to peak amplitude and the distance from a sensor. Then, they used multi-lateration and Support Vector Regression (SVR) based techniques to estimate a location of a person. They set up four PIR sensors on each side of a 7m x 7.5m area and obtained a distance error of 0.65 meters. Alternatively, Liu, et al. [32] proposed azimuth change measurement which is the difference in the azimuth of a person with respect to a PIR sensor at two locations and can be easily determined by a law of cosine. They set up a monitored area of 8m x 8m that has 4 sensors located at 4 corners of the area. Particle filtering was employed to determine a person's location and achieve an accuracy of 0.67 meters. Wei-Han and Hsi-Pin [193] mounted two thermopile sensors that are located 3.3 meters away from each other and are oriented at a 30-degree angle on a wall to monitor an area of 2.35m x 3m. The foreground of the human body is extracted by subtracting background images. Then, the angle of arrival

(AOA) is determined using the foreground from two sensors. The location is estimated by AOA based positioning method. Then, a regression model is used to improve the estimation accuracy. Thus, the mean distance error is achieved at 0.13 meters. Narayana, et al. [194] developed a custom sensor module called LOCI consisting of one thermopile and one PIR sensor. By combing both sensors, 3-dimension locations can be estimated. The thermopile can provide location information across the FoV cone axis and the distance between the sensor module and a person is determined from a PIR sensor. The machine learning, i.e., KNN are training to estimate a possible location of a person. Only one sensor is mounted on wall at 1.2 meters height to monitor a 9m x 8m area and they obtain the best accuracy at 0.12 meters.

For the ceiling-mounted category, Yang, et al. [71] deployed 9 PIR sensors to cover an entire apartment. A practical binary sensor model and an accessibility map were proposed to improve the estimation of the particle filter. As a result, They achieved the distance error within 0.6 meters. Tao, et al. [36] used a binary sensor network consisting of 43 ceiling-based PIR sensors to monitor an office room with a size of 15.0m × 8.5m. Information such as desk s' locations and the moving directions to improve location estimation. Wu, et al. [183] developed the rotationally shuttered PIR (Ro-PIR) sensor to enable the detection of a stationary person. The shuttered PIR sensor is mounted at a height of 2.8 meters and can cover a circle area with a radius of 2 meters. They adopted machine learning algorithm such as KNN and SVM to localize and track a person and achieved the accuracy within 0.44 meters.

From our observation, there are abundant public localization & tracking datasets for a video camera and radio frequency technology such as the UCI wireless indoor localization dataset [195], WiFi RSSI indoor localization dataset[196], etc. These public datasets allow new researchers to improve or invent different techniques at ease, which accelerate several improvements in localization and tracking for their respective technologies. For passive infrared technology, a substantial number of works have been proposed but none of these passive infrared works has made their datasets available to the public. There are still some challenges related to certain areas, such as signal processing, feature engineering, signal modeling, etc., that should be further explored. However, it is not possible due to the lack of a dataset. Therefore, we construct a public PIR-based location estimation dataset for researchers to explore and address challenges.

3.3 Our Proposed System and Dataset

In this section, we describe our proposed PIR-based localization system and a process for constructing our PIR dataset. First, we introduce our hardware and its setup in an environment. Secondly, we describe the process of collecting training data and testing data. Lastly, we explain the format of collected PIR data.

3.3.1 Environment and Hardware setup

To prepare for our data collection, we need to develop a sensor node. From a literature review, three types of sensor deployment include floor-mounted, wall-mounted, and ceiling-mounted. A floor-mounted sensor node is quite simple to deploy in indoor and outdoor areas. However, indoor locations, especially offices or houses, can be cluttered with furniture that obstructs the FOV of a floor-mounted node. In addition, a floor-mounted sensor node can cause tripping hazards. For these reasons, we consider wall-mounted and ceiling-mounted is more appropriate to deploy inside buildings. We can adjust a wall-mounted sensor node to achieve the optimal FOV of a sensor node. However, a preoccupied room can lead to difficulty in sensor deployment. For example, furniture such as a wardrobe can block an installation point. Furthermore, the size and shape of a room can add additional difficulty, especially when we need to arrange sensors in a specific way. Whereas, A ceiling-based node does not have these issues. Although a ceiling height may vary for different buildings, we can adjust a detection range of a PIR sensor to address this issue. For this reason, we choose a ceiling-based sensor node.

We studied various designs of a PIR sensor node from existing works. The hardware designs similar to [7, 45] require a dense deployment of PIR sensor nodes that is not practical in real-life. Some works such as [40, 49] need to have complicated masking, which results in a bulky node and is overly complex. We focus on simplifying a design of a PIR sensor node to be more compact and improve area coverage. For these reasons, we develop a sensor node consisting of 5 SEN-13968 PIR sensors (Model: SEN-13968) from SparkFun. The selected model can generate a maximum voltage of 5 volts. We arrange 5 PIR sensors as shown in Figure 3-1. The middle sensor points perpendicularly to the ground and the surrounding sensors point outwards at 30 degrees from vertical as shown in Figure 3-3.

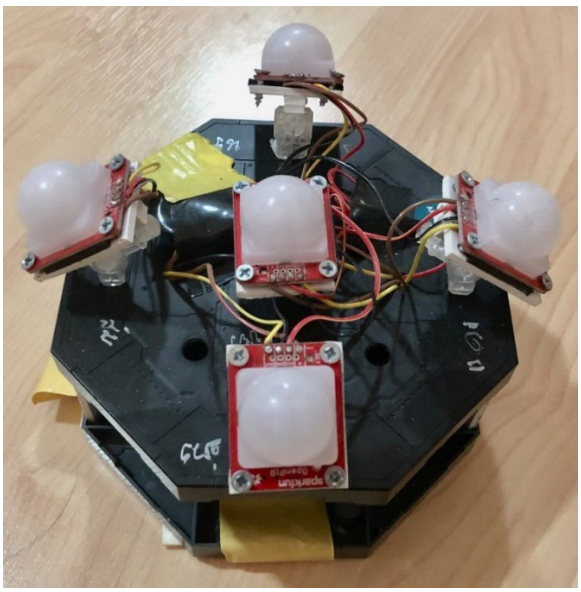


Figure 3-1 A photo of developed PIR sensor node

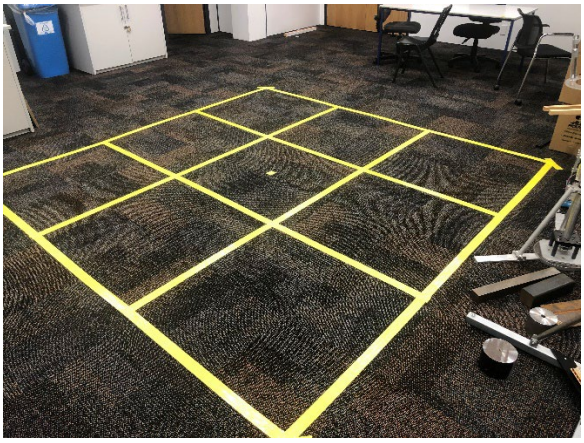


Figure 3-2 A photo of a monitored area

The front FOV and side FOV of each PIR sensor span at 100 degrees and 60 degrees respectively. As a result, overlapping detection zones are formed as illustrated in Figure 3-4. The advantage of the overlapping zones is that they can make a location of a person easier to be distinguished because unique signal patterns are generated from 5 PIR sensors when a person moves in different locations in the monitored area. In addition, the FOVs are aligned symmetrically as much as possible to ensure that we have consistent patterns. All PIR sensors are connected to Arduino MEGA 2560 microcontroller to synchronize the transmission cycle of the five sensors and collect the most recent measurements received from these sensors. Next, we attach the

sensor node to the ceiling in an indoor open office area where its ambient temperature is controlled regularly to ensure that PIR analog outputs are not affected by changes in temperature as much as possible. Then, we set up a monitored area in a square shape as shown in Figure 3-2 that consists of 9 grids and each grid has a size of 1m x 1m. We use an area size of 3m x 3m because a person's movement at the edge of the monitored area is too far from a sensor and is not sufficient enough voltage to trigger a sensor activation. The sensor node is linked to a desktop computer via a serial communication port for processing and storing the collected data. In our work, we set the sampling rate to 60 samples per second because we want to ensure that we do not miss any important features in PIR raw signals, and this sampling rate also simplifies a time synchronization for data labeling.

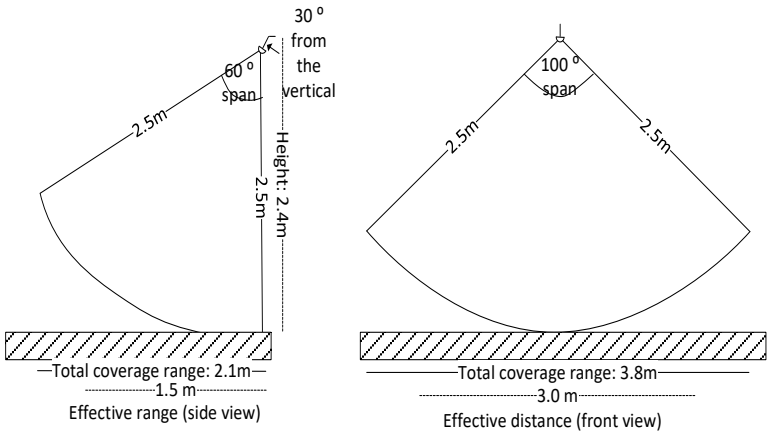


Figure 3-3 An orientation of surrounding PIR sensors

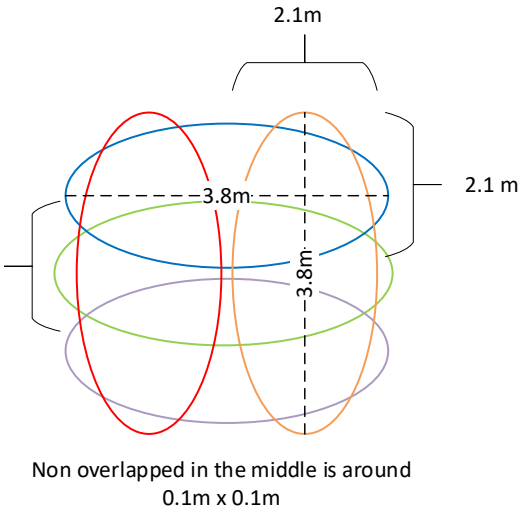


Figure 3-4 An overlapped detecting area of a PIR node

3.3.2 Data Collection

To construct a profile dataset, we further divide each grid in a monitored area into 4 square cells that have 50 cm in width and 50 cm in length. Each cell should be large enough to hold one person because the average shoulder of males and females is less than 50 cm. As a result, we have 36 cells in total. Each grid is labeled using numeric numbers from 1 to 9 and its cells are labeled using numbers from 1 to 4. Figure 3-5 shows numeric labels for all grids and cells in the monitored area. Then, a participant is instructed to walk vertically, horizontally, and diagonally across a monitored area. To be specific, a participant starts walking from the left end to the right end of the monitored area. There are 6 starting locations for a horizontal walking direction. For the vertical walking direction, a person starts from the top end to the bottom end of the monitored area and there are 6 starting locations. For the diagonal walking directions, there are two directions including diagonal left and diagonal right. For the former direction, A participant walks from the top left end to the bottom right end of the monitored area. On the other hand. a participant walks from the top right end to the bottom left end of the monitored area for the latter direction. For each diagonal walking direction, there are 11 starting points. To complete one sample of the profile dataset, a person needs to walk forth and back in every direction. As a result, we have 24 sequences for the horizontal and vertical directions. In addition, we have 44 sequences for the diagonal directions. According to [24], different walking speeds can affect an amplitude and frequency of PIR analog output. However, it is not feasible to collect training data for all walking speeds. To reduce the effect of speed on the PIR output signal, we assume that a participant needs to walk at a constant speed. We employ a metronome that can generate a steady pulse to help a participant control his walking speed to be constant as much as possible. We employ a metronome that can generate a steady pulse to help a participant control his walking speed to be constant as much as possible.

To construct a test dataset, we develop 4 different walking scenarios in the real world as shown in Figure 3-6. Similar to the profile data, we ask a participant to walk at a constant speed. For each scenario, we increase the complexity of walking paths to validate the collected PIR profile. In the first scenario, a participant walks from cell 1-1 to 3-2. Then, he turns and walks from cell 3-2 to 9-4. Finally, he walks from cell 9-4 and finishes at cell 7-3.

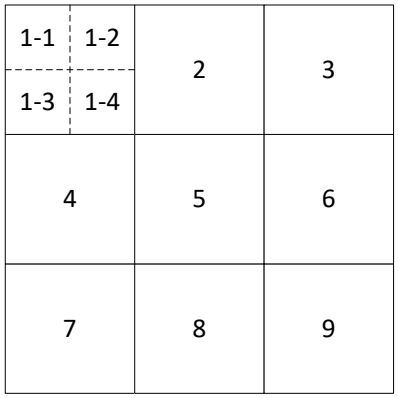
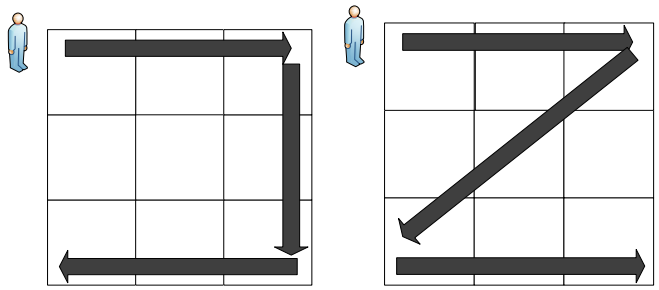
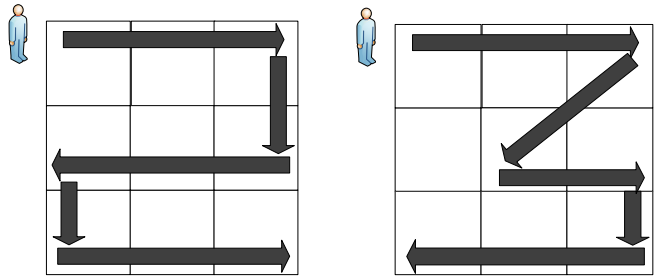


Figure 3-5 Grids and cells of the monitored area with labelling numbers



(a) Scenario 1 (b) Scenario 2



(c) Scenario 3 (d) Scenario 4

Figure 3-6 An illustration of 4 test scenarios

This first scenario is the simplest one because the chosen paths generate signal patterns that are similar to the profile dataset.

In the second scenario, a participant walks from cell 1-1 to 3-2. Next, he walks diagonally from cell 3-2 to 7-3. Finally, he walks from cell 7-3 and finishes at cell 9-4. The second scenario is slightly more complicated than scenario one due to the diagonal walking path. We observe that the signal patterns generated by the diagonal path are significantly varied so it can affect the accuracy of location estimation.

In the third scenario, a participant walks from cell 1-1 to 3-2. Next, he walks down from cell

3-2 to 6-4. Then, he continues walking from cell 6-4 to 4-3. Then, he takes another turn from cell 4-3 to 7-3. Finally, he walks from cell 7-3 and finishes at cell 9-4. In this scenario, we include a few shorter paths to introduce the partial signal patterns because we want to evaluate how well deep learning models can handle the partial pattern.

In the fourth scenario, a participant walks from cell 1-1 to 3-2. Next, he walks diagonally from cells 3-2 to 5-3. Then, he walks from cell 5-3 to 6-4. Then, he walks from cell 6-4 to 9-4. Finally, he walks from cell 9-4 and finishes at cell 7-3. In this fourth scenario, we not only include the partial pattern but also introduce some unexpected patterns generated by a person turning to change his direction under the sensor node. For both datasets, we record the ground truth using a video camera using 60 frames per second and the PIR sensors' signals are synchronized with each recorded frame using a local PC desktop time.

3.3.3 Data Description

Our dataset has 36 classes in total. Each class label represents a location of a cell in a monitored area. For example, cell 1-1 is presented as label 11 represents, cell 3-2 is presented as label 32. Training data contains 1020 sequences of raw PIR output, and each PIR sequence in training data has different lengths depending on the walking distance. Testing data comprises four scenarios with more than 2,000 timesteps. In addition, all PIR sequences have five channels for both training and testing data. The amplitude of raw PIR signals ranges from 0 to 5 volts. When a PIR sensor detects no movement, the output voltage hovers around 2.05 volts. To train machine learning models, we can either utilize raw data directly or extract features such as standard deviation, root mean square, peak to peak amplitude, etc.

3.3.4 Data Format

For this dataset, we focus on a sequence-to-sequence (s2s) classification. To construct our profile data, we divide all collected PIR samples based on walking directions such as left-right, top-down, etc. For each direction, we crop and align the samples using a cross-correlation method because machine learning or deep learning model can learn the discriminative features from the samples correctly if the samples are aligned properly [197].

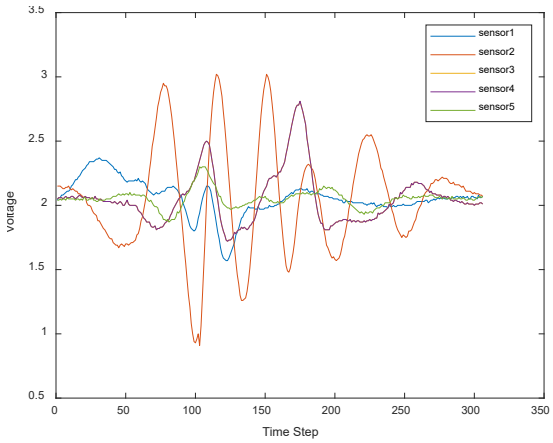


Figure 3-7 A sample of PIR analog signal captured by a PIR node

From one PIR sensor, a collected data is a sequence of output voltage $VT = \{vt_1, vt_2, vt_i, \dots, vt_n\}$ where vt_i represents the *i*th voltage value in a sequence and n is a length of a sequence. We require data sequences from multiple PIR sensors to complete one sample for each walking direction $VT_{sample} = \{VT_1, ; VT_2; \dots ; VT_N\}$ where N is the total number of PIR sensors. In this chapter, we use 5 PIR sensors to collect data and an example of analog signals is shown in Figure 3-7. Therefore, one sample is a 2-dimensional matrix with a size of $5\ x\ n$ where n is a length of a sequence. Next, a sample is associated with a ground truth $L = \{l_1, l_2, l_i \dots\ l_n\}$ where l_i denotes the *i*th label in a sequence and n is a length of a sequence.

For test data, we use the same format as the profile data. Therefore, the test data is also a matrix with a size of $5\ x\ n_{test}$ and is associated with a ground truth with a size of $1\ x\ n_{test}$ where n_{test} is the length of the test sequence. Our dataset is available and ready to download at https://github.com/KanNgamakeur/PIR-localization-dataset

3.4 Test methods

3.4.1 Deep learning methods for classifying a location

For image classification, there are many state-of-the-art architectures such as Alexnet, Inception, ResNet, VGG, etc [198]. For the time series classification, we observe that there is no such gold standard architecture because the characteristics of time series data vary from application to application. Therefore, we review several related papers and find that the reviewed papers

commonly adopt an architecture based on Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), or a combination of CNN and RNN [12, 34, 199-201]. Therefore, we evaluate our dataset using the above-mentioned architectures.

For CNN-based architectures, we test a 1D Convolutional Neural Network (1D-CNN) that has been modified from the standard convolutional neural networks for the image classification (2D-CNN) to operate on 1D signals. In general, 1D CNNs consist of 1D convolution layers that extract discriminative features by sliding a set of 1D filters over the entire sequence of signals and fully connected (dense) layers that use the extracted features to perform a classification task. For the past years, several studies reported that 1D-CNNs give good results, are fast to train, and require low computational power [27]. We also test Temporal convolutional networks (TCNs) which are a variant of convolutional neural networks. They are designed to handle sequential data similar to recurrent neural networks (RNNs). TCNs use causal convolutions to convolve an output at time t with inputs up to time t in the previous layer [202]. However, they require many layers or larger filters to achieve the long effective history sizes. Therefore, dilated convolutions and residual connections are incorporated to address these problems without sacrificing the performance and stability of the networks. Several studies showed that TCNs' performance can match the performance of recurrent networks or even outperform them [203].

For RNN based architectures, we test three recurrent networks including Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU). LSTM can regulate the flow of information entering or exiting a memory cell using the input gate, output gate, and forget gate. These gates enable LSTM to choose what information to be maintained and removed. Thus, it can handle long sequences better than a traditional RNN [12]. Bi-LSTM can use both past and future information to achieve a better prediction for the sequence classification because it consists of two LSTMs that are trained using an original input sequence and a reversed copy of the input sequence [34]. GRU is a lightweight version of LSTM because a cell state is removed and there are only two gates including a reset gate and an update gate. As a result, GRU is faster to train compared with LSTM [204]. We observe that three RNNs can be used as a single layer or can be stacked on top of another. Therefore, we include single-layer RNNs and two stacked RNNs in our tested architectures.

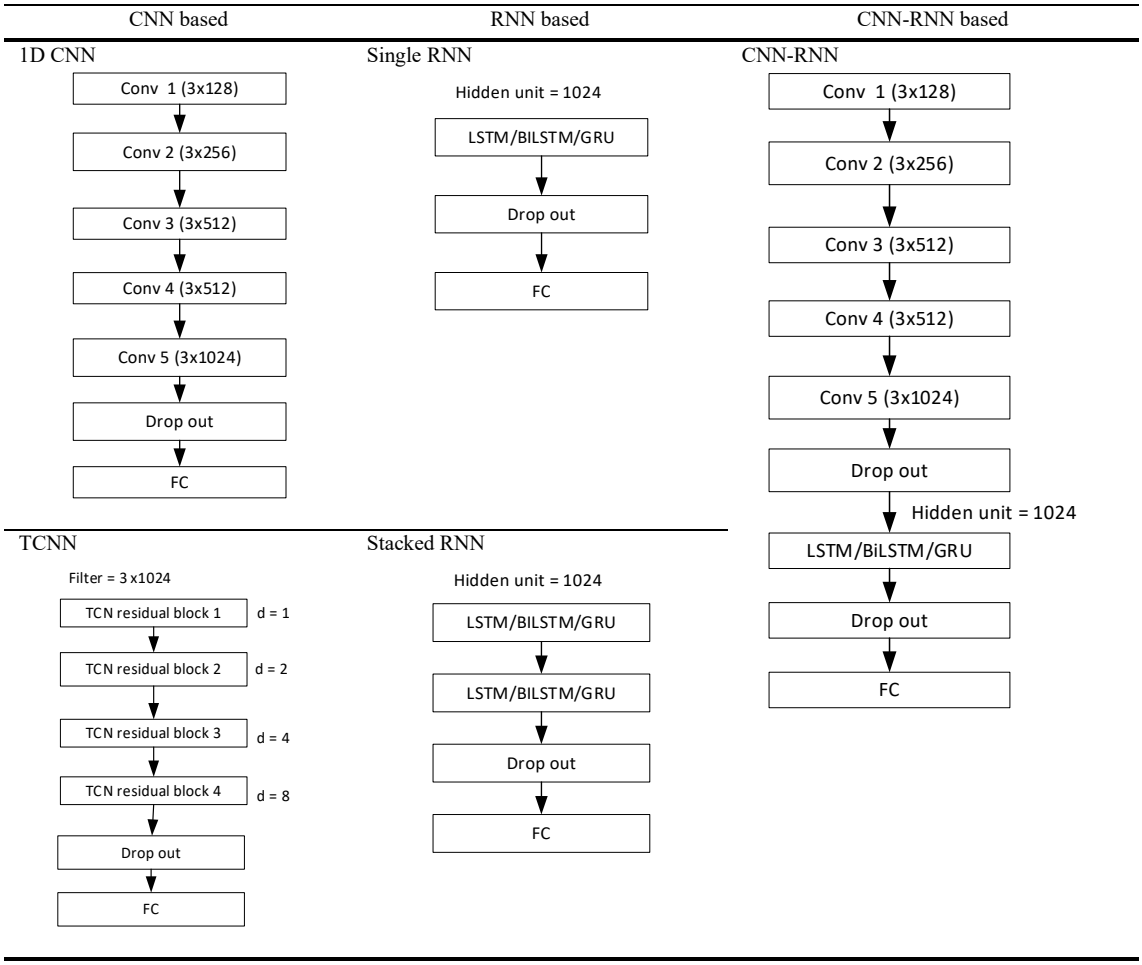


Figure 3-8 The configuration of deep learning architectures for classification

CNN-RNN based architectures are composed of 1D Convolutional Neural Network (CNN) layers and Recurrent Neural Networks (RNNs). 1D-CNN layers perform the feature extraction on input data. Then, extracted features are fed to an RNN layer that is capable of learning temporal dependency between extracted features. The hybrid models have been employed successfully in various applications such as activity recognition [205], Natural language processing [206], speech recognition [207], etc. We test CNN-LSTM, CNN-BiLSTM, and CNN-GRU respectively.

The configurations of our tested architectures are listed in Figure 3-8. 1D CNN consists of 5 CNN layers with filter sizes of 128, 256, 512, 512 and 1024. Each CNN layer performs a 1D convolution operation on PIR signals using a filter size of 3 and a stride length of 1, followed by batch normalization and dropout. Next, the normalized output is passed to the Relu activation function. Finally, a 1D max pooling with a window size of 2 and a stride length of 1 is performed. For both convolution and max pooling, a zero-padding is applied to maintain the

length of the input sequence. For TCNN, we set the number of filters to 1024 for all TCNN blocks. Each block uses a filter size of 3 and a stride length of 1. The dilation factors are set to 1, 2, 4 and 8 respectively. For RNN based, we set the number of hidden units to 1024 for both single and stacked RNN. For CNN-RNN, we use a similar CNN architecture as 1D-CNN architecture and use a single RNN with 1024 hidden units. For all tested architectures, we set a dropout factor to 0.3 and a fully connected layer uses 36 neuron units to generate weighted sum outputs for 36 classes. Lastly, the Softmax function is used to convert the outputs into probabilities that sum to one

To train all tested models, we use 10-fold cross validation to ensure that the trained models are not underfitted or overfit. We employed an ADAM optimizer that is more effective than other optimizers such as RMSprop or SGD. For each tested architecture, we test hyperparameters empirically and use the optimal ones to train deep learning models. All models are implemented using Keras library and trained on i5 3.2Ghz Intel quad-core CPU, Ram 16 Gb, and a GPU of NVIDIA RTX2070 with 8Gb video memory.

To train all tested models, we use 10-fold cross validation to ensure that the trained models are not underfitted or overfit. We employed an ADAM optimizer that is more effective than other optimizers such as RMSprop or SGD[208]. Cross-entropy is chosen as a loss function for this multiclass classification. We train each tested architecture for 100 epochs and set a mini-batch size to 32. A learning rate is set to 0.0001. To reduce an overfitting issue, L2 Regularization is set to 0.0001. All models are implemented using Keras library and trained on i5 3.2Ghz Intel quad-core CPU, Ram 16 Gb, and a GPU of NVIDIA RTX2070 with 8Gb video memory.

Once all models are trained, test data is inputted to these trained deep learning models to classify the locations of a person. However, our test data contains long sequences of analog signals that are generated by a PIR sensor node when a person walks or stops in different walking scenarios. To localize a person using PIR sensors, we are only interested in parts of the signals that are generated by the movements of a person. As a result, we need to perform movement detection. A threshold method is commonly used to indicate a movement or presence of a person because it is a simple yet effective method[25, 79, 81]. This detection method relies on a threshold value to determine whether signal events are caused by a movement of a person or not.

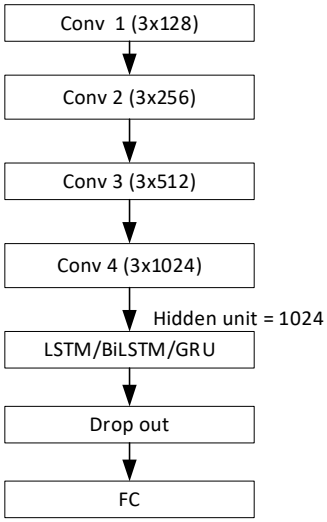


Figure 3-9 The configuration of deep learning architecture for cartesian coordinate estimation

We set a threshold value at 2.1±0.2 volts to select only subsequences that pass a threshold value from test data. As a result, these subsequences contain only PIR analog outputs that are generated by a person's movements. Thus, we can feed the proper subsequences into the trained model to classify person's locations.

3.4.2 Location estimation in cartesian coordinates

We not only want to test our dataset to estimate locations as a classification problem, but we also want to demonstrate an alternative way to utilize our dataset by estimating locations as cartesian coordinates. To achieve this idea, we first need to convert location labels into 2d coordinates for both training data and test data. We simply use the center coordinate of each cell as ground truths in our data. Then, we use deep learning architectures and particle filtering to estimate 2d coordinates.

For the deep learning architecture, we employ CNN-RNN based architecture to estimate cartesian coordinates including CNN-LSTM, CNN-BiLSTM, and CNN-GRU in the demonstration because we have tested a variety of CNN based, RNN based, and a hybrid CNN-RNN based architecture and found that a hybrid one achieves the best results. The architectures of CNN-RNN, shown inFigure 3-9, have 4 CNN layers and a single RNN layer. We set the number of filters in the CNN layer to 128, 256, 512 and 1024. Similarly, we use a filter size of 3 and move the filter across the signal using a stride length of 1. We use the same structure of the

CNN layer as in the 1D CNN model for classification. i.e., each CNN layer is followed by a batch normalization, dropout, Relu and max pool. In addition, a dropout layer is added before a fully connected layer to reduce overfitting. The value of dropout factor is set to 0.3. Instead of outputting only one label like in classification, a fully connected layer produces two outputs which are x and y coordinate respectively. Therefore, we set the number of neuron units to 2. We train the deep learning models with the ADAM optimizer and use the same hyperparameter setting as in the previous subsection. For this regression task, we select Mean Squared Error (MSE) loss function to be minimized by ADAM optimizer.To test the trained models, we follow the same method as described in the previous section to select subsequences that are generated by movements of a person and feed them to the trained models to estimate locations of a person in 2d coordinates.

For Particle filtering, we select this algorithm because it can represent an arbitrary probability distribution and can handle nonlinear processes and measurement models [169]. To use particle filtering with our dataset, we adopt an azimuth change measurement proposed in [32]. According to their paper, we implement a system model that expresses a state transition of a moving person and an observation model that describe a relationship between a person's movement and observed azimuth measurement. For each loop of particle filtering, a non-overlapped window of length 60 is slid across test data and a threshold method is used to check if there is a person's movement or not. If yes, an azimuth change measurement is determined from PIR data in the sliding window. Then, we use it to correct the estimated state i.e., a location of a person.

3.4.3 Evaluation metrics

To validate the performance of classification models, five performance metrics are implemented. The first commonly-used metric is the overall accuracy (Acc) which is the proportion of correct location estimations in the whole dataset:

$$Acc = \frac{TP}{Total\ number\ of\ locations}$$

where TP (true positive) is the total number of the correctly classified locations. However, we cannot rely on overall accuracy to determine the performance of the classification models

because the calculation of the overall accuracy may be influenced by an imbalanced dataset. To give a fair comparison between the tested models, we employ the macro averaged recall, the macro averaged precision, the macro F1 score, and The Cohan's Kappa Score (Kappa) because these additional metrics can provide the performance analysis of all tested models from different perspectives. The macro averaged recall can measure the overall ability of the tested models to identify all the true positives for each location, while the macro averaged precision can determine the overall trustworthiness of the models when they predict a particular location as true positive. Both macro recall and macro precision are calculated by averaging the recall and the precision of each location as follows:

$$macro\_Re = \frac{\sum_{i=1}^{n} Re_i}{n}$$

$$macro\_Pr = \frac{\sum_{i=1}^{n} Pr_i}{n}$$

where Re_i and Pr_i are the recall and the precision of the i^{th} location, and can be calculated as $Re_i = TP_i/(TP_i + FN_i)$ and $Pr_i = TP_i/(TP_i + FP_i)$ respectively. We denote TP_i, FN_i, and FP_i as a true positive, false negative, and false negative of i^{th} location accordingly. We also use the macro F1 score which is the harmonic mean of Macro-Precision and Macro-Recall. The macro F1 can provide a performance indicator that takes precision and recall into account and can be calculated as follows:

$$Macro\_F1\ score = 2 \times \frac{macro\_Pr \times macro\_Re}{macro\_Pr + macro\_Re}$$

where $macro\_Pr$ and $macro\_Re$ are the macro precision and the macro recall respectively. In general, the calculated F1 score is in the range from 0 to 1. A higher macro F1 score indicates that a model has low false positives and low false negatives for all classes. Otherwise, a model performs poorly when the F1 score is low. The Kappa score aims to measure the agreement between a model's prediction and ground truth. Compared with the F1 score, it can provide a better understanding of the performance of the tested models for both multiclass and imbalance

class problems. The Kappa score can be calculated as follows:

$$Kappa = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the observed agreement and P_e is the expected agreement. Kappa values are in the range of -1 and 1. A model's prediction completely agrees with a ground truth when a kappa score is 1. On the other hand, the value is less than or equal to 0 when they do not agree.

For 2d coordinate estimation, we employ two popular metrics including mean distance error (MDE) and cumulative distance function (CDF) of distance errors. The mean distance error can be calculated by averaging the Euclidean distance between the estimated coordinates and the ground truth coordinates as follows:

$$MDE = \frac{\sum_{i=1}^{N} \sqrt{(x_e - x_g)^2 + (y_e - y_g)^2}}{N}$$

where (x_e, y_e) is the estimated coordinate, (x_g, y_g) is the ground truth coordinate, and N is the total number of locations. The cumulative probability function of a distance error is normally used to measure the precision of location estimation. It describes the probability that the distance error is within a particular value and can be expressed as:

$$F(x) = \int_{\infty}^{x} f(t)dt$$

where f is the probability density function.

3.5 Results and Discussion

3.5.1 Cell-level classification results

The results of all tested models are shown in Table 3-1 and Table 3-2. The average accuracy of all models is 77%. For CNN based model, the performance of the 1D-CNN model is the lowest in all metrics compared with the other models. Although the accuracy and kappa score of 1D-CNN is around 71% and 0.694. However, the model suffers from a low f1 score, precision, and recall which are 0.522, 0.540, and 0.538 respectively. We observe that false-positive locations that are not part of ground truth locations are the cause of low f1 score, precision, and recall especially for scenarios 2 and 4. The TCNN model can perform equal to or even outperform some RNN models such as LSTM and GRU. It achieves an accuracy of 78% and a kappa score of 0.768.

Table 3-1 Overall performance for all the tested DL architectures

	Accuracy	Macro f1 score	Macro Recall	Macro Precision	Kappa
1D CNN	0.710	0.522	0.538	0.540	0.694
TCNN	0.780	0.631	0.660	0.650	0.763
LSTM	0.743	0.605	0.626	0.613	0.729
BILSTM	0.844	0.700	0.722	0.690	0.835
GRU	0.754	0.549	0.567	0.583	0.740
Stacked LSTM	0.74	0.578	0.595	0.592	0.724
Stacked BILSTM	0.847	0.727	0.751	0.726	0.840
Stacked GRU	0.722	0.578	0.606	0.579	0.707
CNN-LSTM	0.767	0.637	0.656	0.661	0.754
CNN-BiLSTM	0.849	0.770	0.812	0.760	0.841
CNN-GRU	0.734	0.606	0.628	0.630	0.719

Table 3-2 Detailed performance results for all the tested DL architectures

	Accuracy				Macro f1 score				Macro Recall				Macro Precision				Kappa			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1D CNN	0.83	0.70	0.79	0.56	0.68	0.51	0.70	0.39	0.69	0.53	0.71	0.43	0.69	0.54	0.71	0.42	0.82	0.68	0.78	0.53
TCNN	0.87	0.85	0.83	0.61	0.79	0.75	0.84	0.44	0.81	0.76	0.86	0.46	0.78	0.78	0.84	0.53	0.86	0.83	0.82	0.59
LSTM	0.86	0.81	0.78	0.58	0.77	0.74	0.79	0.46	0.79	0.74	0.81	0.53	0.80	0.76	0.80	0.49	0.85	0.80	0.77	0.56
BILSTM	0.91	0.86	0.88	0.75	0.91	0.82	0.85	0.62	0.92	0.80	0.86	0.65	0.90	0.85	0.85	0.61	0.91	0.85	0.85	0.73
GRU	0.91	0.80	0.77	0.60	0.82	0.64	0.79	0.43	0.83	0.64	0.82	0.45	0.81	0.68	0.80	0.46	0.91	0.79	0.76	0.57
Stacked LSTM	0.91	0.80	0.78	0.55	0.78	0.73	0.75	0.41	0.78	0.73	0.78	0.45	0.79	0.79	0.75	0.46	0.90	0.78	0.77	0.52
StackedBiLSTM	0.91	0.87	0.87	0.77	0.85	0.86	0.88	0.60	0.85	0.84	0.89	0.65	0.88	0.87	0.90	0.61	0.90	0.86	0.86	0.75
Stacked GRU	0.89	0.74	0.79	0.52	0.81	0.61	0.76	0.38	0.82	0.61	0.78	0.43	0.81	0.63	0.78	0.40	0.88	0.72	0.78	0.50
CNN-LSTM	0.90	0.74	0.86	0.61	0.81	0.62	0.83	0.48	0.82	0.63	0.85	0.50	0.81	0.63	0.83	0.54	0.90	0.72	0.85	0.58
CNN-BiLSTM	0.96	0.87	0.87	0.75	0.95	0.83	0.85	0.66	0.96	0.82	0.86	0.70	0.95	0.87	0.86	0.67	0.95	0.86	0.86	0.73
CNN-GRU	0.90	0.72	0.81	0.57	0.77	0.51	0.83	0.46	0.77	0.51	0.86	0.50	0.76	0.54	0.84	0.48	0.89	0.69	0.80	0.54

In addition, the f1 score, precision, and recall are better than the 1D-CNN model because the dilated causal convolution allows features to be extracted from the past time step up to the current time step without any leakage of future data and can cover a long data sequence. Thus, it performs better than 1D-CNN to model sequential data. TCNN achieves fairly good results in scenarios 1 to 3 but performs poorly in scenario 4 because there are a considerable number of

false negatives for cell 53 where a person changes his direction under a sensor node.

For RNN based model, LSTM and GRU achieve accuracy of 74% and 75% respectively but obtain a low recall, precision, and f1 score which is around or less than 0.6 as shown in Table 3-1. LSTM and GRU achieve satisfactory outcomes for scenarios 1 to 3 but obtains poor results in scenario 4 for the similar reason as TCNN. BiLSTM achieves an accuracy of 84% and a kappa score of 0.835 which are the best results compared with that of LSTM and GRU. The f1 score, precision, and recall are around 0.70. the use of past and future information allows a better understanding of the time dependency of PIR outputs. Thus, misclassified locations are reduced for scenarios 1 to 4. When two RNN networks are stacked, we can see that it is not much different in all metrics for all stacked RNN models except stacked BiLSTM which has an improvement in f1 score, precision, and recall. For scenarios 1 to 3, stacked BiLSTM can perform equally or slightly better than BiLSTM. For scenario 4, Stacked BILSTM can improve the number of true positives for cell 53 compared with BiLSTM but the f1 score, precision, and recall do not improve because the number of the misclassified locations that are not part of ground truth locations is relatively unchanged.

For CNN-RNN based model, CNN-BiLSTM obtains an accuracy of 85% and the kappa score of 0.841. In addition, precision, recall, and f1 score are over 0.75. CNN-BiLSTM can perform better than stacked BiLSTM in scenario 1 while its performance is nearly identical to or slightly lower than stacked Bi-LSTM in scenarios 2 and 3. In scenario 4, CNN-BiLSTM achieves a better f1 score, precision, and recall than stacked BiLSTM because the CNN-BiLSTM model obtains perfect precision and recall for more locations and has fewer misclassified locations. The overall performance of CNN-LSTM and CNN-GRU is slightly higher than the stacked version of LSTM and GRU. For scenarios 1,3 and 4, CNN-LSTM and CNN-GRU have a better performance than stacked LSTM and stacked GRU. However, CNN-LSTM and CNN-GRU perform worse than stacked LSTM and stacked GRU in scenario 2. We observe that both CNN-LSTM and CNN-GRU entirely misclassify some locations such as cell 33 as other locations. The extracted features may be similar between these cells but LSTM and GRU can use only the past information to determine locations. Thus, there is not enough information to make a correct estimation.

Based on all result, architectures containing BiLSTM performs better than other RNNs.

Scenario 4 is quite problematic for all architecture due to random patterns of PIR output.

3.5.2 2D Coordinate estimation results

The results of 2D Coordinate estimation are shown in Table 3-3 and Figure 3-10. CNN-LSTM obtains the mean distance error of 0.2359 m and around 75 % of the distance error is less than 0.4 meter, which is the best results compared with other methods. For CNN-BiLSTM, it achieves 28% lower in the mean distance error and 70% of the distance error is less than 0.4 meters. BiLSTM can use past and future information to improve understanding of sequence data.

Table 3-3 Mean distance error and standard deviation for all the tested methods

	Overall		Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Mean	Standard Deviation	Meam	Standard Deviation	Meam	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
CNN-LSTM	0.2359m	0.2255m	0.1388m	0.0970m	0.2005m	0.1546m	0.1860m	0.1383m	0.3708m	0.3129m
CNN-BiLSTM	0.3131m	0.2812m	0.2497m	0.1411m	0.2466m	0.1439m	0.2548m	0.1660m	0.4524m	0.2740m
CNN-GRU	0.5198m	0.2189m	0.4548m	0.2638m	0.5576m	0.3023m	0.4628m	0.2713m	0.5932m	0.2661m
Particle Filter	0.5482m	0.3265m	0.4230m	0.2083m	0.5429m	0.3268m	0.5652m	0.3984m	0.6617m	0.3049m

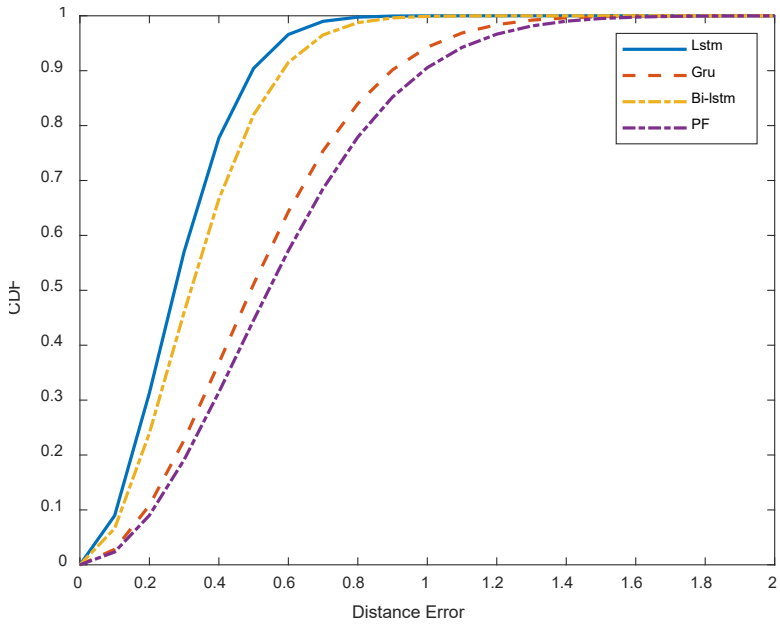


Figure 3-10 CDF of the distance error for all tested method

This characteristic of BILSTM greatly improves the accuracy of certain applications such as language translation and speech recognition. However, it seems that we cannot fully take advantage of this characteristic in this case. CNN-GRU performs considerably worse than both CNN-LSTM and CNN-BiLSTM. It achieves the mean distance error of 0.5198 meters which is 75% higher than CNN-LSTM and 49% higher than CNN-BiLSTM. CDF of CNN-GRU also indicates that around 70 % of the distance error is within 0.6 m. The bad performance of CNN-GRU may be due lack of a cell state. Particle filtering obtains the worst results compared with the deep learning methods. The mean distance error of the particle filter is 0.5482 meters, and its CDF is slightly worse than GRU. The azimuth change measurement is quite ambiguous in some locations in a monitored area which affects the accuracy of the location estimation. From our observation, the ambiguity of azimuth change measurement causes particles to converge at the wrong location and there is no particle at the other locations in a monitored area. Thus, an estimated location cannot be corrected in the correction step. To mitigate this issue, we reinitialize 10% of particles and place them uniformly across the monitored area before we correct an estimated location. According to the above results, features extracted by deep learning methods can provide a better discriminative power than the handcrafted feature used in the particle filter.

3.5.3 Discussion

Although the feasibility of our proposed system is proved and the performance of deep learning models is quite promising, several challenges need to be explored.

Signal ambiguity: Although the deep learning models are capable of extracting features automatically, they do misclassify some locations due to the ambiguity of the extracted features or patterns. It may be necessary to include some extra handcrafted features to improve the differentiation of extracted features and patterns in different locations.

Infeasible location: Location estimations in some time steps can result in infeasible locations. Therefore, it would be preferable to include the motion information into the classification models to enhance their accuracy and reduce infeasible estimations.

Unexpected signal patterns: It is not feasible to collect every combination of PIR output signals because it is time-consuming and requires huge human efforts. It is preferable to have

models that can handle unexpected signal patterns caused by some events, such as a person turning or a person standing still, to some extent to reduce errors in the location estimation.

Length of sequence: We use a full sequence of our test data to evaluate the feasibility of our dataset for deep learning models and the results are quite positive. However, sensor data is sent partially from time to time. Partial data can affect the performance of the deep learning model. Thus, this challenge needs to be further explored.

Possibility for Other methods of location estimation: so far, we only explore location estimation using sequence-to-sequence classification, cartesian coordinate, and particle filter. However, there are a few other possibilities to be explored to use our dataset for location estimation. One can covert PIR signals to images using methods such as Gramian Angular Field (GAF) and Markov Transition Field (MTF) [209]. Then, it is possible to apply the state-of-the-art deep learning models for the image classification such as ResNet, VGGnet, etc.

Scalability of the proposed system: we focus on only a small-scale deployment of sensor nodes to evaluate the feasibility of our proposed system in this work. To improve the scalability of our proposed system. we need to increase the number of installed sensor nodes. It is a common practice to have a sink node to collect data from different sensor nodes. However, there are some factors to be considered such as a reporting frequency, a communication protocol, time synchronization of sensor nodes, etc. These factors can affect our systems in terms of energy consumption, localization accuracy, and real-time capability. A trade-off between these factors must be further explored.

3.6 Conclusion & Future work

In this chapter, we propose a Passive Infrared sensor system and a dataset for indoor localization and tracking. Our sensor node is developed based on commercial-off-the-shelf sensors. The dataset consists of 1,000 walking samples, 4 scenarios, and 36 classes in total. We evaluate our dataset using different deep learning methods such as CNN, RNN, and CNN-RNN. The results confirm the feasibility of our dataset and show that the best deep learning method can achieve an accuracy of 85%, a kappa score of 0.84, and an f1 score of 0.77. Furthermore, our dataset can be converted for a problem of coordinate estimation. To validate this idea, we apply both deep learning methods and the particle filter to estimate coordinates. The best

performer achieves a distance error of 0.25 meters.

In the future, we aim to improve the scalability of the proposed system. In addition, we will increase the variety of our dataset. more scenarios will be included, especially with multiple persons. Different topologies of sensor deployment will be included.

Chapter 4 Deep CNN-LSTM Network for Indoor Location Estimation using Analog Signals of Passive Infrared Sensors

Indoor localization is a crucial component of IoT applications in many areas such as healthcare, energy management, and security control. Passive Infrared (PIR) sensor has been employed for a location estimation due to its cost-effectiveness, low power consumption, and low electromagnetic interference. Compared with its binary output, PIR analog output which is an output voltage generated by a PIR sensor when its sensing elements detect changes in temperature in an environment can provide more information regarding a person's location. However, only a few works focus on using analog signals for location estimation. During the past several years, deep learning approaches have emerged and achieved outstanding results in many applications. In this chapter, we harness the power of deep learning and propose a deep CNN-LSTM architecture for PIR-based indoor location estimation. In our architecture, an upper CNN network can extract features from PIR analog output automatically while a lower LSTM network can learn temporal dependencies between the extracted features. To evaluate the feasibility and performance of our proposed method, we conduct four different sets of experiments. Our results show that the proposed method can efficiently handle complex cases and can achieve the mean distance error of 0.23 meters, and 80% of distance errors are within 0.4 meters.

Chapter 4 has been accepted by IEEE Internet of Things Journal.

4.1 Introduction

The fast improvement of IoT devices has empowered many useful indoor applications in areas such as healthcare, energy management, and security control [3, 210]. To support such applications, accurate person's whereabouts are also crucial and need to be reported so that information such as medical alerts can be analyzed and handled immediately by interested parties [6, 37]. Various surveillance technologies have been employed to enable location estimation in an indoor environment. Wearable or mobile devices, such as smartwatches and mobile phones, allow us to estimate the location of a person by determining the current location of the device. However, it is not always user-friendly for a person to carry a wearable or mobile device around a premise because it is not comfortable, and there is a risk of device loss [8]. On the contrary, these drawbacks can be overcome using radio frequency (RF), WiFi, and sensor devices, such as RFID tags, surveillance cameras, passive infrared sensors, and electric field sensors. When privacy is concerned in certain indoor areas, passive sensing devices and sensors, such as passive infrared and WiFi, are commonly adopted because they sense the presence of a person without recording nor revealing their identities or appearance.

Major advantages of Passive Infrared (PIR) sensors include a fast response time, low power usage, and strong tolerance against electromagnetic interference [35]. Despite these advantages, PIR sensors have relatively low to moderate coverage distance, and their detection accuracy can be affected by a rise and drop in temperature in a surrounding environment. PIR sensors have been employed successfully to locate or track the location of a person. However, most existing work only focuses on the digital outputs of PIR sensors due to their low computational cost. A person's location can be either estimated by analyzing overlapped areas of the activated PIR sensors or located in the proximity of the activated sensors, however, a sophisticated sensors node or dense sensors deployment is needed [1, 18, 22, 211]. As a result, their complicated installation makes it impractical in a real-world setting. On the other hand, as characteristics of PIR analog outputs are influenced by factors such as the distance between a person and a PIR sensor, walking speed, and walking direction [24], it is possible to leverage features of the analog outputs to estimate a person's location. Commonly used filtering algorithms such as Kalman filtering and Particle filtering can be employed to estimate a location based on

characteristics of PIR analog outputs. However, there is not sufficient information on the specifications of commercial off-the-shelf (COTS) PIR sensors to formulate an observation model that describes a relationship between a distance and a PIR analog output [1]. To obtain a specification of a PIR sensor, we require an extensive experiment and simulation. In traditional machine learning methods, one of the problems is to find out the appropriate features for each application which requires us to extract and test features manually. As a result, it is labor-intensive and time-consuming [157]. In addition, analog signals are quite ambiguous. To obtain accurate location estimation, we need to know about the dependency between locations and raw PIR outputs in each time step. However, traditional machine learning algorithms cannot capture this complex dependency very well [12]. Thus, they produce inaccurate estimations. Therefore, it is quite challenging to use analog outputs for location estimation. To the best of our knowledge, only a few works focus on the analog outputs of PIR sensors [30, 32].

During the past several years, various deep learning techniques have been developed and achieved promising results to address the localization problem based on RF signals [12, 157, 184]. Compared with traditional machine learning approaches, deep learning methods can automatically extract discriminative features from raw data without the intervention of domain experts and can learn complex data representations through hierarchical multi-layers even though there are small training samples (e.g., less than 1,000 samples of data [212, 213]) and various characteristics of data[26]. As a result, the performance of the deep learning methods for localization problems generally outperforms traditional machine learning methods[157, 214]. However, some machine learning methods such as Support Vector Regression (SVR) may achieve better performance due to small datasets, less complex problems or requiring extensive feature engineering [96, 214].

In this chapter, we aim to leverage deep learning to address indoor location estimation based on analog signals of PIRs. We propose a two-phase framework for indoor location estimation which consists of offline and online phases. In the offline phase, raw PIR outputs are preprocessed and used as a training dataset. The offline phase consists of our CNN-LSTM architecture, which can process the raw outputs directly to estimate the location of a person. To avoid the need for manual feature extraction, we design the upper CNN level to automatically extract discriminative features from raw PIR outputs. Then, we incorporate an LSTM network in

the lower level of the proposed architecture to capture the sequential correlation of the extracted features from the upper layer. As a result, the dependencies between locations and raw PIR outputs can be modelled appropriately. In the online phase, we adopt a threshold method to detect a person's movement. However, PIR signals are quite noisy and contain non-symmetrical positive and negative phases. Thus, it is hard to set a threshold value properly. To address this issue, we use instantaneous energy instead of raw PIR outputs to simplify threshold value setting and reduce false detection. Once the CNN-LSTM model has been trained to estimate the location of a person, the median absolute deviation method [215] is employed to detect and filter out infeasible estimated locations to enhance the estimation accuracy.

Our main contributions are summarized as follows:

- We present a two-phase framework for indoor location estimation using PIR analog outputs.
- We propose a deep CNN-LSTM architecture that can extract discriminative features from PIR analog outputs and learn temporal patterns of the extracted features to estimate a person’s location.
- We conduct a comprehensive set of experiments mimicking possible real-world scenarios to evaluate the feasibility and effectiveness of our proposed approach

The rest of this chapter is organized as follows. Section 4.2 discusses the related works. Section 4.3 describes the problem formulation. The proposed methodology is presented in Section 4.4. The experiments and their results are reported in Sections 4.5 and 4.6. Finally, Section 4.7 provides some concluding remarks.

4.2 Related works

To realize indoor location estimation, different technologies are employed such as radio frequency (RF), surveillance cameras, and infrared sensors. In this section, we focus our discussion only on the classes of location estimation techniques that can be applied to the PIR technology.

Proximity techniques are the simplest location estimation method. We require a dense sensor deployment in the monitored area where the position of each sensor is known. A threshold method is mainly used to identify affected sensors caused by a person standing or moving nearby sensors. Then, the location of a person can be estimated to be nearby or at the position of an

activated sensor. This type of technique can be used with RFID [58, 130], Zigbee/IEEE 802.15.4 [43], and PIR [21, 22]. Due to a dense sensor deployment, it is rather impractical for deployment in a large area.

Angle of arrival (AOA) based techniques measure angles of arrival of each sensor node, and lines can be drawn w.r.t. the measured angles. Then, the location of a person can be determined from the intersection of these lines. Wang, et al. [7] exploited a multipath problem of RFID tags and extracted AOA using a MUSIC algorithm. Li, et al. [134] measured the AOA of a Wi-Fi signal using the Dynamic Music algorithm. Adib and Katabi [10] employed smoothed music algorithm to extract the AOA of Wi-Fi signals reflected from a moving person to localize a person through walls. Some PIR works utilize binary PIR outputs to measure an angle of detection. Qi, et al. [18] developed two types of infrared modules consisting of 4 PIR sensors in 2 columns. Type I module, which has a field of view of each sensor separated by 16 degrees, is specialized for locating multiple persons. Type II module, which has a field of view of each sensor covered by pseudo-random coded masks, is designed for target identification. They employed a Bayesian tracking framework to track each person while Hidden Markov Model (HMM) was used to identify a person. Yang, et al. [33] created 12 detection zones around a sensor node by overlapping the field of view (FOV) of six PIR sensors. The detection zones are used to determine the detection angle, and a person's location can be determined based on the intersections of these angles. However, their method may produce false intersections. To overcome this drawback, the authors employed a naïve Bayes classifier [174] or artificial neural networks [164] to scope down estimated locations to possible sub-areas.

Model-based techniques can estimate a person's location using mathematical models that describe relationships between the positions of a person and sensor outputs. For the RF technology, two types of RF phenomena including the shadowing effect and diffraction are employed to formulate mathematical models. Shadowing models attempt to describe RF attenuation caused by a person obstructing RF links. There are various variations of shadowing models such as the excess path delay model [53, 135], elliptical model [66, 185], Exponential-Rayleigh model (ERM) [138], Saddle surface model (SaS) [139] and imaging [216]. Diffraction models aim to characterize RF signals that are diffracted by a side of a human body, and a few works focused on models based on the diffraction effect [63, 142]. Recent work in PIR

technology also adopted a model-based approach. The PIRATES localization system was proposed in [32]. This localization system utilized a new azimuth change measurement which is the absolute difference in the azimuth of a person with respect to a PIR sensor at two locations. A mathematical model that describes a relationship between azimuth change and a person's location was presented, and location estimation is performed using a particle filter technique.

Fingerprinting methods are commonly used techniques for radio frequency technology, especially Wi-Fi. It was proposed by [217] to locate a person who carries an RF device. Later, Moussa and Youssef [218] leveraged the received signal strength (RSS) measurement of existing Wi-Fi infrastructure and were the first to adopt a fingerprinting method in a device-free manner. Based on this initial work, Youssef and Agrawala [187] introduced a Horus system that uses a center of mass technique with a time-averaging window to improve the localization's accuracy. Then, the fingerprint method has been further improved in [155] to support a large-scale area. Non-linearity of the impact subjects on RSS is addressed in [106, 158] to minimize time and efforts for fingerprint collection. Channel State Information (CSI) was initially investigated in [219] to locate a person with a mobile Wi-Fi device because it is more stable and accurate compared with RSS. Using CSI amplitudes, a probabilistic method was used in [79] and [220] to map the fingerprints, while SVM was employed in [96]. For PIRs, some works used techniques similar to the fingerprint methods. PIR analog outputs were initially explored in [76] to estimate the distance of a person that walks past two PIR sensors attached to opposite walls of a hallway. Distances between the opposite sensors are grouped into three classes including close-to-sensor 1, middle, and close-to-sensor 2. Then, they employed three classifiers, including Naive Bayes, Support Vector Machines (SVM), and k-Nearest Neighbor (k-NN), to classify the distances. The characteristics of PIR analog outputs were investigated extensively in [24]. This work found that the amplitude, phase, and frequency of PIR analog outputs can be affected by the distance, speed, and direction of a moving person. Some localization experiments are performed to validate the findings in this work using the same classifiers as in [76].

Recent advances in deep learning have shown several advantages for location estimation such as improved accuracy, automatic feature extraction. An autoencoder network was employed in [184], [221], and [157] using CSI amplitudes, CSI phase, and RSS, while a deep CNN was adopted in [186] to estimate a location from CSI images. Chen, et al. [12] extracted sequential

local features using a sliding window and employed a deep LSTM to learn temporal dependency between the extracted features. ResNet was used with Signal to Noise Ratios (SNRs) measurements to estimate a location as a label or coordinate. All of the above works reported an improvement in accuracy of around 20-30% compared with traditional machine learning methods. To the best of our knowledge, the use of a deep learning approach for PIR-based location estimation is still unexplored.

4.3 Problem Formulation

Heat emanated from a moving person can be captured by a PIR sensor. Then, its analog outputs are generated with respect to the orientations of the PIR analog sensor and the distance between the person and the sensor [24]. Thus, we can arrange each PIR sensor to monitor each part of a monitored area and analyze the raw signals from each sensor to estimate a person's location. Based on this finding, we are motivated to locate a moving person using a sensor node that consists of multiple PIR sensors. The sensor node can be attached to a ceiling above the monitored area to avoid its line of sight being obstructed by furniture.

A monitored area A can be divided into a set of grids $G = \{g_i | i = 1, \dots, n\}$ with a center coordinate at $g_i = (x_i, y_i)$. A grid i is further divided into a set of cells $C = \{c_j | j = 1, \dots, 4\}$ with a center coordinate at $c_j = (x_j, y_j)$ where each cell has a size of 50cmx50cm which fit the average size (standing) person. A PIR node is located at the center of the monitored area, and the node consists of m sensors. When a person walks across the monitored area, it generates a set of reference measurements $S_{ref} = \{s_k | k = 1, \dots, m\}$ where s_k is a raw signal vector from a PIR sensor k. The raw signal vector can be represented as $s_k = [s_{k1}, s_{k2}, s_{k3} \dots s_{kt}]$ where t is a time step. Given a set of current measurements $S_{current}$, we want to estimate the location of a person for each time step.

4.4 Methodology

In this section, we discuss the methodology of our deep learning-based PIR location estimation. First, we introduce a location estimation process in which we collect training data to train a model in an offline phase, and a location of a person is estimated in an online phase. Secondly, we present our proposed CNN-LSTM architecture in the offline phase. Lastly, we discuss the underlying

techniques that are employed to support the online phase.

4.4.1 Proposed PIR location estimation process

Our deep learning-based PIR location estimation process consists of two phases as shown in Figure 4-1: 1) the offline phase and 2) the online phase.

In the offline phase, a person walks in a monitored area according to predefined instructions. Raw output signals are generated and collected from a PIR sensor node. Then, the raw outputs are preprocessed into a suitable training data structure. Next, our CNN-LSTM network is trained using the training dataset.

In the online phase, the PIR node generates and sends the raw output signals, with or without a person walking in the monitored area, to a sink node periodically. When we receive the raw outputs, the movement of a person can be detected by a threshold technique, which was successfully applied in several papers [43, 76]. In general, we can directly specify a threshold value to detect the changes in raw signals in a threshold method. However, the signals can contain both positive and negative pulses, and the amplitudes of these pulses are varied. Thus, it is difficult to set a proper threshold value.

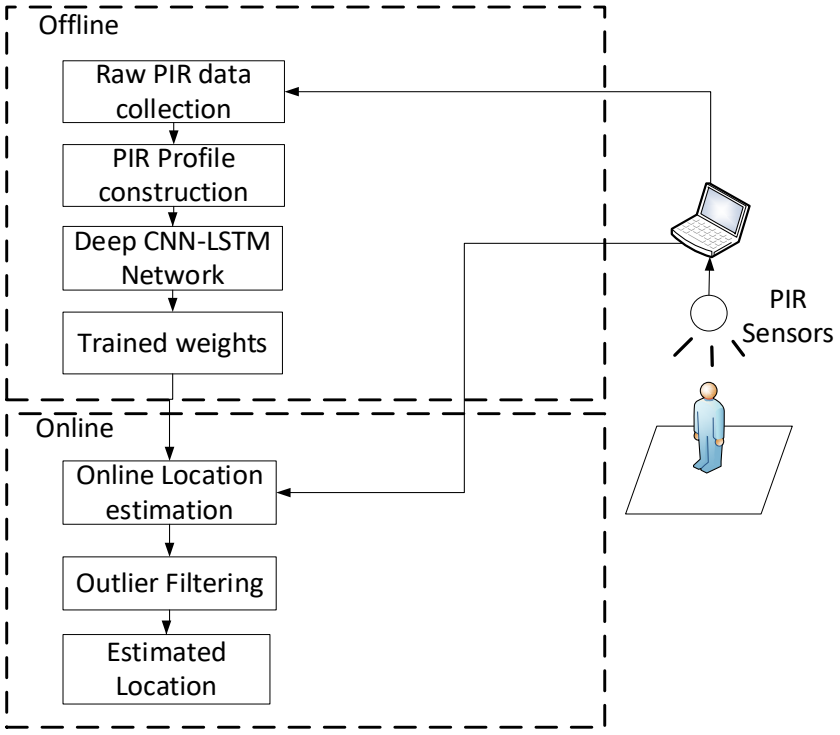


Figure 4-1 An overview of PIR based location estimation.

Instead, we apply Hilbert-Huang Transform (HHT) [222] on the signals to obtain the Hilbert spectrum that captures characteristics of the signals such as the amplitude and the instantaneous frequency as a function of time. As a result, we can extract the instantaneous energy of the signals over time which can be used to indicate a movement of a person. Once a movement is detected, the signals are fed into the trained model to predict estimated locations as 2d coordinates. Then, we filter outliers to improve the accuracy of the location estimation.

4.4.2 Offline phase

In the offline phase, we focus on collecting raw PIR signals for training our deep learning model to estimate the location of a person in a monitored area. First, we describe the process of a PIR profile construction. Second, we discuss the details of our proposed deep learning architecture. Third, a model training process is explained.

4.4.2.1 PIR profile construction

To construct a training dataset, a person walks across a monitor area in different directions, and raw PIR signals are collected. We assume that a person's walking speed is constant to reduce the impact of speed on the PIR signals that is not our main focus to address in this work. For each walking direction, samples of PIR signals are stored in the following format, as shown below, where m represents an index number of each PIR sensor in a sensor node and n represents a time step.

$$V_{Raw} = \begin{Bmatrix} v_{11} v_{21} \cdots v_{m1} \\ v_{12} v_{22} \cdots v_{m2} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ v_{1n} v_{2n} \cdots v_{mn} \end{Bmatrix} \quad (1)$$

Next, the samples are cropped and aligned using a cross-correlation method to ensure that the samples of each direction are aligned as much as possible. A better alignment of the samples should help deep learning models to learn the correct features of the signal generated by a person moving in a certain location.

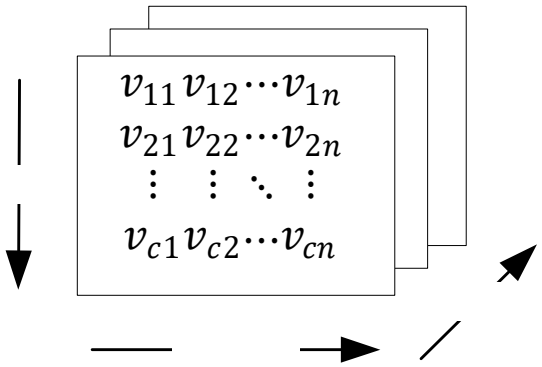


Figure 4-2. A matrix of training data

Once the alignment is completed, we then form a training data matrix M with a size of $c \times n \times b$ where we treat sensors in a sensor node as c channel, n represents a time step of the PIR signals in each channel, and b is the number of samples, as shown in Figure 4-2. In our case, the origin of the raw signals does not start at 0 volts but at 2.05 volts. Unscaled inputs can cause unstable learning during a training process, and our model learns inappropriate weights, which can affect the accuracy of our deep learning model. Therefore, raw PIR output signals are rescaled to the origin of zero. However, we do not need to normalize the raw signal data because there is no difference in the scaling of readings from PIR sensors. In addition, a zero-padding is applied to the PIR profile because different walking directions result in different lengths of the profile signals. We also need to ensure the lengths of the profile data are equal because an inconsistent length of the data can affect the performance of our deep learning model [34]. Finally, we associate the center coordinates of reference locations to each time step to serve as a ground truth.

4.4.2.2 Proposed network architecture

To estimate a person's location using PIR analog signals, we adopt a deep learning-based approach based on a convolutional network (CNN) and Long Short-Term Memory (LSTM) network. Our proposed architecture consists of two levels as shown in Figure 4-3.

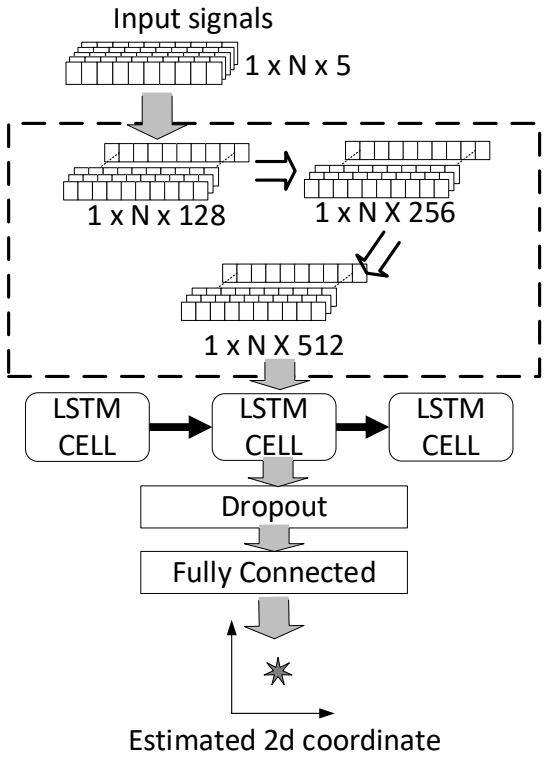


Figure 4-3 Proposed deep CNN-LSTM architecture

The upper part is for feature extraction via a CNN containing multiple convolutional layers, whose inputs are provided by the PIR node. There are 3 layers with the filter numbers 128, 256, and 512. Each convolutional layer gradually extracts dominant features which can be utilized to estimate a person's whereabouts. The lower part is a single LSTM layer that takes features extracted by the upper CNN. The LSTM layer focuses on learning temporal patterns of the extracted features to help differentiate a person's location at different time steps. We also use a dropout layer to reduce overfitting. Finally, a fully connected layer is employed to output the estimated coordinates for each time step. The details of each part are discussed in the following sections.

4.4.2.3 CNN model

A CNN model consists of multiple convolutional layers. The core operation so-called "convolution" is performed in each convolution in the CNN model. The convolution is a linear operation that involves the dot product multiplication of a filter or a kernel that has its size less than the input data and is a set of weights with the input data.

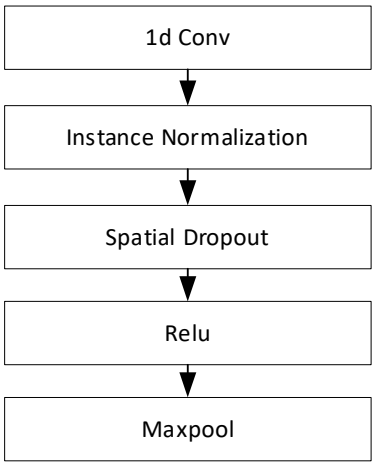


Figure 4-4 Details of each CNN layer

During the convolutional operation, the filter is slid systematically across the input to perform a multiplication at different parts of the input. As a result, this systematic operation allows the filter to discover some features in the input data. In our model, for each convolutional layer, a 1d convolution operation is performed along a time axis to extract features from the sequences of observation, as shown in Figure 4-4. The size of the filter is set to 3, and the stride length is set to 1. We extract the feature maps according to the number of filters. A zero padding is applied to maintain the length of the input sequence. Then, instance normalization [223] is applied to the feature maps of 1d convolution. For each observation, the mean u and variance σ^2 is calculated, and the input is normalized across each channel using the following formula:

$$\hat{X} = \frac{X-u}{\sqrt{\sigma^2+\epsilon}} \qquad (2)$$

where ϵ is added in the denominator for numerical stability and is an arbitrarily small constant. Compared with batch normalization, instance normalization can improve the performance (such as training time and convergence) of a deep learning model for time-series data [223]. After that, a dropout operation is performed on the normalized feature maps. This operation randomly ignores or drops some layer outputs to prevent overfitting and improve generalization error. Similar to [224], we consider that time-series data are highly correlated, and a normal dropout can lead to slow training time but does not prevent overfitting. In our model, we adopt a spatial dropout that extends a dropout operation across the entire feature map. Next, the feature maps are passed to the rectified linear activation (ReLU) function which is the most widely used activation function in a deep learning approach. The function returns the same value if it receives

a positive value, otherwise, it returns zero if the value of an input is negative. It can be formulated as:

$$f(x) = \max(0, x) \qquad (3)$$

There are three benefits for using the ReLU function [225]. First, it improves computational efficiency. Second, it prevents the vanish gradient problems. Lastly, the advantages of sparsity can be exploited.

After that, down-sampling or pooling is performed to reduce the resolution of the feature maps while important features of the maps are maintained through translational and rotational invariants. As a result, pooling operation can improve the computational cost due to the reduced number of parameters and minimize the effect of overfitting. A 1D max pooling with a window size of 2 and a stride length of 1 is performed across each feature map. Max pooling divides the feature map into sub-regions, and the maximum value of each region is calculated. Similar to the 1D convolution, the zero-padding method is also applied for the same reason.

4.4.2.4 LSTM network

Long Short-Term Memory (LSTM) is one type of recurrent neural network that specializes in learning temporal patterns from time series or sequential data. Similar to a traditional recurrent neural network, LSTM processes inputs and retains its memory of the received information. Then, it makes predictions according to information in its memory. LSTM has a mechanism that allows it to determine which information needs to be maintained or deleted. As a result, it is more superior to a traditional RNN in handling long sequences. A basic structure of vanilla LSTM is shown in Figure 4-5.

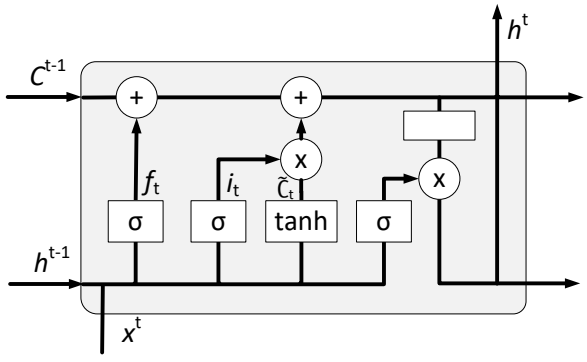


Figure 4-5 Basic LSTM structure

At a time step t, an input is denoted as x_t and h_t is a hidden state. Weights are denoted as w_f, w_i, w_C,and w_o. Biases are represented as b_f, b_i, b_C, and b_o. Once an input is received at time t, the previous time step information in the cell state at a time step $t-1$, denoted as C_{t-1}, is evaluated by a forget gate which can be expressed as:

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right) \quad (4)$$

The purpose of the forget gate is to determine which information should be deleted from a cell state C_{t-1}. It returns either 1 or 0 as its output. If $f_t = 1$, the previous time step information is kept. Otherwise, the information is removed when $f_t = 0$. Next, two prerequisite tasks are executed before the cell state is updated. In the first task, an input gate determines which value needs to be updated and can be formulated as follows:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (5)$$

where $\sigma(\cdot)$ is a sigmoid function. A candidate state is generated using the following equation:

$$\tilde{C}_t = tanh(w_C[h_{t-1}, x_t] + b_C) \quad (6)$$

where tanh $(\cdot)$ represents a tangent function. Then, the current state C_t is updated using the following equation:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (7)$$

a multiplication of f_t and C_{t-1} may result in deleting some information in a cell state C_{t-1}while the multiplication of i_t and $\tilde{C}_t$ determine which information is important to maintain from $\tilde{C}_t$. Therefore, the current state can be updated by the addition of outputs of these two multiplications. To calculate the final hidden unit output h_t, we need both the current state C_t and the output of a sigmoid function o_t, which indicates what information the hidden state should keep, and its equation can be expressed as follows:

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (8)$$

Finally, the current hidden state h_t is computed using the following formula:

$$h_t = o_t tanh(C_t) \quad (9)$$

The current cell state C_t and the current hidden h_t are then carried over to the next time step. By utilizing LSTM in our architecture, extracted PIR signal features can be stored in memory cells systematically while temporal dependencies between the features are also preserved. This memory cell concept is very crucial for PIR-based location estimation because the information

of the past time steps can help improve the location estimation [12]. In our research, the sequences of the features extracted by the upper CNN level are fed into a single LSTM network, as shown in Figure 4-3. We do not stack one LSTM layer on the other LSTM layer because it performs worse than a single layer due to the effect of overfitting [199]. Our LSTM uses a Multiple-Input Multiple-Output structure (MIMO) in which it processes the input sequences sequentially and returns the output sequences of the same length as its input. Finally, the outputs of the LSTM layer are passed to the fully connected layer to estimate the location for each time step.

4.4.2.5 Model training

To train our CNN-LSTM, network parameters need to be optimized by minimizing a loss function. Specifically, the distance error between the estimated location and the actual location needs to be minimized to ensure the performance of a trained model. The Adaptive Moment Estimation (Adam) optimizer [208] is employed for this task. This optimizer can be viewed as a combination of Root Mean Square Propagation (RMSprop) and Stochastic Gradient Descent (SGD) with momentum because the learning rate is scaled using the squared gradients like RMSprop, and it takes advantage of momentum by using the moving average of the gradient instead of the gradient itself like SGD with momentum. The Adam optimizer is widely used in several works due to its excellent convergence and computational efficiency for training a deep neural network [12, 34, 199]. Given that θ_t is the parameter to be optimized and the gradient is denoted as g_t, the optimizer updates the parameter θ_{t+1}as described in the following steps:

$$
\begin{aligned}
m_t &= \beta_1 \cdot m_{t-1} + (1-\beta_1) \cdot g_t \\
v_t &= \beta_2 \cdot v_{t-1} + (1-\beta_2) \cdot g_t^2 \\
\widehat{m}_t &= m_t/(1-\beta_1^t) \\
\hat{v}_t &= v_t/(1-\beta_2^t) \\
\theta_{t+1} &= \theta_t - \alpha \cdot \widehat{m}_t/(\sqrt{\hat{v}_t} + \epsilon)
\end{aligned}
\qquad (10)
$$

where m_t and v_t are estimates of the 1st moment and the 2nd moment of the gradient, $\widehat{m}_t$and $\hat{v}_t$ represents the corrected estimates, the decay factors are denoted as $\beta_1$and β_2 respectively, α is a learning rate, and a small scalar denoted as ϵ is used to prevent division by zero. Mean square error is selected as our loss function to be minimized by the Adam optimizer. It can be expressed

as follows:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \quad (11)$$

where n represents the number of samples, y_i is an actual coordinate, and $\hat{y}_i$ indicates an estimated coordinate.

4.4.3 Online phase

We use the trained CNN-LSTM network to estimate a person's location in the online phase. In this section, we discuss the process of the online phase. First, we explain our technique of movement detection, followed by the technique for location estimation.

4.4.3.1 Movement detection

For PIR-based location estimation, a person's movement affects the temperature of the surrounding environment. PIR sensors can sense this change and generate output signals. By analyzing the signals, we can leverage its features for location estimation. However, PIR signals are sensitive to a small temperature change, thus causing false detection and estimation even though there are no movements. Therefore, movement detection is the first crucial step in PIR-based location estimation. A threshold method is commonly used to determine each signal event to be either active or non-active using a threshold value [43, 76]. PIR signals have both positive and negative phases. Based on our observation, the amplitudes of both phases are not always symmetrical. In addition, multiple PIR sensors may have different sensitivities. As a result, it is difficult to set the best suitable threshold values for each sensor.

To find the most appropriate threshold value, we rely on using instantaneous energy to provide a simple viewpoint to support motion detection. We observe that the instantaneous energy rises and falls according to a movement of a person, and its value is always positive. As a result, the energy is more suitable than voltage to indicate a movement of a person. The Hilbert-Huang transform technique [222], which is used to perform time-frequency analysis of a nonstationary signal, is adopted to measure instantaneous energy. To the best of our knowledge, this is the only suitable technique that can calculate the energy of signals. For each sensor, its signal is decomposed, via an iterative sifting process, into a finite number of intrinsic mode functions (IMF) by the empirical mode decomposition [226] as follows:

$$X(t) = \sum_{i=1}^{N} IMF_i(t) + r_n(t) \quad (12)$$

where $X(t)$ is an original signal, $IMF_i(t)$ is the ith IMF component, and $r_n(t)$ is a residual. Then, the Hilbert transform (HT) that is used to determine the properties of a vibration system such as amplitude, instantaneous phase, and time-varying frequency is applied to each IMF component. In the Hilbert transform, the analytical signal $Z(t)$ is computed as follows:

$$Y(t) = H[X(t)] = \int_{-\infty}^{\infty} \frac{x(u)}{\pi(t-u)} du \quad (13)$$

$$Z(t) = X(t) + iY(t) = A(t)e^{i\theta(t)} \quad (14)$$

where $H[X]$ is the Hilbert operator, the imaginary unit is denoted as i, the amplitude and the instantaneous phase angle are represented as $A(t)$ and $\theta(t)$, respectively. Based on $A(t)$ and $\theta(t)$, the instantaneous energy and the instantaneous frequency can be expressed as:

$$E(t) = |\mathrm{A(t)}|^2 \quad (15)$$

$$\omega(t) = \frac{\mathrm{d\theta(t)}}{dt} \quad (16)$$

For each sensor, the instantaneous energy is calculated by averaging the energy of all IMF components. Then, we calculate the instantaneous energy across all sensors to provide a single energy value that can help define a threshold value by calculating the mean energy:

$$\bar{E} = \frac{\sum_{i=1}^{n} E_i}{n} \quad (17)$$

where n is the total number of sensors, E_i is the instantaneous energy of the i^{th} sensor. Once we have the mean energy across all the sensors, it is possible to determine a single threshold value for all the sensors as follows:

$$P(S, e) = \begin{cases} 1 & \text{if } \bar{E} > \mathrm{T} \\ 0 & otherwise \end{cases} \quad (18)$$

where a threshold value is T, an event is denoted as e, S is a set of sensors, and $\bar{E}$ is the calculated mean energy.

4.4.3.2 Location estimation

Once a movement is detected, the PIR signals are passed to the trained CNN-LSTM model to

estimate the location of a person. Unlike the training phase, the input data is not the entire sequence generated by a person walking across the monitored area, i.e., the input data is a subset of the profile sequence. The length of the input data depends on the frequency of the location estimation, e.g., 0.5 seconds. In the first location estimation, a hidden state and cell state are initialized. Then, the CNN-LSTM model processes the input data to estimate the location as well as update both the hidden state and the cell state. The output of the model is a sequence of the estimated location that has the same length as its input. Each value in the output sequence is represented as a 2D or x-y coordinate to indicate the location of a person at each time step, as shown below:

$$O = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ y_1 & y_2 & y_3 & \cdots & y_n \end{bmatrix} \tag{20}$$

The updated hidden state and cell state of the LSTM layer are returned to be used by the next estimation, and the median value of the output sequence is stored. Then, we determine the next location estimation using the new input data, the updated hidden state, and the updated cell state. However, the hidden and cell states are updated based on the input data of the current time step only, but a relationship between the input data in consecutive time steps is not captured. Thus, the next estimated location may be less accurate if the input data of the next time step is ambiguous or does not have enough information. To address this issue, we update the hidden and cell states using two consecutive time steps, i.e., the current time step and the previous time step. As a result, we can obtain better location estimations compared with when we use only the current time step to update states of the LSTM layer. We found that two consecutive time steps yield the best results amongst longer time steps. When the number of time steps increases beyond three-time steps, the accuracy of location estimation starts to decline due to too much irrelevant information updated to hidden and cell states. When a person turns his body to change his walking direction or stops moving for a short moment, we observe that a person's movement may not be detected. If the person is detected in the previous time step and his recent position is known, we can consider that the person's location remains unchanged and the stored median value from the previous estimation is assigned as the estimated location.

From our initial experimental results, some estimated locations are not feasible for a person to travel from the previous location, and these errors can affect the overall accuracy of location

estimation. To improve the accuracy, infeasible locations need to be filtered out. However, we cannot filter the infeasible locations using a boundary of a walking distance because we do not have explicit knowledge of a person walking speed to determine the possible walking distance. To detect outliers, the method of the mean and standard deviation is commonly used, but the mean and standard deviation are strongly affected by outliers [215]. Thus, they are not a good indicator for outlier detection. Therefore, we apply a statistical method to detect outliers based on a scaled median absolute deviation (MAD) that is insensitive to outliers, and MAD can be calculated as follows:

$$Scaled\ MAD = c * MAD \tag{21}$$

where MAD can be defined as $median(|x_i - median(x)|)$, and c is calculated as $-1/\left(\sqrt{2} \times \left(erfc^{-1}(\frac{3}{2})\right)\right)$. The estimated locations that exceed three-scaled MAD are identified as outliers. Then, we replace identified outliers with the median value.

4.5 Experimental Setup

In this section, we first describe our hardware and environmental setups. Then, we present our data collection process and scenarios used to evaluate our model.

4.5.1 Hardware and environmental setup

We develop a sensor node that consists of five PIR sensors (Model SEN-13968) from SparkFun and each sensor can generate the maximum voltage output of 5 volts. We arrange the five sensors on a sensor node and their (overlapped) FOVs are projected to cover the monitored area, as shown in Figure 4-6. Our designed FOVs create overlapping detection zones that can help in the generation of distinct signal patterns when a person moves in different zones. The FOVs are aligned symmetrically as much as possible to ensure that we have consistent patterns. We use Arduino MEGA 2560 microcontroller to synchronize the transmission cycle of the five sensors and combine the most recent measurements received from these sensors.

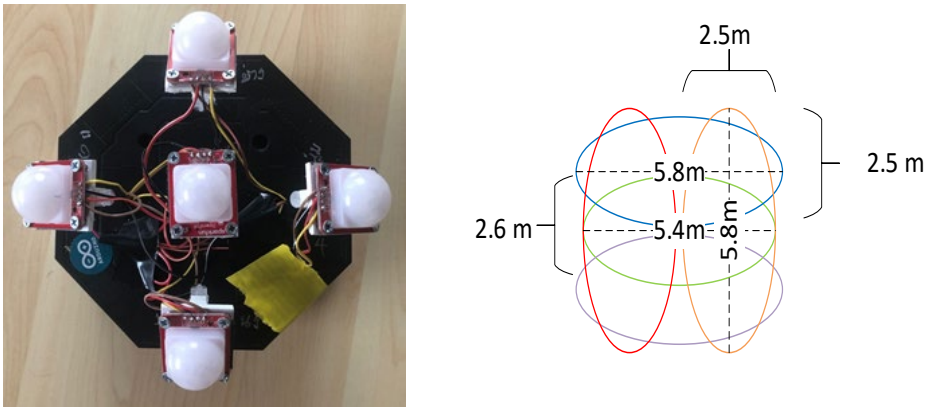


Figure 4-6 PIR sensor node and its FOV

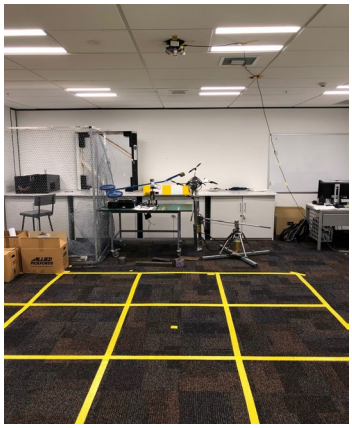


Figure 4-7 The area monitored by a PIR node on the ceiling

The information from the microcontroller is transmitted to a desktop computer, which acts as a sink node that stores all voltages measurements for processing, via serial communication. The sampling rate is set to 60 samples per second because we want to ensure that we do not miss any important features in PIR raw signals, and this sampling rate facilitates a time synchronization for data labeling.

In our setup, we attach our PIR node on the ceiling with a distance of 2.5 meters above the ground level, as shown in Figure 4-7. We empirically evaluate the node's coverage and find that it has an optimal detection area of 9 m^2. The monitored (squared) area consists of 9 grids, of which has a size of 1m×1m. We also divide each grid into 4 smaller cells that have the same area size of 0.5m×0.5m. We record the ground truth using a video camera using 60 frames per second, and the PIR sensors' signals are synchronized with each recorded frame using a local desktop time.

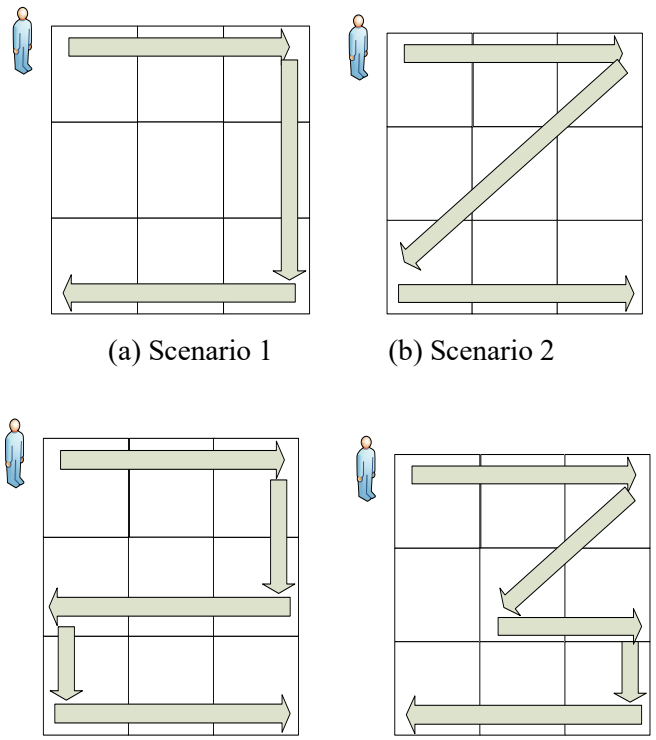
4.5.2 Training and test data collection

We collect a training dataset by instructing a person to walk within the monitored area in different directions. Specifically, the person needs to walk forward and backward horizontally and vertically from 6 starting locations, thus resulting in 24 directions. Besides, the person needs to walk forth and back diagonally from the top left/right corner to the bottom right/left corner of the monitored area. In this case, there are 11 starting locations. Based on the above instruction, we record 44 directions, and we have 1020 sequences of PIR signals as our training data in total. To address the assumption of the constant walking speed, a metronome is employed to generate a steady pulse, which is set according to the person's walking speed preference. As a result, the person can use the pulse as a guideline to control a constant walking speed. Each collected PIR signal sequence is labelled using ground-truth coordinates extracted from the video camera. Our CNN-LSTM model is trained using Matlab 2020b on i5 3.2Ghz Intel quad-core CPU, RAM 16 GB, and a GPU of NVIDIA RTX2070 with 8GB video memory. We vary the model's hyperparameters to achieve the optimal results, and they are shown in Table 4-1.

For performance evaluation, we use mean distance error and the cumulative distance function (CDF) of distance errors as a metric to measure the accuracy of our CNN-LSTM model. We design 4 walking scenarios that range from simple walking patterns to more complex ones, as shown in Figure 4-8. In the complex scenarios, we include a few short walking paths to test how well our model handles sharp turns and random walking directions.

Table 4-1 CNN-LSTM training parameter setting

Parameter	Value
Max epoch	100
Mini batch size	32
Learning rate	0.0005
Learning rate Drop factor	0.1
Gradient threshold	1
Gradient Threshold Method	l2norm
L2 Regularization	0.0001



(a) Scenario 1 (b) Scenario 2

(c) Scenario 3 (d) Scenario 4

Figure 4-8 Illustration of the walking scenarios

In addition to the location estimation accuracy, we are interested in the other factors that can affect the performance including the impact of a sequence length in each observation and the impact of limited training data. We also compare different RNN models in our model including LSTM, bi-LSTM, and GRU to compare the performance of each recurrent network. Besides the deep learning models, we select Support Vector Regression (SVR) in the comparison because it can provide accurate results compared with the other traditional methods and can even outperform deep learning in some cases [96, 182]. In this chapter, a linear kernel for SVR is used because it performs better than the other kernels in our cases. In addition, we compare our proposed method with particle filtering (PF) which is a commonly-used method for location estimation. We implement a system model and observation model based on the azimuth change measurement proposed in [32] for a particle filtering method. In addition, we set the number of particles to 2,000 particles because we have varied the number of particles and found that 2,000 particles are the optimal one. Lastly, a systematic method is employed to resample particles due to its simplicity.

4.6 Results and Discussion

4.6.1 Location estimation accuracy

The accuracy of location estimation can be determined based on the Euclidean distance error between the ground truth's coordinates and estimated coordinates. Our CNN-LSTM model can achieve the overall mean distance error and the standard deviation of 0.2379 meters and 0.1876 meters, respectively, as shown in Table 4-2. In addition, our results show that 80% of the distance errors are less than 0.4 meter, as shown in Figure 4-9, which proves the feasibility of our approach for PIR-based location estimation. From Figure 4-10, we can see that the estimated locations (red X) follow the actual location (blue dot) closely. For each scenario, the results are shown in Column 2 of Table 4-3. Scenario 1 achieves the mean distance error of 0.1542 meter and the standard deviation of 0.0948 meter.

Table 4-2 The mean distance errors and their standard deviations for different model structure

STRUCTURE	MEAN	Standard Deviation	Training Time	Estimation Time
CNN-LSTM	0.2379 m	0.1876 m	6.10 mins	0.043 secs
CNN-BiLSTM	0.3222 m	0.2316 m	10.12 mins	0.063 secs
CNN-GRU	0.5172 m	0.1924 m	5.34 mins	0.037 secs
SVR	0.6906 m	0.4303 m	0.23 mins	0.003 secs
Particle filter	0.4588 m	0.2718 m	N/A	0.320 secs

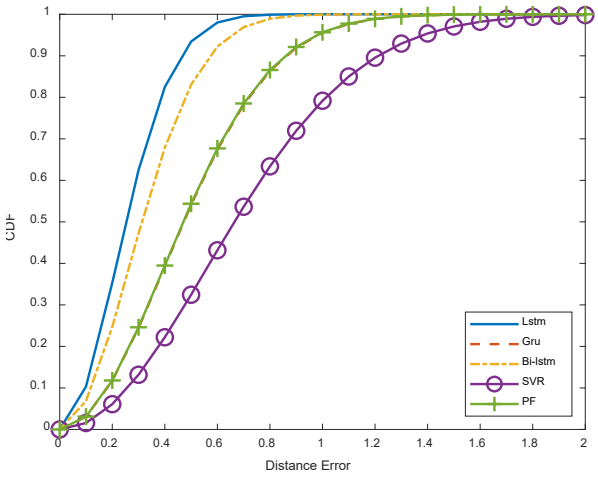
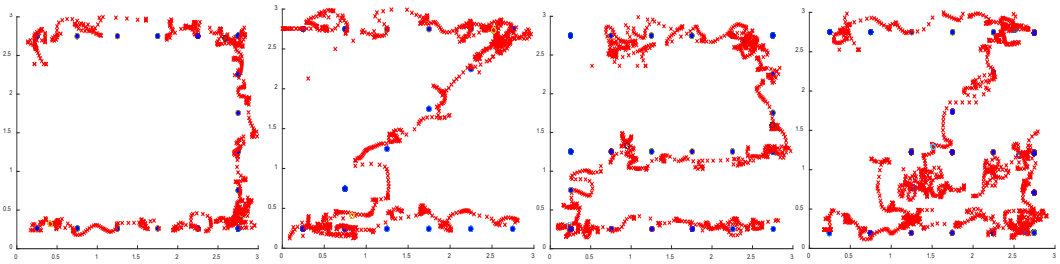


Figure 4-9 CDF of distance error for each model structure (all scenarios)



(a)Scenario 1 (b)Scenario 2 (c)Scenario 3 (d)Scenario 4

Figure 4-10 PIR localization and tracking results. Red X represent estimated location at each time step. Blue dots represent the ground truth location

The CDF plot of Scenario 1 in Figure 4-11 shows that 80% of the distance errors are less than 0.3 meter. In the second scenario, we achieve the mean distance error of 0.3373 meters and the standard deviation of 0.2050 meters. The CDF of the second scenario shows that 70% of the errors are less than 0.4 meters. In the third scenario, the mean distance error of 0.1772 meters and the standard deviation of 0.1450 meters is obtained. 70% of the distance errors are less than 0.3 meters as shown in Figure 4-11. The fourth scenario can achieve the mean distance error of 0.3016 meters and the standard deviation of 0.1968 meters. In addition, 70% of the mean distance errors are less than 0.4 meters. From these four scenarios, our model is proven capable of localizing and tracking a person without the knowledge of the initial location. Although estimation is not accurate at the beginning, the more information is received, the better the model corrects itself and makes a better estimation. For each time step, the received signal data may be ambiguous in some cases, such as when a person starts moving or when a person walks under the PIR sensor node. We observe the amplitudes of raw signals are quite low, and the signal patterns seem to be significantly varied in comparison to the raw signal generated by a person walking perpendicular to the sensor node. Using the data from a single time step, the length of the data is not enough to provide any meaningful insight. As a result, the information stored in an updated hidden state, and an updated cell state is too ambiguous to make an accurate location estimation. This problem affects the accuracy of location estimation in every walking scenario, especially Scenario 2 that we obtain the highest mean distance error of 0.5223 meters as shown in Column 3 of Table 4-3.

Table 4-3 The mean distance errors and their standard deviations for different scenarios

Scenario	2 consecutive time steps		single time step	
	Mean	Standard Deviation	Mean	Standard Deviation
1	0.1542 m	0.0948 m	0.2267m	0.1220m
2	0.3564 m	0.2076 m	0.4492m	0.2768m
3	0.2244 m	0.1855 m	0.3647m	0.2473m
4	0.3138 m	0.1791 m	0.3572m	0.1726m

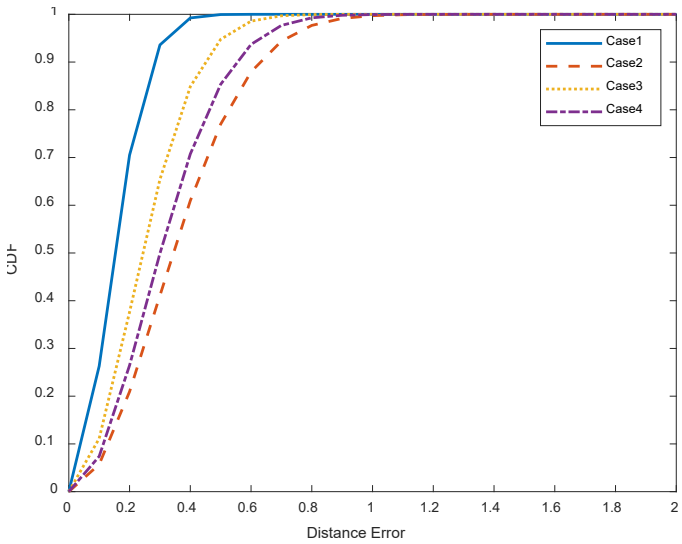


Figure 4-11. CDF of distance error for each scenario

Once we update the states of the LSTM layer using data from two consecutive time steps, we obtain the improved result for every scenario as shown in Table 4-3. This result indicates that the information stored in the hidden and cell states is less ambiguous and can capture a relationship between the signal data for each time step. As a result, the locations of a person can be estimated accurately.

Next, we compare our model with the other well-known recurrent networks including Bi-LSTM, GRU, and PF. Figure 4-9 and Table 4-2 show that our model can achieve the best performance for both mean distance error and CDF. Our CNN-LSTM is about 30% better than CNN-Bi-LSTM. Although Bi-LSTM has an advantage over LSTM by taking into account both past and future information, this is not for our case because each sensor observation is received in real-time so no future information available for BI-LSTM. The performance of CNN-GRU is about 73% lower than ours. GRU aims to improve training speed by simplifying the cell design

to reduce the number of parameters, and the input and forget gates are combined in an update gate. In addition, GRU controls the flow of information without using a cell memory unit. However, it becomes less expressive compared with LSTM, and we find that GRU is not suitable for modeling correlations of PIR output sequences, and its results are inferior to LSTM in our cases. SVR can achieve promising results with the distance error of 0.69 meters in our experiments, but its accuracy is still lower than the deep learning models used in our study. Lastly, particle filtering's accuracy is 63% lower than our CNN-LSTM. We observe that azimuth change measurements can be similar in several positions in the monitored area. Thus, distance errors increase considerably in several time steps due to the ambiguity of the measurement. In addition, the azimuth change requires extensive efforts to extract and enhance to make it usable, whereas deep learning methods can automatically extract and learn discriminative features. Thus, our proposed deep learning method is more practical.

4.6.2 Computational time

Computational time is one important factor for a localization and tracking system because it can affect both the offline and online training phases. In the offline training, a long training time can delay the startup of the system. During the online phase, a long computational time can cause a delay in a location estimation, thus making results inaccurate and less useful. Our proposed model takes roughly 6 mins to train for 50 epochs, as shown in Table 4-2. The training duration is 50% faster than BiLSTM but it is 13% longer than the GRU model. However, our model is still favorable due to its better accuracy. For an estimation time, our model takes the average of only 0.043 sec to estimate a location of a person at each time step, as shown in Table 4-2. The CNN-GRU and CNN-BiLSTM take 0.037 and 0.063 sec, respectively. There is no significant difference in estimation times among these three models. SVR requires 0.23 mins and 0.003 secs to train and to estimate a location, respectively. Compared with the other methods, SVR is the fastest method in our comparison, but the difference is insignificant. The particle filtering obtains the highest execution time because we need to perform calculations for all particles. Even so, the execution time of the particle filter is acceptable for real-time location estimation.

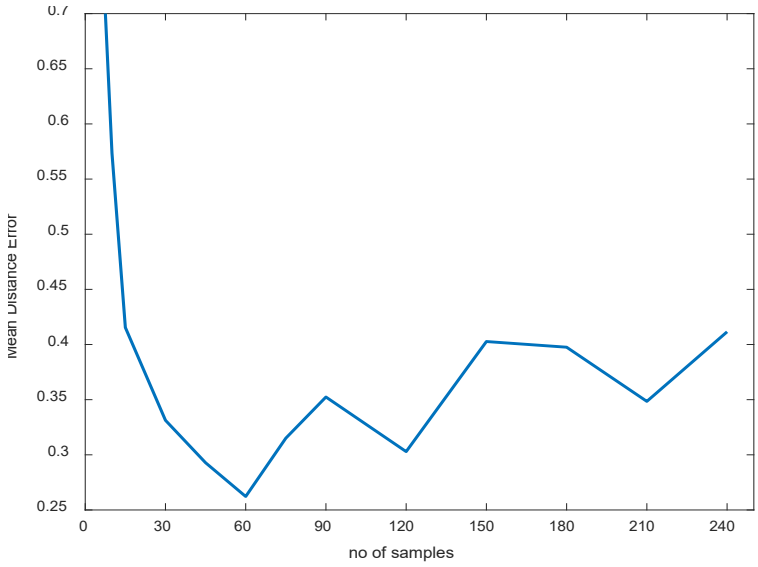


Figure 4-12 Impact of number of samples in each observation on accuracy (all scenarios)

4.6.3 Impact of the number of samples in each observation

One main issue that keeps the deep learning methods away from deployment in real life is the limitation of training data. Constructing a large dataset requires considerable manpower, which is time-consuming and is not practical in real-life. Therefore, we evaluate our approach with different sizes of training data, as shown in Figure 4-12, to see how it affects the estimation accuracy. We found that the accuracy continues to improve with the increase in the number of samples in each observation. Scaling from 10 samples to 60 samples, the mean distance errors are decreased by 75%. However, the accuracy of our model is gradually increased when a sequence length increases above 60 samples. As our model utilizes two consecutive time steps to update the states of LSTM layers, when the length of data is too long, some irrelevant data are included in the hidden and cell states. As a result, these irrelevant data can affect the accuracy of the model. Based on these results, we can say that the selection of 60 samples in a sequence is optimal.

4.6.4 Impact of the amount of training data

One main issue that keeps the deep learning methods away from deployment in the real life is the limitation of training data. Constructing a large dataset requires considerable manpower, is time-consuming, and is not practical in the real-life. Therefore, we evaluate our approach with different sizes of training data, as shown in Figure 4-13.

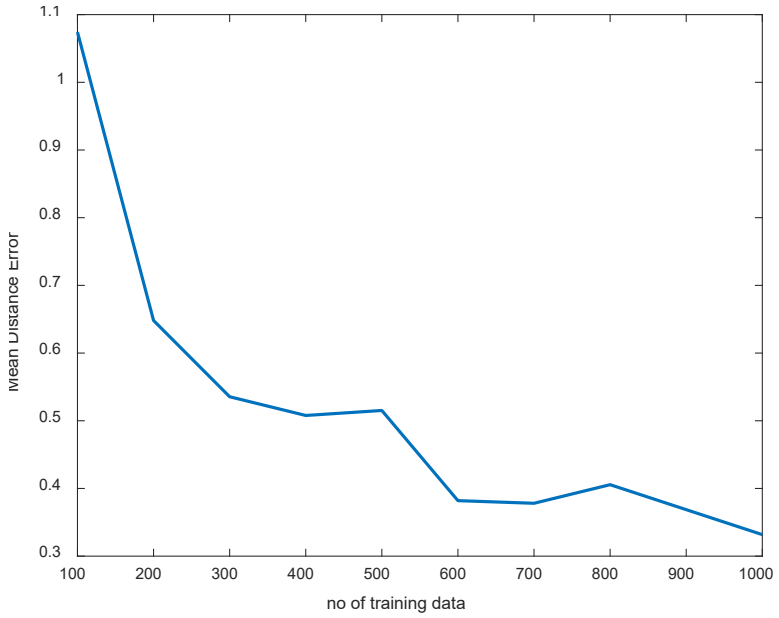


Figure 4-13 Impact of size of training data on accuracy (all scenarios)

The mean distance error of our model improves significantly with a training data size of 200. In addition, the distance error decreases steadily with the increase in the number of training data without the drastic rise and drop in the distance error. Therefore, we can use small and limited training data to initialize our model and gradually add new data to retrain the model. This result can prove that our model is feasible in a real-life environment.

4.6.5 Model Hyperparameters

To discover the optimal hyperparameters of the proposed CNN-LSTM model, we conduct several experiments on the hyperparameters extensively. We test notable hyperparameters that affect the accuracy of the location estimation. The results of the hyperparameter experiments are illustrated in Figure 4-14. For the number of CNN layers, we vary the number of CNN layers from 1 to 6 layers. An increment of the number of CNN layers can improve the accuracy of the location estimation using the validation data. However, the results of the test scenarios show otherwise. More layers can reduce the accuracy considerably due to overfitting. The optimal number of CNN layers is 3 in our case.

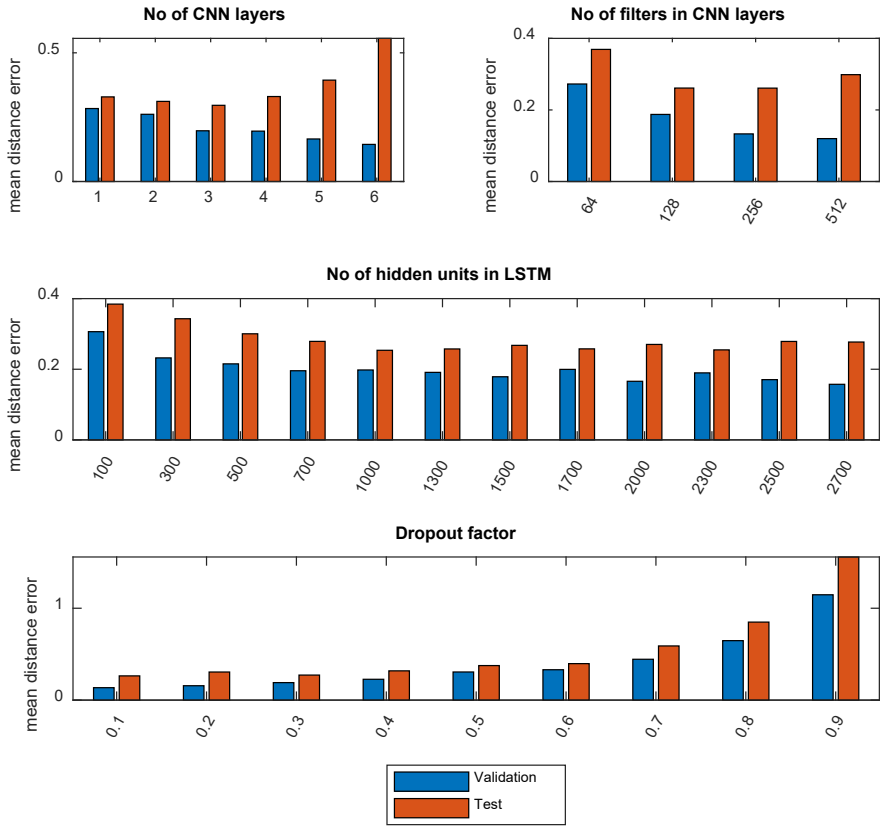


Figure 4-14 Hyperparameter testing of CNN-LSTM model

The number of filters can affect the accuracy of the proposed model. We vary a set of filters including 64-126-256, 128-256-512, 256-512-1024, and 512-1024-2048. A higher number of filters can improve the accuracy of the proposed CNN-LSTM model using the validation data, but the model can perform poorly in the test scenario because of the overfitting issue. The optimal number of filters is 128-256-512 in this case. The results of LSTM hidden units show the same trend for both validation and test. The higher the number of hidden units, the better the accuracy. We use 1024 units because the accuracy of the location estimation does not change significantly beyond this point. Lastly, we vary a dropout factor from 0.1 to 0.9. The results show that the higher dropout value can reduce the accuracy of the location estimation because the model becomes too generalized. We set the drop factor to 0.3 because it gives the best results in our experiments.

4.7 Conclusion & Future works

In this chapter, we propose a deep CNN-LSTM architecture for indoor location estimation using PIR analog signals. Based on an assumption of a constant walking speed, our proposed model not only extract discriminate features from PIR signals without the need for handcrafted features but also capture the temporal pattern of the extracted features to accurately and timely estimate the location of a person. A set of experiments mimicking real-world walking scenarios are conducted to evaluate our model, and the results show that we can achieve a mean distance error of 0.23 meters, and 80% of distance errors are within 0.4 meters.

The presence of multiple persons in the monitored area can generate ambiguous signal patterns. As a result, it affects the accuracy of our proposed method. To handle multiple persons, some challenges need to be addressed, such as signal separation and data association. In this chapter, we scope our research problems to only a single person and will extend our approach to support multiple persons in the next chapter.

Chapter 5 Multi-person Localization using Analog Signals of Passive Infrared Sensors

The Internet-of-Things (IoT) technology has been improved over the past decade. As a result, device-free localization has emerged and become a crucial part of application areas such as healthcare, energy management, etc. Passive Infrared (PIR) sensors detecting changes in temperature in an environment are suitable options for human localization due to a lower cost, low energy consumption, and electromagnetic tolerance. The literature shows that PIR analog outputs can be analyzed to detect multiple persons because it can provide more information such as phase and amplitude Deep learning approaches have been proved in many applications such as activity recognition, surveillance system, localization for their outstanding results. In this chapter, we propose a novel deep learning method for multi-person localization using channel separation and template matching techniques Our method is based on a deep CNN-LSTM architecture. Instead of using a single deep learning model, we integrate the results of multiple models using the mean bagging technique to achieve more accurate localization results. Our results show that the proposed method can estimate the locations of two participants simultaneously with the mean distance error of 0.53 m and 70% of distance errors are within 0.8 m

Chapter 5 has been submitted to IEEE Internet of Things Journal.

.

5.1 Introduction

Indoor localization is a key component of many useful application areas such as healthcare, energy management, and security management [3, 6]. Localization information enables us to respond to any crucial event such as medical or security events appropriately and efficiently. In general, we can categorize indoor localization into device-based and device-free localization. For device-based localization, people are required to wear wearable devices that can provide location information such as mobile phones, smartwatches, and tracking bracelets. However, this requirement is not suitable when people do not cooperate due to intrusiveness or where everyone cannot wear a device [8]. These issues give rise to the device-free localization approaches where persons can be located or localized without requiring them to wear or hold any device. In addition, Internet-of-Things (IoT) technology has been vastly improved over the past decade. As a result, the improvement further increases the capability of device-free localization such as accuracy, coverage distance, sensitivity, etc. Various technologies can be employed to implement device-free localization. Images and Radiofrequency (RF) are commonly used by many works. However, there are a few drawbacks including privacy issues, energy consumption, and hardware cost [18, 227].

Passive Infrared (PIR) sensors have a low unit price and are easy to acquire off the shelf. They can operate at any lighting condition, consume less energy, and are not affected by electromagnetic interference [1, 35]. On the other hand, PIR sensors cannot detect through walls and their detection sensitivity is affected by a rise and fall of background temperature. Despite these drawbacks, PIR sensors have been successfully employed as a low-cost indoor localization solution. We observe that existing works require either a dense deployment of PIR sensors [23, 36] or sophisticated PIR nodes [18, 20] which are not only difficult to implement in real-life but also do not achieve good results in multi-person scenarios. In this chapter, we aim to leverage analog output's features to improve localization for multiple persons. Compared with binary outputs, PIR's analog outputs are feature-rich. By utilizing analog outputs, we can reduce the number of deployed PIR sensors [24].

Deep learning approaches have emerged over the past years and have been adopted successfully in many areas such as image recognition, activity recognition, localization [27, 199,

207]. There are several advantages of using deep learning over traditional machine learning such as automatic feature extraction and efficient at learning complex features [157]. In this chapter, we propose PIR-based localization for multi-person using a deep learning approach. However, some challenges need to be addressed to realize multi-person localization with a deep learning approach.

First, data collection for multiple persons can be cumbersome and time-consuming. In addition, we cannot possibly collect data for a deep learning model to cope with every walking scenario of multi-person. In this work, we utilize an existing profile signal to generate a set of segmented profile signals and combine the multiple profile signals through a summation to generate a signal profile of multiple persons without the need for real data collection.

Second, a PIR sensor may detect an individual or multiple persons together. In a case of a sensor node consisting of multiple PIR sensors, raw PIR outputs have multiple channels. Each channel may contain information from a single person or multiple persons. As a result, we cannot directly use these multi-channel outputs. To address this challenge, we use a template technique to identify whether data in each channel is generated by a single person or multiple persons and then generate inputs for each person accordingly.

Third, the presence of multiple persons can introduce ambiguity and interference into raw PIR outputs. Thus, unexpected patterns can occur and a trained deep learning model cannot recognize them To address this challenge, we combine the localization results of multiple trained models using the mean bagging method to improve the results. Our main contributions are summarized as follows:

- We present a robust localization approach for multi-person without requiring extra data collection.
- We propose a channel separation method to generate inputs for multiple persons through a summation of an existing dataset.
- We leverage multiple deep learning models to handle the ambiguity of signals caused by the presence of multiple persons and our model outperforms the existing methods.

The rest of this chapter is organized as follows. Section 5.2 discusses related works. Section 5.3 presents our proposed methodology for multi-person localization. Sections 5.4 describes the setup of our experiment for both hardware and data collection. Section 5.5 discusses the results of our experiments. Finally, the summary of our work is presented in Section 5.6.

5.2 Related works

We can divide types of localization technology into two categories including non infrared-based technology and infrared-based technology.

For non infrared-based technology, Radiofrequency (RF) technology including Zigbee, WIFI, RFID, UWB, and mmWave has been employed in many works due to its coverage distance and obstacle penetrating capability [49]. However, RF localization is susceptible to electromagnetic interference, a multipath effect that can affect the accuracy of localization [9]. In addition, hardware cost and power consumption are considerably expensive in some RF technology such as WiFi and mmWave [8]. The commonly used localization technique for RF technology includes proximity, triangulation, imaging, and fingerprinting method. The proximity method can localize multiple persons that are located at different sensors but may require a dense deployment to improve the accuracy of localization [44, 130]. The triangulation method utilizes time of arrival (TOA) and angle of arrival (AOA) to estimate locations of multiple persons but reflected signals interfere with each other due to the presence of multiple persons [10, 100, 180]. Thus, we need to separate the signals so separated signals belong to their respective sources. Fingerprinting technique can localize multiple persons that stay in different zones [155]. In addition, localizing different people in the same zone can be achieved by analyzing an impact of a single person and subtracting it from the received measurement recursively [158]. The imaging method can locate multiple persons that cause signal attenuation by obstructing radio links at different locations [107, 144]. Capacitive floor technology can determine the locations of different persons who step directly on the capacitive mats at different locations. it can give an accurate result but may not be practical due to the difficulty of floor mat installation[13, 73]. Acoustic or vibration can localize a person using triangulation techniques similar to RF technology and requires us to separate sources to localize multiple targets [1, 57].

For Passive infrared technology, both passive infrared (PIR) sensors and thermopile sensors are commonly used for localization. In general, changes in temperature on the sensing element of both sensors are converted to either analog output or binary output through multiple steps of the thermal-to-voltage conversion. In general, Passive infrared sensors can be acquired at a low price and have energy consumption but they can achieve the localization error that is less than 1

meter when we use them for localization. There are some drawbacks such as low coverage distance and their sensitivity can be affected by background temperature. However, they are still quite suitable for indoor environments such as residential homes and office buildings. We classify PIR-based localization approaches according to a sensor deployment including *floor-mounted*, *wall-mounted*, and *ceiling-mounted*.

For the floor-mounted category, Qi, et al. [18] developed a PIR sensor module that has a modulated FOV for multi-person tracking. The design of a modulated FOV enables the detection of multiple moving persons and facilitates data association. Kalman, HMM, and Gaussian particle filters are employed to track two targets in a 9m x 9m room. They achieve a distance error of 0.5 meters for two targets. Yang and Zhang [35] arranged six sensors to form 12 overlapped detection zones around a sensor node. For each detection area, a coding scheme is assigned to indicate a detection line. They proposed the credit-based method to filter out false measurement points, i.e. intersection points of activated detection lines. They set up a 10m x10m area and deployed 9 sensor nodes uniformly. Kalman filter is employed to track a target and they obtained an average distance error of 0.42 meters for 2 persons.

For the wall-mounted category, Kemper and Hauschildt [192] attached 4 thermopile sensors on walls of a 4.9m x 6.2m area. Probability Hypothesis Density (PHD) filter was adopted to perform location estimation for multiple targets, They achieved the mean distance error of 0.3 meters for 2-3 persons. Liu, et al. [32] proposed azimuth change measurement that is the difference of the azimuth of a person with respect to a PIR sensor at two locations. They adopted particle filtering to localize multiple persons. In their experiment, 4 PIR sensors were installed on a 8m x 8m area. The localization result for two persons is around 0.8-0.9 meters for a distance error. In addition, they adopted CNN-BiLSTM to separate signals for each person and BiLSTM to estimate a location for each person in [29]. They obtained a mean distance error of 0.73 meters for two persons. LOCI sensor modules were developed by Narayana, et al. [194]. The module contains one thermopile and one PIR sensor. Thermopiles are responsible for providing location information across the FoV cone axis while a PIR sensor focuses on estimating the location between a person and itself. K nearest neighbour (KNN) is trained to estimate a possible location of a person. In their experiment, only one sensor is mounted on a wall at 1.2 meters height to monitor a 9m x 8m area and their result shows that 82% of the distance error is less

than 0.88 meters for 2-3 persons

For the ceiling-mounted category, Tao, et al. [36] deployed 43 ceiling-based PIR sensors to localize multiple persons in an office room with a size of 15.0m × 8.5m. They integrate knowledge such as desk s' locations and the moving directions to improve location estimation. The result shows that around 75% of the distance error is within 0.5 m. Wu, et al. [183] developed the rotationally shuttered PIR (Ro-PIR) sensor. The rotating shutter allows the developed sensor node to detect a stationary person. In their experiments, the PIR sensor was attached to a 2.8 m height ceiling to monitor a circle area with a radius of 2 meters. Machine learning algorithms such as KNN and SVM were employed to localize individuals and achieved the distanace error within 0.44 meters.

5.3 Methodology

In this section, we present the methodology of our deep learning-based PIR localization for multiple persons. First, we introduce an overview of our localization scheme for multiple persons. Secondly, we discuss an underlying technique for data separation. Lastly, we discuss our deep CNN-LSTM model for localization of multi persons and a method that we use to improve the accuracy of localization.

5.3.1 Overview of localization method for multiple persons

To localize locations of multiple persons, our method consists of two main steps: channel separation and person localization, as shown in Figure 5-1. The two steps method is inspired by PIRNet [28]. A monitored area A is consisted of g grids. Each grid consists of c cells. Once persons' movements are detected by a threshold method, raw PIR output X_k is generated at time step k. The first step aims to separate a raw PIR output into m single person inputs according to m persons. According to [29], PIR outputs of a single person can be summed to approximate the outputs of multiple persons. We have existing profile sequences U collected from a single person, we segment existing PIR profile sequences U by cells into smaller profile sequences u_i where i is a cell index. Then, we can combine the smaller sequences to generate N combinations.

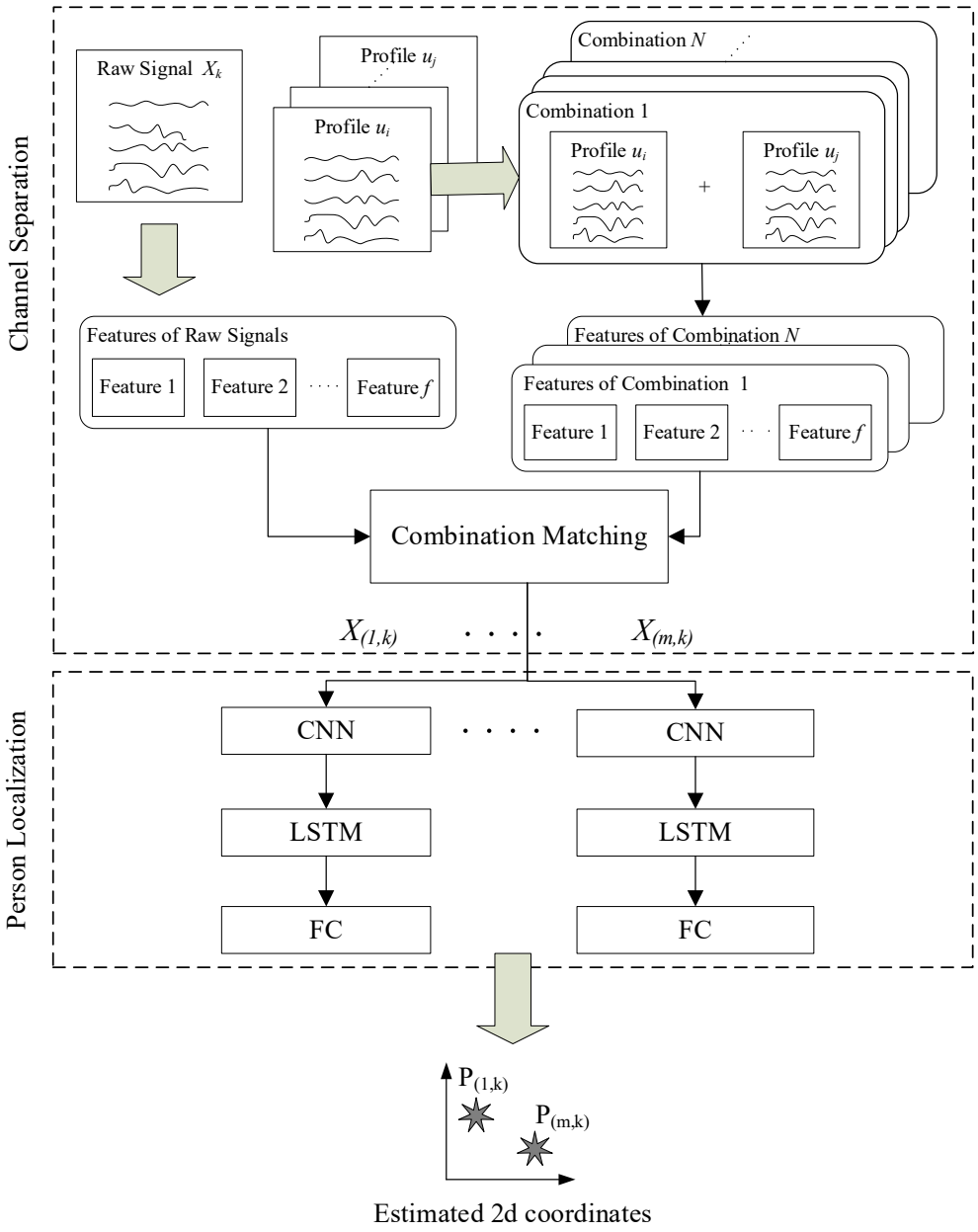


Figure 5-1 An overview of PIR based localization for multiple persons

The raw output X_k is compared to these combinations to find the closet match one. As a result, the matched combination can help us dividing the input X_k into a set of single person inputs [$X_{1,k}$ ... $X_{m,k}$] for our next step.

In the second step, the extracted inputs are fed to a trained deep learning model to estimate the locations of multiple persons as shown in Figure 5-1. The upper CNN level focuses on extracting features. Then, an LSTM level receives the extracted features from the upper layer and learns time dependency between extracted features. Lastly, the fully connected layer is used to output the estimated coordinates for m persons. In addition, we also train multiple models and combine their results to improve the accuracy of localization for multiple persons. The details

for each step will be further discussed in the following section.

5.3.2 Channel Separation

Once persons' movements are detected, a PIR node generates and send raw analog signals that can be represented as a 2d matrix with a dimension of $C\ x\ T$ where C stands for a channel and T is the number of samples. In our work, each output has 5 channels and contains 60 timesteps. The raw signal X_k is used as an input for the location estimation. The signal may contain information generated by multiple persons. For each channel, a signal needs to be analyzed to determine whether it is generated for whom. Thus, our goal of this step is to divide the input into a set of inputs for m persons.

When a person walks across different cells in a monitored area, we observe that distinct patterns of raw PIR signals are generated from different cells. Signal patterns can also indicate which sensors in the node are activated for each cell. Sensor activation is the information that we can utilize to separate the input data for each person. Motivated by [29], we sum PIR outputs of a single person to approximate the outputs of multiple persons. Thus, we should be able to combine profile signals from different cells to estimate the input signals. Then, we can find a combination that is closely matched. As a result, the profile signals that contribute to the combination can be used to indicate a set of sensors activated by each person. The process of this step is illustrated in Figure 5-2. Next, we segment PIR profile sequences into smaller profile sequences according to cells in the monitored area. In total, we have the number of segmented profile signals according to the total number of cells. Each segmented profile is associated with a particular set of activated sensors. We assume that the number of persons and their starting locations is available for us. For each time step, we select sets of segmented profiles from their nearby cells only because computational efficiency can be considerably affected if we select too many profiles to generate the combinations. In general, a standard walking speed is around 1.2-1.3 meters per second and the size of a cell is 0.5 meters x 0.5 meters. At one time step, a person can walk across two cells at a maximum. Thus, two surrounding adjacent cells around a person's location are chosen.

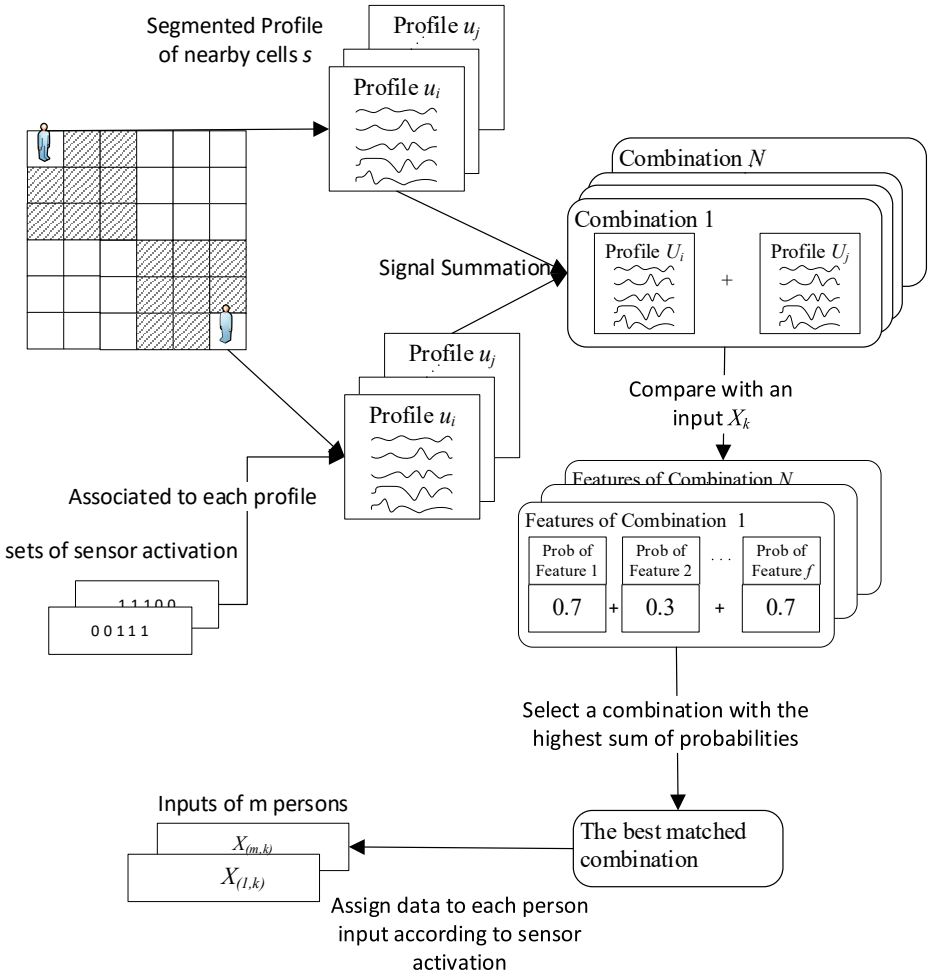


Figure 5-2 An overview of PIR based localization for multiple persons

Next, we generate signal combinations using the chosen sets of segmented profiles and the total number of combinations N is calculated as follows.

$$N = s_1 \times s_2 \ldots \times s_m \quad (1)$$

where s_i denote the number of chosen segmented profiles for i^{th} person and m is the number of persons. For each combination, the selected segmented profiles are combined through a summation. Then, we compare the input X_k with all combinations to find the closely matched combination. To compare the similarity between the input signal and profile combinations, we employ cross-correlation and dynamic time warping. In addition, we use statistical features including peak-to-peak amplitude, root mean square, variance, standard deviation, mean, mean frequency in this process as well. For each feature, we calculate the difference between the input

and profile combinations. Then, we convert cross-correlation, dynamic time warping and the difference in measurements into probability values as follows.

$$pb_{i,n} = \frac{v_{i,n}}{\sum_{n=1}^{N} v_{i,n}} \quad (2)$$

where $pb_{i,n}$ and $v_{i,n}$ denote a probability value and feature value of i^{th}feature of n^{th} combination. For each combination, we sum the probability values up and a profile combination that has the highest probability value is selected as shown below.

$$result_{comb} = \arg max(\ sum(n)\ ) \quad (3)$$

where $sum(n)$ is a sum function for probability values of n^{th} combination

Once the possible combination is selected, the associated sets of sensor activation are utilized to generate inputs for multiple persons. We create m single person inputs with a dimension of $C\ x\ T$ according to the number of persons. For each single person input, we assign a signal in each channel of the input X_k to a corresponding channel of a single person input based on a set of activated sensors.

5.3.3 Person localization with a deep learning model

In this step, we use the channel separated data to estimate the locations of *m* persons. We propose a deep learning network comprised of CNN level and LSTM level as shown in Figure 5-1. The upper CNN level consists of 3 CNN layers with the number of filters 128, 256, and 512 respectively as shown in Figure 5-3. For each CNN layer, we employ a 1d convolutional neural network that has good performance on time series data. The feature extraction is performed in 1d CNN by sliding a filter that has its size smaller than an input systematically across 1d sequence data to perform a dot multiplication at different parts of the sequence to discover some important features. In our work, we set a size of a filter to 3 and set a stride length to 1. In addition, a zero-padding is applied to maintain the length of the input sequence. The outputs or feature maps of 1d CNN are normalized using an instance normalization that improves training time and convergence of a deep learning model[223].

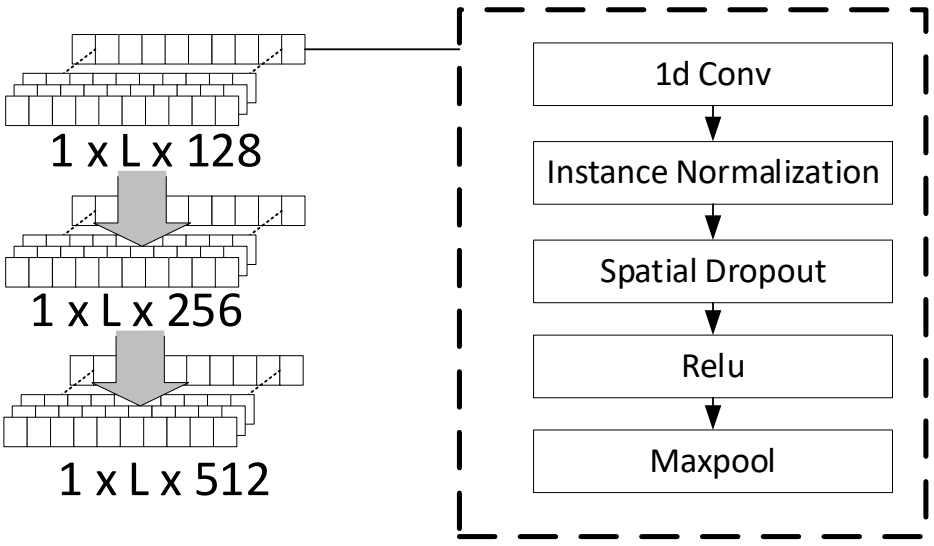


Figure 5-3 The CNN layers and the details of each layer

To avoid overfitting and improve generalization, we apply a spatial dropout on the normalized outputs. We choose this type of dropout operation because it has been reported in [224] that a spatial dropout can improve overfitting and training time. Next, the feature maps are sent to the activation function. rectified liner action (RELU) function is chosen in our work because it is computationally efficient and can mitigate vanish gradient problems[225]. Finally, 1d max pooling is adopted to reduce the number of parameter and improve computational efficiency while discriminative features are maintained. In our work, a pooling window size is set to 2 and a stride length is set to 1. The zero padding is also applied for the same reason as in 1d CNN. The lower level focuses on learning temporal patterns from time series or sequential data. LSTM is employed for this task. Compared with traditional Recurrent Neural Network (RNN), LSTM can handle longer sequences because it utilizes the concept of input gate, forget gate, and output gate to determine which information to be maintained or removed. For this reason, it is suitable to apply to a sequence of PIR output that is considerably long and complex. The LSTM unit is shown in Figure 5-4 and its operation can be expressed as follows.

$$f_t = \sigma\big(w_f[h_{t-1}, x_t] + b_f\big) \tag{4}$$

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{5}$$

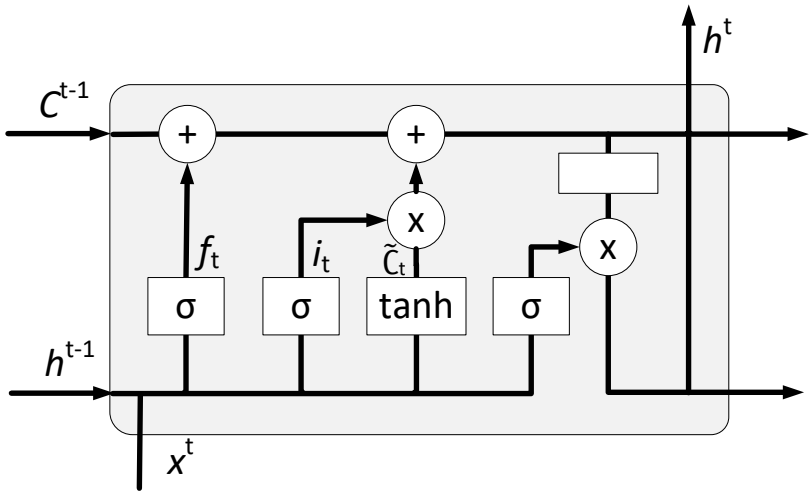


Figure 5-4 LSTM unit and its details

$$\tilde{C}_t = tanh(w_C[h_{t-1}, x_t] + b_C) \quad (6)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (7)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t tanh(C_t) \quad (9)$$

f_t , i_t and o_t stand for a forget gate, input gate, and output gate respectively. an input is expressed as x_t and h_t is a hidden state. Weights are denoted as w_f, w_i, w_C, and w_o. Biases are represented as b_f, b_i, b_C, and b_o. A sigmoid function and a tangent function is represented as $\sigma(.)$ and $tanh(C_t)$ accordingly. Lastly, C_t is a cell state. After the LSTM layer, we add a dropout layer to improve the generalization of our deep learning model. Finally, The fully connected layer is used to estimate locations of *m* persons for each time step.

To train our deep CNN-LSTM model, we employ the Adaptive Moment Estimation (Adam) [208] to minimize a loss function to find the most optimal parameters of the deep learning model because it has an excellent convergence and a good computational efficiency. A loss function, namely Mean square error (MSE) is chosen for the Adam optimizer. It can be calculated as follow:

$$MSE = \frac{\sum_{i=1}^{j}(y_i - \hat{y}_i)^2}{j} \quad (10)$$

where y_i and $\hat{y}_i$ are a ground-truth coordinate and predicted coordinate respectively. The number of samples is represented as j.

We observe that multiple-person scenarios can affect the location estimation considerably due to the ambiguity and diversity of signal patterns. Thus, it is difficult for a single model to give an accurate result. To further improve the location estimation, we adopt 10-fold cross-validation to reduce the bias of the location estimation and avoid an overfitting issue. We divide the training dataset into 10 slices. For every training process, 9 slices are used to train a deep learning model and 1 slice is used for the model evaluation. As a result, we receive 10 different models. Then, we integrate these models using mean bagging which is calculated as follows.

$$O_{mean\ bagging} = \frac{O_1 + O_2 + \ldots + O_n}{n} \quad (11)$$

where O_i is the output of i^{th} model and n is the total number of models. The technique can help reduce the variance of the localization estimation. By combining these two techniques, we can achieve better localization results as reported in [34].

5.4 Experiment Setup

In this section, we first describe our hardware and environment setup. Then, we discuss our data collection process and scenarios used to evaluate our model.

5.4.1 Hardware and environment setup

To collect training and testing data, a PIR sensor node is developed using five Sparkfun PIR sensors (Model SEN-13968) and Arduino Mega 2560 microcontroller. Each PIR sensor is equipped with a fresnel lens that has 100-degree horizontal and 60 degrees vertical FOV. When a movement is detected, it can generate analog outputs at a maximum of 5 volts. We arrange the FOV of each sensor as illustrated in Figure 5-5.

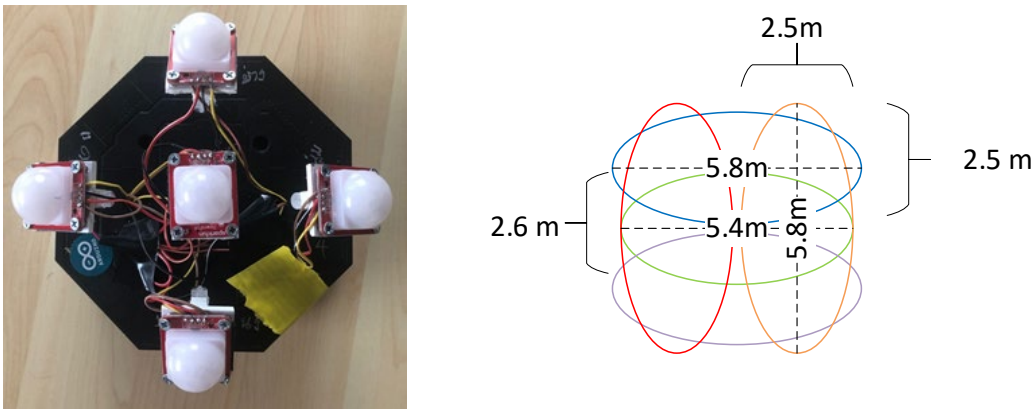


Figure 5-5 PIR sensor node and its FOV

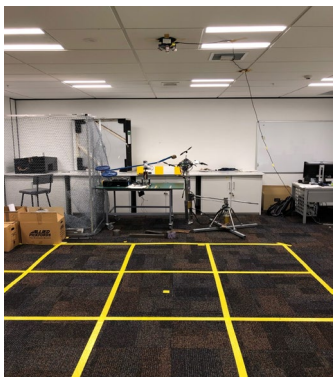
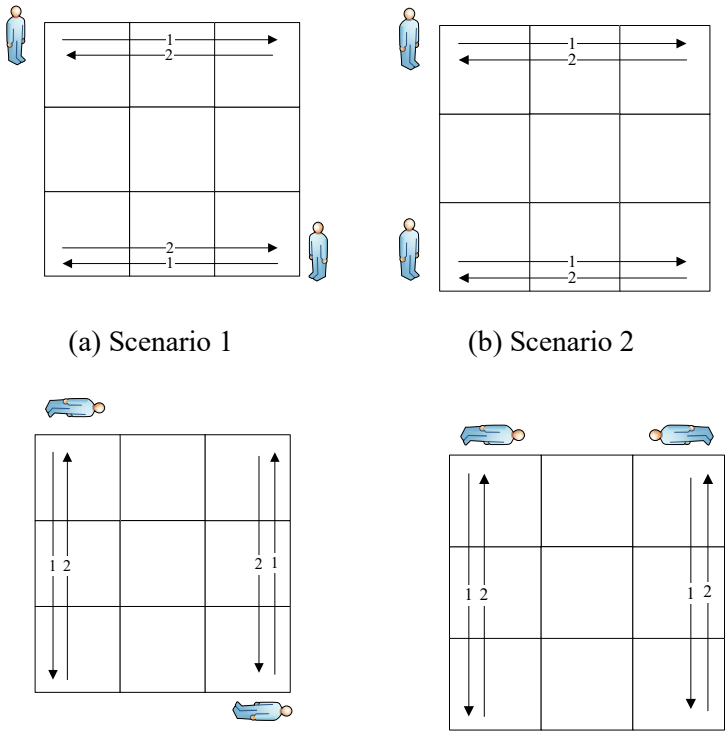


Figure 5-6 A photo of the area monitored by a PIR node on the ceiling

The middle sensor points directly to the ground while the surrounding sensor is angled at 30 degrees from the vertical. As a result, overlapped detection zones are formed symmetrically, which can improve the differentiation of a person's location. The developed node is connected to a desktop via serial communication and transmits collected data at a rate of 60 samples per second to ensure that we can capture and simplify a time synchronization for data labelling.

Then, we select a temperature-controlled environment such as an office room to set up a monitored area because we want to avoid frequent changes in an environment that can affect the quality of our data collection. We prepare a square monitor area with a size of 3 meters x 3 meters as shown in Figure 5-6. We divide the monitored area into 9 grids, i.e. a grid has a size of 1 $m^{2,}$ and we label each grid using a number from 1 to 9. Furthermore, each grid is further separated into 4 smaller cells that have a size of 0.25 m^2. Each cell is labelled with a number from 1 to 4. The sensor node is attached to 2.5 meters height ceiling at the center of the monitored area.



(a) Scenario 1 (b) Scenario 2

(c) Scenario 3 (d) Scenario 4

Figure 5-7 Walking scenarios for multi-person localization

5.4.2 Training and Test data collection

Our training data is collected using a single participant in this work. We instruct a participant to walk horizontally, vertically, and diagonally across the monitored area. In this work, we assume that a person walks at a constant speed because tackling the impact of walking speed is not our scope in this work. To maintain the speed, we use a metronome to give a signal to a participant for each walking step. There are 6 starting locations for both horizontal and vertical paths, so there are a total of 12 sequences. For a diagonal path, there are 22 sequences to be collected as we can walk diagonally from 11 starting locations for both left and right sides of the area. In addition, a participant needs to walk back and forth for each starting location. Therefore, we collect 68 sequences to complete one set of training data. We repeat 15 times and we have a total of 1020 sequences of PIR signals as training data.

To evaluate our localization method, we design different walking scenarios, as shown in Figure 5-7. Two participants participated in our testing data collection. For each scenario, both persons are instructed to walk back and forth across the monitored area at a constant speed.

Table 5-1 CNN-LSTM training parameter setting

Parameter	Value
Max epoch	100
Mini batch size	32
Learning rate	0.0005
Learning rate Drop factor	0.1
Gradient threshold	1
Gradient Threshold Method	l2norm
L2 Regularization	0.0001

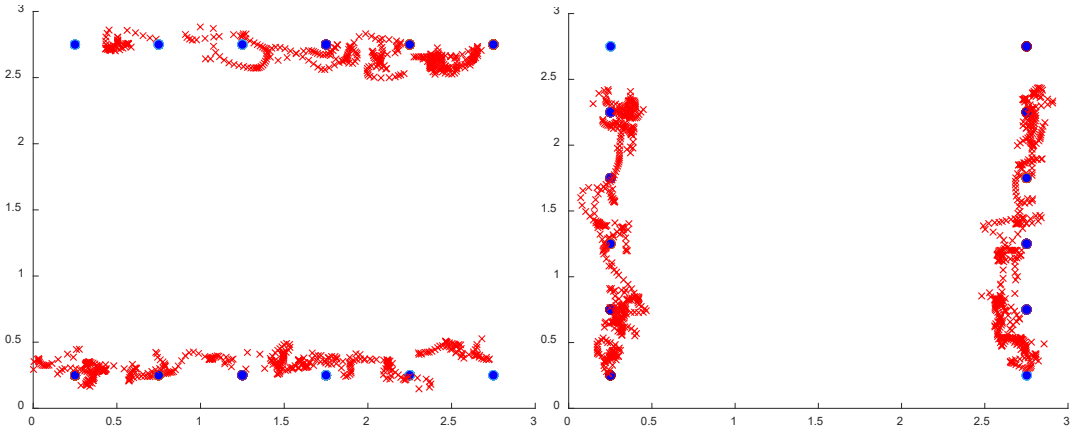
For Scenarios 1 and 2, two persons walk alongside horizontally across the area. However, a moving direction for each person is varied between these two scenarios. Two test subjects walk in the opposite direction in Scenario 1. In contrast, they walk in the same direction in Scenario 2. For Scenarios 3 and 4, they need to walk vertically across the area. Similarly, we also vary their walking directions in the same way as Scenarios 1 and 2. In addition, we vary the proximity between two participants to 2 meters and 0.5 meters. The settings of these scenarios can help us analyze the overall performance of our method and factors that can affect the accuracy of localization such as moving direction and proximity.

For both training and testing data, a video camera is employed to record video footage for labelling. Then, we extract coordinates from these footages. We train and test a deep learning model and our methods using Matlab 2020b on i5 3.2Ghz Intel quad-core CPU, Ram 16 Gb, and a GPU of NVIDIA RTX2070 Super with 8Gb video memory. The above Table 5-1. illustrates the optimal hyperparameter that we use to train a deep learning model.

5.5 Result and discussion

5.5.1 Overall Localization accuracy

The visualization of localization and tracking for two persons are shown in Figure 5-8 where the red paths and blue dots represent estimated locations and ground truth, respectively. We can see that the estimated locations for 2 persons follow along with the ground truth locations as both persons walk back and forth in the monitored area.



(a)Scenario 1 (b)Scenario 3

Figure 5-8 Example of PIR localization and tracking results. Red X represent estimated location at each time step. Blue dots represent the ground truth location

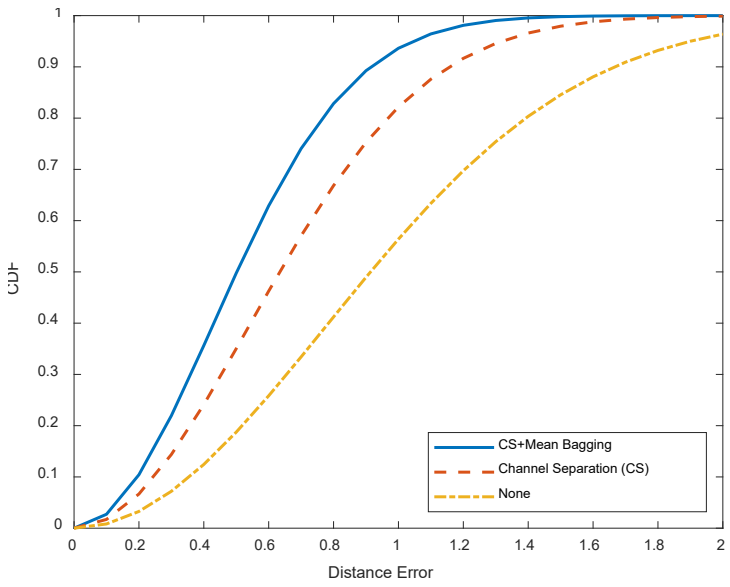


Figure 5-9 CDF of distance error for different appraoches

This indicates that our method can localize and track two persons simultaneously. The overall accuracy of our proposed method is 0.5269 meters for the distance error with a standard deviation of 0.2925 meters. In addition, CDF of the distance error shows that 70% of the distance error is less than 0.8 meters as shown in Figure 5-9. From Table 5-2, the location estimation results without applying any technique are shown in the first row. As expected, we receive the highest distance error of 1.0218 meters with the standard deviation of 0.4010 meters and CDF of the distance error shows that 70% error is less than 1.2 meters because input signals for each person contain information generated by the other person.

Table 5-2 The mean distance errors and their standard deviations for different settings

Setting	Distance Error	
	Mean	Std
None	1.0218m	0.4010m
Channel Separation (CS)	0.6668m	0.3677m
CS + Mean Bagging	0.5269m	0.2925m

As a result, a deep learning model incorporates irrelevant information into its calculation and results in inaccurate results. Once we apply our technique based on a signal combination to separate input signals for both persons, the localization results are improved approximately by 42%. The mean distance error is at 0.6668 meters and the standard deviation is 0.3677 meters. In addition, 70% of the distance error is less than 0.8 meters. The above results are estimated by the best-trained model. Despite the best model, we observe that poor estimation can be made because the presence of multiple persons can obscure patterns of the raw signals. Therefore, a single model is not sufficient to estimate accurate results. After we utilize a mean bagging technique, the localization results are further improved by 23% compared to the result without the mean bagging as shown in the third row of Table 5-2. Thus, we can see that the use of multiple models is more favourable than a single model in multi-person scenarios because we can take an advantage of multiple models to handle the ambiguity and unexpected patterns of the signals from multi-person and integrate their results to generate better final results.

5.5.2 Impact of the walking direction

As we can see two test subjects walk parallelly towards each other in Scenarios 1 and 3. Whereas, they walk parallelly in the same direction in Scenarios 2 and 4. Based on the results, it indicates that directions of moving persons can affect the accuracy of localization or location estimation. From Table 5-3 and Figure 5-10, the results of Scenarios 1 and 3 are more accurate than the results of Scenarios 2 and 4 at 2 meters. The mean distance error of Scenarios 1 and 3 is approximately 0.29 meters compared to the error of 0.47 meters for Scenarios 2 and 4.

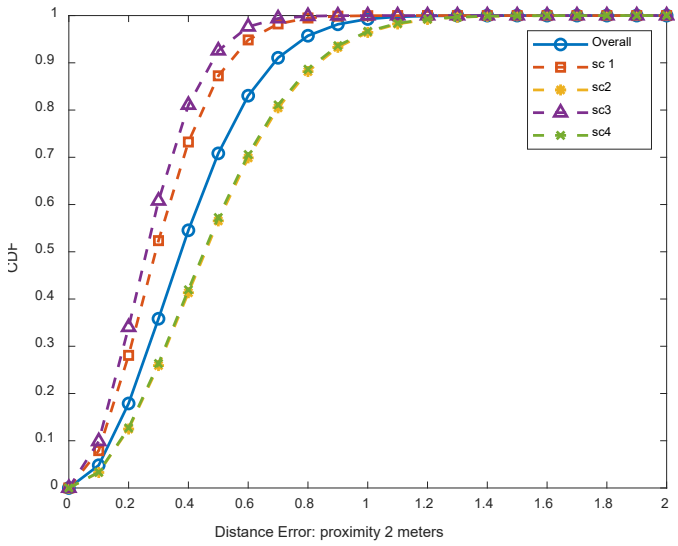


Figure 5-10 CDF of distance error for different scenario with proximity of 2 meters

Table 5-3 The mean distance errors and their standard deviations for different scenarios

Scenario	Distance Error with Different Proximity			
	2m		0.5m	
	Mean	Std	Mean	Std
1	0.3085m	0.2405m	0.6795m	0.3723m
2	0.4815m	0.2809m	0.6403m	0.4112m
3	0.2809m	0.1909m	0.6228m	0.4117m
4	0.4669m	0.3265m	0.6501m	0.3086m

When two persons walk toward each other, we observe that each person is detected by a different set of PIR sensors at each time step, i.e., at least two distinct sensors are required to detect each person. Thus, patterns of raw PIR signals are less likely to be interfered by the other person. On the other hand, two persons that walk alongside in the same direction cause interference in patterns of PIR signals because the same set of PIR sensors detect both persons at the same time.

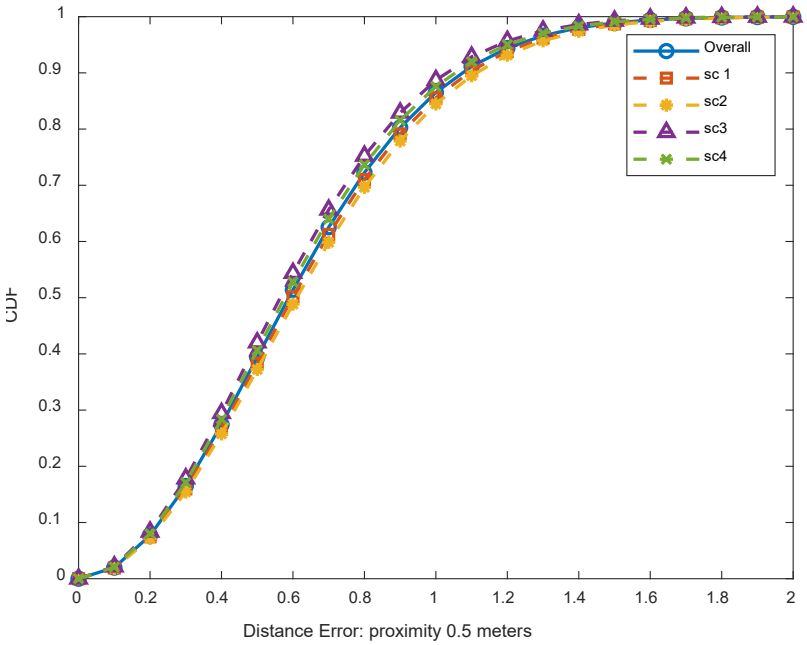


Figure 5-11 CDF of distance error for different scenario with proximity of 0.5 meters

5.5.3 Impact of the proximity between persons

The proximity between persons is one factor that can affect the accuracy of localization or location estimation. In our experiment, we vary the proximity between two persons to 0.5 meters and 2 meters. From Table 5-3, the overall mean distance error at 2 meters proximity is around 50% lower than the overall error at 0.5 meters. At 2 meters away, both persons can be detected separately by different sets of PIR sensors. Patterns of raw PIR signals are reasonably clear for a deep learning model to give accurate results. On the other hand, both persons have to share the same PIR sensors at 0.5 meters. The presence of both persons under the same sensor leads to ambiguity in PIR signal patterns. Thus, we obtain less accurate results. According to Table 5-3 and Figure 5-11, the results show that a walking direction doesn't affect localization accuracy significantly when a gap between two persons is 0.5 meters. Movements of a person with a better position can be easily captured compared to the other person with a bad position. Thus, signal patterns generated by movements at the better position become dominant in the raw PIR signal. From our results, we notice that a person with a better position tends to have better accuracy. From the above outcome, our method can handle multiple persons with proximity up to 0.5 meters and still obtain acceptable accuracy.

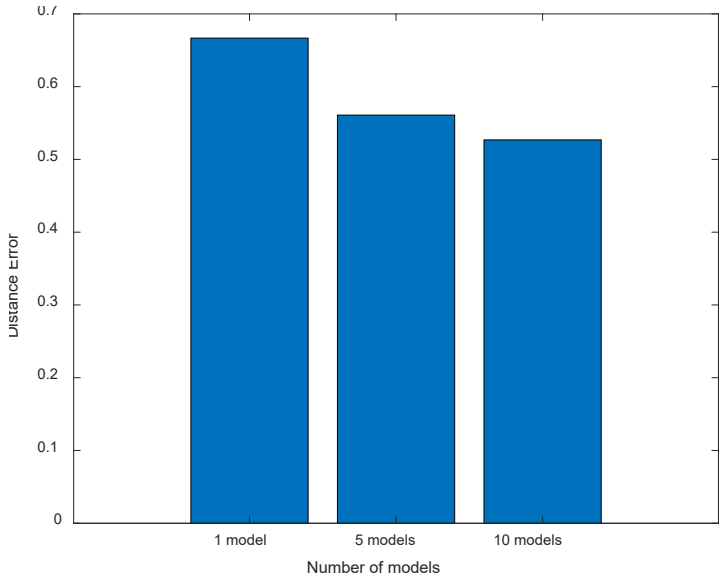


Figure 5-12 The effect of the number of models used in mean bagging

5.5.4 Impact of the number of models in mean bagging

We vary the number of models to see the effect on the accuracy. The result shows that an increase in the number of models can gradually improve the accuracy of location estimation. A single model can provide reasonably accurate estimations for a single person but its performance is affected considerably in a presence of multiple persons. In our experiment, it shows that a single model performs inconsistently from scenario to scenario. As a result, a single model produces the highest distance error as shown in Figure 5-12. When the number of models is increased to 5, we can see that the mean distance error is reduced by 17%. We observe that 5 models can improve the results of scenarios that a single model performs poorly. Using 10 models, we can improve the results by another 6%. According to the above results, an increase in the number of models in mean bagging can further reduce the variance of the location estimation. As a result, better localization results can be obtained compared to the estimation made by a single model.

5.5.5 Comparison with the baseline methods

In this section, we compare the performance of our deep CNN-LSTM model with other baseline methods including CNN-BiLSTM, CNN-GRU, and SVR. The results of different methods are shown in Table 5-4. CNN-LSTM achieves the distance error of 0.5269 meters which is the best

accuracy compared with other methods. CNN-BiLSTM can estimate locations of two persons with a distance error of 0.5734 meters which is around 8.5 % higher than CNN-LSTM. CNN-GRU obtains the distance error of 0.7752 meters and performs worse than both CNN-LSTM and CNN-BiLSTM. Due to a lack of a cell memory unit, GRU cannot capture complex time dependency of PIR analog signals very well, unlike the other RNN methods. SVR achieves the distance error of 0.8370 meters for the location estimation of two persons. The overall accuracy of SVR is 45%, 37%, and 7.6% lower than CNN-LSTM, CNN-BiLSTM, and CNN GRU respectively. Compared with deep learning methods, SVR relies on handcrafted features to estimate locations. Two people can be detected by the same PIR sensor at the same time thus introducing unexpected patterns into PIR analog outputs. The handcrafted features may be ambiguous and are not discriminative enough for SVR to achieve accurate results. In addition, SVR cannot capture the complex time dependency of PIR signals. As a result, SVR is found to achieve less accuracy than deep learning models. When a gap between two persons is at 2 meters, CNN-LSTM and CNN-BiLSTM can obtain a distance error of 0.4 meters approximately. In contrast, both CNN-GRU and SVR obtain worse results even when two persons stay 2 meters away from each other. When a gap is closer to 0.5 meters, all the methods are affected considerably for the same reasons as discussed previously in subsection 5.5.3. Our CNN-LSTM obtains a distance error of 0.6481 meters which is 9.7% more accurate than CNN-BiLSTM, 18% more accurate than CNN-GRU, and 35% more accurate than SVR. Therefore, our proposed model achieves the best result given this smaller gap.

Table 5-4 The mean distance errors for baseline deep learning and machine learning methods

methods	Overall Mean Distance Error	Mean Distance Error at 2m	Mean Distance error at 0.5m
CNN-LSTM	0.5269m	0.3841m	0.6481m
CNN-BiLSTM	0.5734m	0.4132m	0.7142m
CNN-GRU	0.7752m	0.7606m	0.7806m
SVR	0.8370m	0.7725m	0.8953m

5.6 Conclusion & Future works

In this chapter, we explore a deep learning approach for multi-person localization using PIR analog signals. Based on an assumption of a constant walking speed, our proposed method can separate raw PIR outputs generated by multiple persons into a set of inputs according to the number of persons. then, our deep CNN-LSTM architecture is used to estimate a coordinate for each person. The strategy to integrate multiple models using the mean bagging technique in our work can improve the localization results for multi-person. We conduct a set of experiments with 2 participants to evaluate our method and the results show that we can achieve a mean distance error of 0.53 meters and 70% of distance errors are within 0.8 meters.

However, there are some areas of improvement in our works. Even though we can estimate locations of 2 persons with a gap of 0.5 meters, it is still an issue because the location estimation of one person is much more accurate than the other one due to the ambiguity of PIR analog signals. In future work, we aim to investigate this issue to further improve the accuracy of our multi-person localization scheme. Our data collection was conducted in a temperature-controlled room. If there are changes in a room's temperature, the accuracy of our localization method can be affected. Other factors such as walk speed can also affect the accuracy because our trained deep learning model does not account for these factors. We will address these issues in our future work. In addition, our experiment is only conducted with 2 persons in a small-scale open area. The performance of our method in real-life locations such as residential homes or apartments is still questionable for us. Thus, we would like to extend our experiments to different locations.

Chapter 6 Conclusion and Future Work

In this thesis, three research questions regarding PIR-based multi-person localization have been addressed. In this chapter, the first section summarizes the contributions of this thesis. Then, the limitations of this research are discussed in the second section. Finally, the third section describes the future works.

6.1 Summary of Contributions

This thesis focused on PIR-based Localization using analog output. Three research questions regarding sensor design and data collection, single person localization, multi-person localization have been addressed. The contributions of the thesis can be summarized as follows:

Chapter 2 presents a complete survey of device-free indoor localization and tracking. The first section of chapter 2 describes the comprehensive overview of the device-free indoor localization and tracking as well as the taxonomy and classification of this field. Then, Device-free technologies, such as RF, infrared, etc., are described. In addition, the advantages and disadvantages of these technologies are highlighted. Then, techniques of multi-person localization are discussed in detail and the performance metrics are presented. Based on the metrics, related works are compared. Finally, the research challenges and future directions are discussed.

Chapter 3 proposes a novel design of a ceiling-based PIR sensor node that is easy to install in an indoor environment and has a good coverage range. The prototype of the proposed sensor node contains 5 PIR sensors. Every sensor is arranged so that its field of view overlaps with others to increase location distinguishability. As a result, the PIR node can cover a 9 m^2 area. Then, a testing monitored area is set up, and a set of walking scenarios is conducted to collect both training data and testing data. This thesis selects a set of deep learning methods, such as CNN, RNN, and CNN-RNN, to evaluate the applicability and feasibility of data collected from the proposed PIR node. The classification results show that the best deep learning method can achieve an accuracy of 85%, a kappa score of 0.84, and an f1 score of 0.77. Furthermore, this thesis further tests the data for coordinate estimation using deep learning methods and the particle filter approach. The best performer achieves a distance error of 0.25 meters. Therefore, the proposed PIR node is feasible and practical for indoor localization. Finally, the dataset of PIR-based localization is also available online on a public data repository for downloading.

Chapter 4 proposes a deep learning framework for single person indoor localization. The proposed framework comprises two major phases. The offline phase focuses on the training of a deep learning model. In this phase, A participant is instructed by a researcher to walk across a monitored area in different directions to collect a training dataset. This thesis proposes the deep CNN-LSTM architecture that consists of the upper CNN level and the lower LSTM level. The

purpose of the upper CNN level is to automate feature extraction from raw PIR outputs. Then, the lower LSTM level learns the temporal correlation of extracted features. The online phase focuses on estimating a person's location using the trained model. This thesis improves the accuracy of location estimation by updating the hidden state and cell state of the LSTM network with input signals in two consecutive time steps. As a result, the hidden state and cell state hold enough information, and the LSTM network can handle a shorter sequence efficiently in the online phase. Four walking scenarios are conducted to evaluate the proposed method. In addition, the proposed method is compared with other methods including CNN-GRU, CNN-BiLSTM, and Particle filtering. The results show that the accuracy of CNN-LSTM is considerably better than other tested methods. The proposed method achieves the mean distance error of 0.26m and 80% of distance errors are within 0.4 m which is better than existing works[22, 32, 33]

Chapter 5 proposes the PIR-based localization for multiple persons. In this chapter, the two-step localization method consists of channel separation and Person localization adopted. The first step focuses on differentiating raw PIR outputs generated by multiple persons and separating them into a set of single person inputs. Existing training sequences are segmented into a set of segmented profile sequence that is associated with a set of sensor activations. When raw PIR outputs are received, a set of signal combinations is generated through a summation of the candidate segmented sequences. Finally, the best match combination is used to identify a set of possible sensor activations and generate inputs for each person. The Person localization step takes inputs from the previous step and estimate the locations of multiple persons. The previously proposed CNN-LSTM architecture is employed for this task. Instead of using a single model, the mean bagging method is employed to integrate the results of multiple CNN-LSTM models. As a result, the proposed method can handle ambiguity and unexpected patterns from the presence of multiple persons and achieves improved results. A comprehensive set of experiments is conducted to evaluate the performance of the proposed method and to identify factors that affect multi-person localization. The results show that the proposed method can estimate the locations of two participants simultaneously with a mean distance error of 0.55 meters and 80% of distance errors are within 0.8 meters. Especially, the proposed method can locate two persons that stay 0.5 meters away from each other with a mean distance error of 0.68 meters. Lastly, the proposed method achieves similar or more accurate results than the existing works[32, 35, 36]. This is proof that

the proposed multi-person can achieve desirable localization results without any extra data collection.

6.2 Limitations

In this thesis, PIR-based localization has been addressed for both single and multiple persons. Although the proposed methods can deliver promising results, it is too early to consider this work complete because some limitations prevent this work from real-life use. By knowing these limitations, it can provide an outline for future works. The limitations of this thesis and their details are discussed below:

I. Variation of environment temperature and human factors

Each person has a different height, weight, and speed. Room temperature can also change depending on the season. These variations can affect raw PIR outputs. For example, a fast-moving person results in higher frequency outputs. A higher room temperature can reduce the amplitude of raw PIR outputs due to the small temperature gap between a human body and a room temperature. In this thesis, the localization methods are developed with an assumption of a constant walking speed and a room temperature is controlled. Specifically, the deep CNN-LSTM model is not trained to cope with these variations. As a result, the proposed localization methods achieve lower localization accuracy.

II. The number of samples in a dataset

The deep learning approach is adopted in this work. In general, it requires a large volume of data to learn and solve problems appropriately. In this research, human resource is very limited for both hardware development and data collection. It is not possible to collect a large amount of training data to train a deep CNN-LSTM model. As a result, it is quite difficult to determine whether the model is trained optimally or not.

III. Scalability of the proposed hardware and methods

Scalability is an important feature of every localization system. To be specific, the systems need to maintain their accuracy even in a large area or can handle a different number of persons. In this work, only one to two persons participate in the experiments in a small area. Thus, it is not possible to evaluate the scalability of the proposed system due to this limitation.

6.3 Future Work

PIR-based localization for multiple persons is challenging. The defined challenges have been addressed successfully in this thesis. However, there is room for improvements in the proposed localization method. To be specific, future works aim to overcome the limitations of this research. The details of the future works are outlined as follows:

I. Variations of environment temperature and human factors can affect the performance of the proposed PIR-based localization method. The proposed deep CNN-LSTM must be trained using a training dataset containing PIR sensor data from every situation. However, it is not practical to collect such diverse and very large training data. To solve this issue, Data augmentation can be a potential solution that can improve the performance of neural networks because this technique uses some prior knowledge to simulate more training datasets [228]. The future work will explore and apply this method to a PIR dataset.

II. The proposed hardware and localization method can estimate the locations of multiple persons accurately in a small-scale area. However, scalability is one of the desirable features for localization systems. The future work focuses on improving the scalability of the proposed PIR-based localization system. Multiple PIR nodes will be deployed in a larger area such as an apartment room and/or office room. When people move under multiple PIR nodes sequentially, multiple measurements or messages are generated from different nodes and are sent to a sink node. Thus, node cooperation will be investigated to ensure the correctness and efficiency of communication between nodes. In addition, data association will be addressed to assign a piece of information to a correct target.

III. In this thesis, prior knowledge such as starting locations and the number of occupants is required in the proposed PIR-based localization for multi-person. However, this information may not be available every time. Future work will address this issue to enable PIR-based localization for multi-person without any prior knowledge.

References

[1] F. Alam, N. Faulkner, and B. Parr, "Device-Free Localization: A Review of Non-RF Techniques for Unobtrusive Indoor Positioning," *IEEE Internet of Things Journal,* vol. 8, no. 6, pp. 4228-4249, 2021.

[2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal,* vol. 1, no. 1, pp. 22-32, 2014.

[3] L. Zhao, H. Huang, C. Su, S. Ding, H. Huang, Z. Tan, and Z. Li, "Block-Sparse Coding-Based Machine Learning Approach for Dependable Device-Free Localization in IoT Environment," *IEEE Internet of Things Journal,* vol. 8, no. 5, pp. 3211-3223, 2021.

[4] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A Survey on Wireless Indoor Localization from the Device Perspective," *ACM Computing Surveys,* vol. 49, no. 2, pp. 1-31, 2016.

[5] R. Zhang, X. Jing, S. Wu, C. Jiang, J. Mu, and F. R. Yu, "Device-Free Wireless Sensing for Human Detection: The Deep Learning Perspective," *IEEE Internet of Things Journal,* vol. 8, no. 4, pp. 2517-2539, 2021.

[6] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, "A Survey on Ambient Intelligence in Health Care," *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers,* vol. 101, no. 12, pp. 2470–2494, 2013.

[7] J. Wang, J. Xiong, H. Jiang, X. Chen, and D. Fang, "D-Watch: Embracing "bad" Multipaths for Device-Free Localization with COTS RFID Devices," in The 12th International Conference on emerging Networking EXperiments and Technologies, Irvine, CA, 2016, pp. 253-266.

[8] K. Ngamakeur, S. Yongchareon, J. Yu, and S. U. Rehman, "A Survey on Device-free Indoor Localization and Tracking in the Multi-resident Environment," *ACM Comput. Surv.,* vol. 53, no. 4, pp. Article 71, 2020.

[9] S. Palipana, B. Pietropaoli, and D. Pesch, "Recent advances in RF-based passive device-free localisation for indoor applications," *Ad Hoc Networks,* vol. 64, pp. 80-98, 2017.

[10] F. Adib, and D. Katabi, "See ThroughWalls with Wi-Fi!," in ACM SIGCOMM'13, Hong Kong, , 2013, pp. 75-86.

[11] D. Zhang, and L. M. Ni, "Dynamic clustering for tracking multiple transceiver-free objects," in 2009 IEEE International Conference on Pervasive Computing and Communications, Galveston, TX, USA, 2009, pp. 1-8.

[12] Z. Chen, H. Zou, J. Yang, H. Jiang, and L. Xie, "WiFi Fingerprinting Indoor Localization Using Local Feature-Based Deep LSTM," *IEEE Systems Journal,* vol. 14, no. 2, pp. 3001-3010, 2020.

[13] M. Valtonen, T. Vuorela, L. Kaila, and J. Vanhala, "Capacitive indoor positioning and contact sensing for activity recognition in smart homes," *J. Ambient Intell. Smart Environ.,* vol. 4, no. 4, pp. 305-334, 2012.

[14] T. Grosse-Puppendahl, X. Dellangnol, C. Hatzfeld, B. Fu, M. Kupnik, A. Kuijper, M. R. Hastall, J. Scott, and M. Gruteser, "Platypus—Indoor Localization and Identification through Sensing Electric Potential Changes in Human Bodies," in The 14th ACM International Conference on Mobile Systems, Applications, and Services, Singapore, 2016, pp. 17-30.

[15] J. D. Poston, R. M. Buehrer, and P. A. Tarazaga, "Indoor footstep localization from

structural dynamics instrumentation," *Mechanical Systems and Signal Processing,* vol. 88, pp. 224-239, 2017/05/01/, 2017.

[16] S. Drira, Y. Reuland, and I. F. C. Smith, "Model-based Interpretation of Floor Vibrations for Indoor Occupant Tracking," in 26th International Workshop on Intelligent Computing in Engineering, Leuven, Belgium, 2019.

[17] M. Shankar, J. B. Burchett, Q. Hao, B. D. Guenther, and D. J. Brady, "Human-tracking systems using pyroelectric infrared detectors," *Optical Engineering,* vol. 45, no. 10, pp. 106401, 2006.

[18] H. Qi, H. Fei, and X. Yang, "Multiple Human Tracking and Identification With Wireless Distributed Pyroelectric Sensor Systems," *IEEE Systems Journal,* vol. 3, no. 4, pp. 428-439, 2009.

[19] R.-S. Hsiao, D.-B. Lin, H.-P. Lin, S.-C. Cheng, and C.-H. Chung, "Indoor Target Detection and Localization in Pyroelectric Infrared Sensor Networks," in the 8th IEEE VTS Asia Pacific Wireless Communications Symposium, Singapore, Singapore, 2011.

[20] W. Zhou, F. Li, D. Li, X. Liu, and a. B. Li, "A human body positioning system with pyroelectric infrared sensor," *Int. J. Sensor Networks,* vol. 21, no. 2, pp. 108-105, 2016.

[21] K. N. Ha, K. C. Lee, and S. Lee, "Development of PIR sensor based indoor location detection system for smart home," in SICE-ICASE International Joint Conference, Busan, 2006, pp. 2162-2167.

[22] S. Tao, M. Kudo, H. Nonaka, and J. Toyama, "Person Localization and Soft Authentication Using an Infrared Ceiling Sensor Network," *Computer Analysis of Images and Patterns.* pp. 122-129.

[23] A. Milan, S. Roth, K. Schindler, and M. Kudo, "Privacy Preserving Multi-target Tracking," *Computer Vision - ACCV 2014 Workshops. ACCV 2014. Lecture Notes in Computer Science*, pp. 519-530: Springer, Cham, 2014.

[24] S. Narayana, R. V. Prasad, V. S. Rao, T. V. Prabhakar, S. S. Kowshik, and M. S. Iyer, "PIR Sensors: Characterization and Novel Localization Technique," in Proceedings of the 14th International Conference on Information Processing in Sensor Networks, New York, NY, USA, 2015, pp. 142-153.

[25] P. Zappi, E. Farella, and L. Benini, "Tracking Motion Direction and Distance With Pyroelectric IR Sensors," *IEEE Sensors Journal,* vol. 10, no. 9, pp. 1486-1494, 2010.

[26] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review,* vol. 40, pp. 100379, 2021/05/01/, 2021.

[27] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing,* vol. 151, pp. 107398, 2021/04/01/, 2021.

[28] T. Yang, P. Guo, W. Liu, X. Liu, and T. Hao, "Enhancing PIR-Based Multi-Person Localization Through Combining Deep Learning With Domain Knowledge," *IEEE Sensors Journal,* vol. 21, no. 4, pp. 4874-4886, 2021.

[29] T. Yang, P. Guo, W. Liu, X. Liu, and T. Hao, "DeepPIRATES: A Training-Light PIR-Based Localization Method With High Generalization Ability," *IEEE Access,* vol. 9, pp. 86054-86061, 2021.

[30] P. Zappi, E. Farella, and L. Benini, "Enhancing the spatial resolution of presence detection in a PIR based wireless surveillance network," in 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, London, 2007, pp. 295-300.

[31] J. Xiong, F. Li, N. Zhao, and N. Jiang, “Tracking and recognition of multiple human targets moving in a wireless pyroelectric infrared sensor network,” *Sensors (Basel),* vol. 14, no. 4, pp. 7209-28, Apr 22, 2014.

[32] X. Liu, T. Yang, S. Tang, P. Guo, and J. Niu, "From relative azimuth to absolute location: pushing the limit of PIR sensor based localization," *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, p. Article 1: Association for Computing Machinery, 2020.

[33] B. Yang, J. Luo, and Q. Liu, “A novel low-cost and small-size human tracking system with pyroelectric infrared sensor mesh network,” *Infrared Physics & Technology,* vol. 63, pp. 147-156, 2014.

[34] S. Fan, Y. Wu, C. Han, and X. Wang, “A Structured Bidirectional LSTM Deep Learning Method For 3D Terahertz Indoor Localization,” in IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, 2020, pp. 2381-2390.

[35] B. Yang, and M. Zhang, “Credit-Based Multiple Human Location for Passive Binary Pyroelectric Infrared Sensor Tracking System: Free From Region Partition and Classifier,” *IEEE Sensors Journal,* vol. 17, no. 1, pp. 37-45, 2017.

[36] S. Tao, M. Kudo, B.-N. Pei, H. Nonaka, and J. Toyama, “Multiperson Locating and Their Soft Tracking in a Binary Infrared Sensor Network,” *IEEE Transactions on Human-Machine Systems,* vol. 45, no. 5, pp. 550-561, 2015.

[37] A. L. Ballardini, L. Ferretti, S. Fontana, A. Furlan, and D. G. Sorrenti, “An Indoor Localization System for Telehomecare Applications,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 46, no. 10, pp. 1445-1455, 2016.

[38] J. Han, C. Qian, X. Wang, D. Ma, J. Zhao, P. Zhang, W. Xi, and Z. Jiang, “Twins: Device-free Object Tracking using Passive Tags,” in IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, Toronto, ON, 2014, pp. 469-476.

[39] F. Li, C. Zhao, G. Ding, J. Gong, C. Liu, and F. Zhao, “A reliable and Accurate Indoor localization method using phone inertial sensors,” in Proceedings of the 2012 ACM Conference on Ubiquitous Computing, New York, NY, USA, 2012, pp. 421–430.

[40] Z. Jiang, J. Zhao, J. Han, S. Tang, J. Zhao, and W. Xi, “Wi-Fi Fingerprint Based Indoor Localization without Indoor Space Measurement,” in 2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems, Hangzhou, 2013, pp. 384-392.

[41] V. Moreno, M. A. Zamora, and A. F. Skarmeta, “A Low-Cost Indoor Localization System for Energy Sustainability in Smart Buildings,” *IEEE Sensors Journal,* vol. 16, no. 9, pp. 3246-3262, 2016.

[42] S. Tao, M. Kudo, and H. Nonaka, “Privacy-preserved behavior analysis and fall detection by an infrared ceiling sensor network,” *Sensors (Basel),* vol. 12, no. 12, pp. 16920-36, Dec 07, 2012.

[43] D. Zhang, J. Ma, Q. Chen, and L. M. Ni, “An RF based system for tracking transceiver-free objects,” in Fifth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'07), White Plains, NY, USA, 2007, pp. 135-144.

[44] D. Lieckfeldt, J. You, and D. Timmermann, “Passive Tracking of Transceiver-Free Users with RFID,” in Intelligent Interactive Assistance and Mobile Multimedia Computing, 2009, pp. 319–329.

[45] H. Kavalionak, E. Carlini, A. Lulli, C. Gennaro, G. Amato, C. Meghini, and L. Ricci, “A prediction-based distributed tracking protocol for video surveillance,” in 2017 IEEE 14th

International Conference on Networking, Sensing and Control (ICNSC), 2017, pp. 140-145.

[46] Q. Hao, D. J. Brady, B. D. Guenther, J. B. Burchett, M. Shankar, and S. Feller, "Human Tracking With Wireless Distributed Pyroelectric Sensors," *IEEE Sensors Journal,* vol. 6, no. 6, pp. 1683-1696, 2006.

[47] Y. Kasama, and T. Miyazaki, "Simultaneous Estimation of the Number of Humans and their Movement Loci in a Room Using Infrared Sensors," in 2012 26th International Conference on Advanced Information Networking and Applications Workshops, Fukuoka, Japan, 2012, pp. 508-513.

[48] T. W. Hnat, E. Griffiths, R. Dawson, and K. Whitehouse, "Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors," in Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, Toronto, Ontario, Canada, 2012, pp. 309-322.

[49] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews),* vol. 37, no. 6, pp. 1067-1080, 2007.

[50] F. Gu, X. Hu, M. Ramezani, D. Acharya, K. Khoshelham, S. Valaee, and J. Shang, "Indoor Localization Improved by Spatial Context—A Survey," *ACM Comput. Surv.,* vol. 52, no. 3, pp. 1-35, 2019.

[51] F. Khelifi, A. Bradai, A. Benslimane, P. Rawat, and M. Atri, "A Survey of Localization Systems in Internet of Things," *Mobile Networks and Applications,* vol. 24, no. 3, pp. 761-785, June 01, 2019.

[52] M. Youssef, M. Mah, and A. Agrawala, "Challenges Device-free Passive localization for wireless Environment," in the 13th annual ACM international conference on Mobile computing and networking, Montréal, Québec, Canada, 2007, pp. 222-229

[53] D. Lieckfeldt, J. You, and D. Timmermann, "Exploiting RF-Scatter: Human Localization with Bistatic Passive UHF RFID-Systems," in 2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Marrakech, Morocco, 2009, pp. 179-184.

[54] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint Activity Recognition and Indoor Localization With WiFi Fingerprints," *IEEE Access,* vol. 7, pp. 80058-80068, 2019.

[55] Y. Kilic, H. Wymeersch, A. Meijerink, M. J. Bentum, and W. G. Scanlon, "An Experimental Study of UWB Device-Free Person Detection and Ranging," in 2013 IEEE International Conference on Ultra-Wideband (ICUWB), 2013, pp. 43-48.

[56] I. Vlasenko, I. Nikolaidis, and E. Stroulia, "The Smart-Condo: Optimizing Sensor Placement for Indoor Localization," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 45, no. 3, pp. 436-453, 2015.

[57] W. Chen, M. Guan, L. Wang, R. Ruby, and K. Wu∗, "FLoc: Device-free Passive Indoor Localization in Complex Environments," in IEEE ICC 2017 Ad-Hoc and Sensor Networking Symposium, Paris, France, 2017, pp. 1-6.

[58] Y. Liu, Y. Zhao, L. Chen, J. Pei, and J. Han, "Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays," *IEEE Transactions on Parallel and Distributed Systems,* vol. 23, no. 11, pp. 2138-2149, 2012.

[59] W. Ruan, Q. Z. Sheng, L. Yao, T. Gu, M. Ruta, and L. Shangguan, "Device-free Indoor

Localization and Tracking through Human-Object Interactions," in 2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), Coimbra, Portugal, 2016, pp. 1-9.

[60] W. Ruan, Q. Z. Sheng, L. Yao, X. Li, N. J. G. Falkner, and L. Yang, "Device-free human localization and tracking with UHF passive RFID tags: A data-driven approach," *Journal of Network and Computer Applications,* vol. 104, pp. 78-96, 2018.

[61] L. Ma, M. Liu, H. Wang, Y. Yang, N. Wang, and Y. Zhang, "WallSense: Device-Free Indoor Localization Using Wall-Mounted UHF RFID Tags," *Sensors (Basel),* vol. 19, no. 1, pp. 1-16, Dec 25, 2018.

[62] M. Moussa, and M. Youssef, "Smart devices for smart environments: Device-free passive detection in real environments," in 2009 IEEE International Conference on Pervasive Computing and Communications, Galveston, TX, USA, 2009, pp. pp. 1-6.

[63] J. Wang, H. Jiang, J. Xiong, X. Chen, D. Fang, K. Jamieson, 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, New York, 2016, pp. 234-256.

[64] D. Zhang, Y. Liu, and L. M. Ni, "RASS: A Real-time, Accurate and Scalable System for Tracking Transceiver-free Objects," in 2011 IEEE International Conference on Pervasive Computing and Communications, Seattle 2011, pp. 197-204.

[65] J. Wilson, and N. Patwari, "Radio Tomographic Imaging with Wireless Networks," *IEEE Transactions on Mobile Computing,* vol. 9, no. 5, pp. 621-632, 2010.

[66] J. Wang, D. Fang, X. Chen, Z. Yang, T. Xing, and L. Cai, "LCS: Compressive Sensing based Device-Free Localization for Multiple Targets in Sensor Networks," in 2013 Proceedings IEEE INFOCOM, Turin, 2013, pp. 145-149.

[67] Y. Kilic, H. Wymeersch, A. Meijerink, M. J. Bentum, and W. G. Scanlon, "Device-Free Person Detection and Ranging in UWB Networks," *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING,* vol. 8, pp. 1-12, 2013.

[68] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances," *Sensors (Basel),* vol. 16, no. 5, May 16, 2016.

[69] B. Gulmezoglu, M. B. Guldogan, and S. Gezici, "Multiperson Tracking With a Network of Ultrawideband Radar Sensors Based on Gaussian Mixture PHD Filters," *IEEE Sensors Journal,* vol. 15, no. 4, pp. 2227-2237, 2015.

[70] F. Liang, F. Qi, Q. An, H. Lv, F. Chen, Z. Li, and J. Wang, "Detection of Multiple Stationary Humans Using UWB MIMO Radar," *Sensors (Basel),* vol. 16, no. 11, Nov 16, 2016.

[71] D. Yang, W. Sheng, and R. Zeng, "Indoor Human Localization Using PIR Sensors and Accessibility Map," in The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, Shenyang, China, 2015, pp. 577-581.

[72] X. Luo, T. Liu, B. Shen, Q. Chen, L. Gao, and X. Luo, "Human Indoor Localization Based on Ceiling Mounted PIR Sensor Nodes," in 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, 2016, pp. 868-874.

[73] A. Braun, H. Heggen, and R. Wichert, "CapFloor – A Flexible Capacitive Indoor Localization System," *Evaluating AAL Systems Through Competitive Benchmarking.*

Indoor Localization and Tracking: International Competition, EvAAL 2011, Competition in Valencia, Spain, July 25-29, 2011, and Final Workshop in Lecce, Italy, September 26, 2011. Revised Selected Papers, S. Chessa and S. Knauth, eds., pp. 26-35, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[74] B. Fu, F. Kirchbuchner, J. von Wilmsdorff, T. Grosse-Puppendahl, A. Braun, and A. Kuijper, "Performing indoor localization with electric potential sensing," *Journal of Ambient Intelligence and Humanized Computing,* vol. 10, no. 2, pp. 731-746, 2018.

[75] X. Bian, G. D. Abowd, and J. M. Rehg, "Using Sound Source Localization in a Home Environment," in In Proceedings of the Third international conference on Pervasive Computing, Berlin, Heidelberg, 2005, pp. 19-36.

[76] P. Zappi, E. farella, and L. Benini, "Pyroelectric infrared sensors based distance estimation," in 2008 IEEE Sensors, Lecce, Italy, 2008, pp. 716-719.

[77] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "FreeDetector: Device-Free Occupancy Detection with Commodity WiFi," in 2017 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops), San Diego, CA, USA, 2017, pp. 1-5.

[78] A. E. Kosba, A. Saeed, and M. Youssef, "RASID: A robust WLAN device-free passive motion detection system," in 2012 IEEE International Conference on Pervasive Computing and Communications, Lugano, Switzerland, 2012, pp. 531-533.

[79] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Pilot: Passive Device-Free Indoor Localization Using Channel State Information," in 2013 IEEE 33rd International Conference on Distributed Computing Systems, Philadelphia, PA, USA, 2013, pp. 236-245.

[80] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "FIMD: Fine-grained Device-free Motion Detection," in 2012 IEEE 18th International Conference on Parallel and Distributed Systems, Singapore, Singapore, 2012, pp. 229-235.

[81] L. Gong, W. Yang, Z. Zhou, D. Man, H. Cai, X. Zhou, and Z. Yang, "An adaptive wireless passive human detection via fine-grained physical layer information," *Ad Hoc Networks,* vol. 38, pp. 38-50, 2016.

[82] T. Xin, B. Guo, Z. Wang, P. Wang, J. C. K. Lam, V. Li, and Z. Yu, "FreeSense: A Robust Approach for Indoor Human Detection Using Wi-Fi Signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.,* vol. 2, no. 3, pp. 1-23, 2018.

[83] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," in 2014 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS), Hsinchu, Taiwan, 2014, pp. pp. 1-8.

[84] J. Lv, D. Man, W. Yang, X. Du, and M. Yu, "Robust WLAN-Based Indoor Intrusion Detection Using PHY Layer Information," *IEEE Access,* vol. 6, pp. 30117-30127, 2018.

[85] Z. Yuan, S. Wu, X. Yang, and A. He, "Device-Free Stationary Human Detection with WiFi in Through-the-Wall Scenarios," *Wireless and Satellite Systems.* pp. 201-208.

[86] Y. Zhao, N. Patwari, J. M. Phillips, and S. Venkatasubramanian, "Radio tomographic imaging and tracking of stationary and moving people via kernel distance," in 2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Philadelphia, PA, USA, 2013, pp. 229-240.

[87] S. Palipana, P. Agrawal, and D. Pesch, "Channel State Information Based Human

Presence Detection using Non-linear Techniques," in BuildSys '16 Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, Palo Alto, CA, USA, 2016, pp. 177-186.

[88] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Towards Omnidirectional Passive Human Detection," in 2013 Proceedings IEEE INFOCOM, Turin, 2013, pp. 3057-3065.

[89] F. Xiao, X. Xie, H. Zhu, L. Sun, and R. Wang, "Invisible Cloak Fails: CSI-based Passive Human Detection," in CSAR '15 Proceedings of the 1st Workshop on Context Sensing and Activity Recognition, Seoul, South Korea, 2015, pp. 19-23.

[90] C. Li, J. Ling, J. Li, and J. Lin, "Accurate Doppler Radar Noncontact Vital Sign Detection Using the RELAX Algorithm," *IEEE Transactions on Instrumentation and Measurement,* vol. 59, no. 3, pp. 687-695, 2010.

[91] N. Patwari, J. Wilson, S. Ananthanarayanan, S. K. Kasera, and D. R. Westenskow, "Monitoring Breathing via Signal Strength in Wireless Networks," *IEEE Transactions on Mobile Computing,* vol. 13, no. 8, pp. 1774-1786, 2014.

[92] O. J. Kaltiokallio, s. Yigitler, R. Jntti, and N. Patwari, "Non-invasive respiration rate monitoring using a single COTS TX-RX pair," in Proceedings of the 13th international symposium on Information processing in sensor networks, Berlin, Germany, 2014, pp. 59-70.

[93] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals," in 2014 IEEE Real-Time Systems Symposium, 2014, pp. 346-355.

[94] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-Invasive Detection of Moving and Stationary Human With WiFi," *IEEE Journal on Selected Areas in Communications,* vol. 33, no. 11, pp. 2329-2342, 2015.

[95] X. Liang, and J. Deng, "Detection of stationary human target via contactless radar networks," *Journal of Ambient Intelligence and Humanized Computing,* vol. 10, no. 8, pp. 3193-3200, August 01, 2019.

[96] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-Free Presence Detection and Localization With SVM and CSI Fingerprinting," *IEEE Sensors Journal,* vol. 17, no. 23, pp. 7990-7999, 2017.

[97] C. Han, Q. Tan, L. Sun, H. Zhu, and J. Guo, "CSI Frequency Domain Fingerprint-Based Passive Indoor Human Detection," *Information,* vol. 9, no. 4, pp. 1-14, 2018.

[98] S.-H. Fang, C.-C. Li, W.-C. Lu, Z. Xu, and Y.-R. Chien, "Enhanced Device-Free Human Detection: Efficient Learning From Phase and Amplitude of Channel State Information," *IEEE Transactions on Vehicular Technology,* vol. 68, no. 3, pp. 3048-3051, 2019.

[99] S. D. Domenico, M. D. Sanctis, E. Cianca, and M. Ruggieri, "WiFi-based through-the-wall presence detection of stationary and moving humans analyzing the doppler spectrum," *IEEE Aerospace and Electronic Systems Magazine,* vol. 33, no. 5-6, pp. 14-19, 2018.

[100] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation, Oakland, CA, 2015, pp. 279-292.

[101] T. Li, Y. Wang, L. Song, and H. Tan, "On Target Counting by Sequential Snapshots of Binary Proximity Sensors," in Wireless Sensor Networks. EWSN 2015, Cham, 2015, pp. 19-34.

[102] L. Song, and Y. Wang, "Multiple Target Counting and Tracking using Binary Proximity

Sensors: Bounds, Coloring, and Filter," in Proceedings of the 15th ACM international symposium on Mobile ad hoc networking and computing, Philadelphia, Pennsylvania, USA, 2014, pp. 397-406.

[103] Y. Wang, L. Song, Z. Gu, and D. Li, "IntenCT: Efficient Multi-Target Counting and Tracking By Binary Proximity Sensors," in 2016 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), London, UK, 2016, pp. 1-9.

[104] F. Wahl, M. Milenkovic, and O. Amft, "A Distributed PIR-based Approach for Estimating People Count in Office Environments," in 2012 IEEE 15th International Conference on Computational Science and Engineering, 2012, pp. 640-647.

[105] F. Wahl, M. Milenkovic, and O. Amft, "A green autonomous self-sustaining sensor node for counting people in office environments." pp. 203-207.

[106] I. Sabek, M. Youssef, and A. V. Vasilakos, "ACE: An Accurate and Efficient Multi-Entity Device-Free WLAN Localization System," *IEEE Transactions on Mobile Computing,* vol. 14, no. 2, pp. 261-273, 2015.

[107] Q. Wang, H. Yigitler, R. Jantti, and X. Huang, "Localizing Multiple Objects Using Radio Tomographic Imaging Technology," *IEEE Transactions on Vehicular Technology,* vol. 65, no. 5, pp. 3641-3656, 2016.

[108] S. Nannuru, Y. Li, Y. Zeng, M. Coates, and B. Yang, "Radio-Frequency Tomography for Passive Indoor Multitarget Tracking," *IEEE Transactions on Mobile Computing,* vol. 12, no. 12, pp. 2322-2333, 2013.

[109] N. Zamzami, M. Amayri, N. Bouguila, and S. Ploix, "Online Clustering for Estimating Occupancy in an Office Setting," in 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), 2019, pp. 2195-2200.

[110] Y. P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, and M. A. Little, "Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction," in In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), New York, NY, USA, 2016, pp. 1016-1027.

[111] D. Wu, D. Chen, K. Xing, and X. Cheng, "A Statistical Approach for Target Counting in Sensor-Based Surveillance Systems," in 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 2012, pp. 226-234.

[112] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy Estimation Using Only WiFi Power Measurements," *IEEE Journal on Selected Areas in Communications,* vol. 33, no. 7, pp. 1381-1393, 2015.

[113] S. D. Domenico, M. D. Sanctis, E. Cianca, and G. Bianchi, "A Trained-once Crowd Counting Method Using Differential WiFi Channel State Information," in Proceedings of the 3rd International on Workshop on Physical Analytics, Singapore, Singapore, 2016, pp. 37-42.

[114] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using WiFi," in IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, Toronto, ON, Canada, 2014, pp. 361-369.

[115] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, "Sparse Target Counting and Localization in Sensor Networks Based on Compressive Sensing," in 2011 Proceedings IEEE INFOCOM, Shanghai, China, 2011, pp. 2255-2263.

[116] J. Wang, D. Fang, Z. Yang, H. Jiang, X. Chen, T. Xing, and L. Cai, "E-HIPA: An Energy-

Efficient Framework for High-Precision Multi-Target-Adaptive Device-Free Localization," *IEEE Transactions on Mobile Computing,* vol. 16, no. 3, pp. 716-729, 2017.

[117] B. Sun, Y. Guo, N. Li, and D. Fang, "Multiple Target Counting and Localization Using Variational Bayesian EM Algorithm in Wireless Sensor Networks," *IEEE Transactions on Communications,* vol. 65, no. 7, pp. 2985-2998, 2017.

[118] Y. Guo, B. Sun, N. Li, and D. Fang, "Variational Bayesian Inference-based Counting and Localization for Off-Grid Targets with Faulty Prior Information in Wireless Sensor Networks," *IEEE Transactions on Communications*, pp. 1-1, 2017.

[119] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based Person Identification in Smart Spaces," in 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Vienna, 2016, pp. 1-12.

[120] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "WiFi-ID: Human Identification Using WiFi Signal," in 2016 International Conference on Distributed Computing in Sensor Systems (DCOSS), 2016, pp. 75-82.

[121] F. Hong, X. Wang, Y. Yang, Y. Zong, Y. Zhang, and Z. Guo, "WFID: Passive Device-free Human Identification Using WiFi Signal," in MOBIQUITOUS 2016 Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Hiroshima, Japan, 2016, pp. 47-56.

[122] J. Lv, W. Yang, D. Man, X. Du, M. Yu, and M. Guizani, "Wii: Device-free Passive Identity Identification via WiFi Signals," in GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Singapore, Singapore, 2017.

[123] J. Lv, W. Yang, and D. Man, "Device-Free Passive Identity Identification via WiFi Signals," *Sensors (Basel),* vol. 17, no. 11, Nov 2, 2017.

[124] J. Yun, and S. S. Lee, "Human movement detection and identification using pyroelectric infrared sensors," *Sensors (Basel),* vol. 14, no. 5, pp. 8057-81, May 05, 2014.

[125] J. Xiong, F. Li, and J. Liu, "Fusion of Different Height Pyroelectric Infrared Sensors for Person Identification," *IEEE Sensors Journal,* vol. 16, no. 2, pp. 436-446, 2016.

[126] J. Yan, P. Lou, R. Li, J. Hu, and J. Xiong, "Research on the Multiple Factors Influencing Human Identification Based on Pyroelectric Infrared Sensors," *Sensors (Basel),* vol. 18, no. 2, Feb 16, 2018.

[127] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling Identification and Behavioral Sensing in Homes using Radio Reflections," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk, 2019, pp. 1-13.

[128] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mID: Tracking and Identifying People with Millimeter Wave Radar," in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2019, pp. 33-40.

[129] K. Hyun Hee, H. Kyoung Nam, L. Suk, and L. Kyung Chang, "Resident Location-Recognition Algorithm Using a Bayesian Classifier in the PIR Sensor-Based Indoor Location-Aware System," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 39, no. 2, pp. 240-245, 2009.

[130] D. Zhang, J. Zhou, M. Guo, J. Cao, and T. Li, "TASA: Tag-Free Activity Sensing Using RFID Tag Arrays," *IEEE Transactions on Parallel and Distributed Systems,* vol. 22, no. 4, pp. 558-570, 2011.

[131] T. Hosokawa, M. Kudo, H. Nonaka, and J. Toyama, "Soft authentication using an infrared ceiling sensor network," *Pattern Analysis and Applications,* vol. 12, no. 3, pp. 237-249, 2008.

[132] R. Peng, and M. L. Sichitiu, "Angle of Arrival Localization for Wireless Sensor Networks," in 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, Reston, VA, 2006, pp. 374-382.

[133] D. Niculescu, and B. Nath, "Ad Hoc Positioning System (APS) Using AOA," in IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies, 2003, pp. 1734-1743.

[134] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-MUSIC: Accurate Device-Free Indoor Localization," in the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, HEIDELBERG, GERMANY, 2016, pp. 196-207.

[135] D. Lieckfeldt, J. You, and D. Timmermann, "Characterizing the Influence of Human Presence on Bistatic Passive RFID-System," in 2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Marrakech, Morocco, 2009, pp. 338-343.

[136] Z. Yang, K. Huang, X. Guo, and G. Wang, "A real-time device-free localization system using correlated RSS measurements," *EURASIP Journal on Wireless Communications and Networking,* vol. 2013, no. 1, pp. 186, 2013/07/09, 2013.

[137] J. Wang, Q. Gao, Y. Yu, P. Cheng, L. Wu, and H. Wang, "Robust Device-Free Wireless Localization Based on Differential RSS Measurements," *IEEE Transactions on Industrial Electronics,* vol. 60, no. 12, pp. 5943-5952, 2013.

[138] Y. Guo, K. Huang, N. Jiang, X. Guo, Y. Li, and G. Wang, "An Exponential-Rayleigh Model for RSS-Based Device-Free Localization and Tracking," *IEEE Transactions on Mobile Computing,* vol. 14, no. 3, pp. 484-494, 2015.

[139] J. Wang, Q. Gao, M. Pan, X. Zhang, Y. Yu, and H. Wang, "Toward Accurate Device-Free Wireless Localization With a Saddle Surface Model," *IEEE Transactions on Vehicular Technology,* vol. 65, no. 8, pp. 6665-6677, 2016.

[140] Z. Wang, H. Liu, S. Xu, X. Bu, and J. An, "A Diffraction Measurement Model and Particle Filter Tracking Method for RSS-Based DFL," *IEEE Journal on Selected Areas in Communications,* vol. 33, no. 11, pp. 2391-2403, 2015.

[141] C. Liu, D. Fang, Z. Yang, H. Jiang, X. Chen, W. Wang, T. Xing, and L. Cai, "RSS Distribution-Based Passive Localization and Its Application in Sensor Networks," *IEEE Transactions on Wireless Communications,* vol. 15, no. 4, pp. 2883-2895, 2016.

[142] J. Wang, B. Xie, D. Fang, X. Chen, C. Liu, T. Xing, and W. Nie, "Accurate Device-Free Localization with Little Human Cost," in SmartObjects '15 Proceedings of the 1st International Workshop on Experiences with the Design and Implementation of Smart Objects, Paris, France, 2015, pp. 55-60.

[143] M. Nicoli, V. Rampa, S. Savazzi, and S. Schiaroli, "Device-free Localization of Multiple Targets," in 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 2016, pp. 1-5.

[144] B. Wagner, N. Patwari, and D. Timmermann, "Passive RFID tomographic imaging for device-free user localization," in 9th Workshop on Positioning, Navigation and Communication, Dresden, 2012, pp. 120-125.

[145] J. Wilson, and N. Patwari, "See-Through Walls: Motion Tracking Using Variance-Based

Radio Tomography Networks," *IEEE Transactions on Mobile Computing,* vol. 10, no. 5, pp. 612-621, 2011.

[146] O. Kaltiokallio, M. Bocca, and N. Patwari, "Enhancing the accuracy of radio tomographic imaging using channel diversity," in 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012), Las Vegas, NV, USA, 2012, pp. 254-262.

[147] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Propagation Modeling for Radio Frequency Tomography in Wireless Networks," *IEEE Journal of Selected Topics in Signal Processing,* vol. 8, no. 1, pp. 55-65, 2014.

[148] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian, "Multiple Target Tracking with RF Sensor Networks," *IEEE Transactions on Mobile Computing,* vol. 13, no. 8, pp. 1787-1800, 2014.

[149] W. Jie, G. Qinghua, C. Peng, Y. Yan, X. Kefei, and W. Hongyu, "Lightweight Robust Device-Free Localization in Wireless Networks," *IEEE Transactions on Industrial Electronics,* vol. 61, no. 10, pp. 5681-5689, 2014.

[150] B. Wei, A. Varshney, N. Patwari, W. Hu, T. Voigt, and C. T. Chou, "dRTI: directional radio tomographic imaging," in Proceedings of the 14th International Conference on Information Processing in Sensor Networks, Seattle, Washington, 2015, pp. 166–177.

[151] M. A. Kanso, and M. G. Rabbat, "Compressed RF Tomography for Wireless Sensor Networks: Centralized and Decentralized Approaches," in Distributed Computing in Sensor Systems. DCOSS 2009. , Berlin, Heidelberg, 2009, pp. 173–186.

[152] T. Liu, X. Luo, and Z. Liang, "Enhanced Sparse Representation-Based Device-Free Localization with Radio Tomography Networks," *Journal of Sensor and Actuator Networks,* vol. 7, no. 1, pp. 7, 2018.

[153] S. Xu, H. Liu, F. Gao, and Z. Wang, "Compressive Sensing Based Radio Tomographic Imaging with Spatial Diversity," *Sensors (Basel),* vol. 19, no. 3, Jan 22, 2019.

[154] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless RSSI fingerprinting localization," *Signal Processing,* vol. 131, pp. 235-244, 2017.

[155] M. Seifeldin, A. Saeed, A. E. Kosba, A. El-Keyi, and M. Youssef, "Nuzzer: A Large-Scale Device-Free Passive Localization System for Wireless Environments," *IEEE Transactions on Mobile Computing,* vol. 12, no. 7, pp. 1321-1334, 2013.

[156] C. Xu, B. Firner, Y. Zhang, R. Howard, J. Li, and X. Lin, "Improving RF-based device-free passive localization in cluttered indoor environments through probabilistic classification methods." pp. 209-220.

[157] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-Free Wireless Localization and Activity Recognition: A Deep Learning Approach," *IEEE Transactions on Vehicular Technology,* vol. 66, no. 7, pp. 6258-6267, 2017.

[158] C. Xu, B. Firner, W. Trappe, R. S. Moore, R. Howard, F. Zhang, Y. Zhang, and NingAn, "SCPL: Indoor Device-Free Multi-Subject Counting and Localization Using Radio Signal Strength," in 2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Philadelphia, 2013, pp. 79-90.

[159] J. Wang, X. Zhang, Q. Gao, X. Ma, X. Feng, and H. Wang, "Device-Free Simultaneous Wireless Localization and Activity Recognition With Wavelet Feature," *IEEE Transactions on Vehicular Technology,* vol. 66, no. 2, pp. 1659-1669, 2017.

[160] J. He, Y. Hu, X. Liu, C. Liu, Y. Peng, and Xianjia Meng, "LiReT: An Fine-Grained Self-Adaption Device-Free Localization with Little Human Effort," in 2017 IEEE

International Conference on Smart Computing (SMARTCOMP), Hong Kong, 2017, pp. 1-3.

[161] É. L. Souza, E. F. Nakamura, and R. W. Pazzi, "Target Tracking for Sensor Networks," *ACM Computing Surveys,* vol. 49, no. 2, pp. 1-31, 2016.

[162] V. Fox, J. Hightower, L. Lin, D. Schulz, and G. Borriello, "Bayesian filtering for location estimation," *IEEE Pervasive Computing,* vol. 2, no. 3, pp. 24-33, 2003.

[163] X. Luo, B. Shen, X. Guo, G. Luo, and G. Wang, "Human tracking using ceiling pyroelectric infrared sensors," in 2009 IEEE International Conference on Control and Automation, Christchurch, New Zealand, 2009, pp. 1716-1721.

[164] B. Yang, X. Li, and J. Luo, "A novel multi-human location method for distributed binary pyroelectric infrared sensor tracking system: Region partition using PNN and bearing-crossing location," *Infrared Physics & Technology,* vol. 68, pp. 35-43, 2015.

[165] Z. Wang, H. Liu, S. Xu, X. Bu, and J. An, "Bayesian Device-Free Localization and Tracking in a Binary RF Sensor Network," *Sensors (Basel),* vol. 17, no. 5, Apr 27, 2017.

[166] J. Wilson, and N. Patwari, "A Fade-Level Skew-Laplace Signal Strength Model for Device-Free Localization with Wireless Networks," *IEEE Transactions on Mobile Computing,* vol. 11, no. 6, pp. 947-958, 2012.

[167] J. Wang, Q. Gao, Y. Yu, X. Zhang, and X. Feng, "Time and Energy Efficient TOF-Based Device-Free Wireless Localization," *IEEE Transactions on Industrial Informatics,* vol. vol. 12, no. 1, pp. 158-168, Feb. 2016, 2015.

[168] B. Yang, Q. Wei, and M. Zhang, "Multiple human location in a distributed binary pyroelectric infrared sensor network," *Infrared Physics & Technology,* vol. 85, pp. 216-224, 2017.

[169] M. F. Bugallo, T. Lu, and P. M. Djuric, "Target Tracking by Multiple Particle Filtering," in 2007 IEEE Aerospace Conference, Big Sky, MT, USA, 2007, pp. 1-7.

[170] D. Zhang, K. Lu, R. Mao, Y. Feng, Y. Liu, Z. Ming, and L. M. Ni, "Fine-Grained Localization for Multiple Transceiver-Free Objects by using RF-Based Technologies," *IEEE Transactions on Parallel and Distributed Systems,* vol. 25, no. 6, pp. 1464-1475, 2014.

[171] T. S., K. M., N. H., and T. J, "Person Localization and Soft Authentication Using an Infrared Ceiling Sensor Network," *Computer Analysis of Images and Patterns. CAIP 2011. Lecture Notes in Computer Science*, D.-P. D. Real P., Molina-Abril H., Berciano A., Kropatsch W., ed., pp. 122-129: Springer, Berlin, Heidelberg, 2011.

[172] J. Wang, X. Chen, D. Fang, C. Q. Wu, Z. Yang, and T. Xing, "Transferring Compressive-Sensing-Based Device-Free Localization Across Target Diversity," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 4, pp. 2397-2409, 2015.

[173] T. Liu, and J. Liu, "Feature-specific biometric sensing using ceiling view based pyroelectric infrared sensors," *EURASIP Journal on Advances in Signal Processing,* vol. 2012, no. 1, pp. 206, 2012/09/25, 2012.

[174] B. Yang, Y. Lei, and B. Yan, "Distributed Multi-Human Location Algorithm Using Naive Bayes Classifier for a Binary Pyroelectric Infrared Sensor Tracking System," *IEEE Sensors Journal,* vol. 16, no. 1, pp. 216-223, 2016.

[175] J. Wang, Q. Gao, M. Pan, and Y. Fang, "Device-Free Wireless Sensing: Challenges, Opportunities, and Applications," *IEEE Network,* vol. 32, no. 2, pp. 132-137, 2018.

[176] V. Losing, B. Hammer, and H. Wersing, "Incremental on-line learning: A review and

comparison of state of the art algorithms," *Neurocomputing,* vol. 275, pp. 1261-1274, 2018.

[177] H. Zou, X. Lu, H. Jiang, and L. Xie, "A fast and precise indoor localization algorithm based on an online sequential extreme learning machine," *Sensors (Basel),* vol. 15, no. 1, pp. 1804-24, Jan 15, 2015.

[178] X. Jiang, J. Liu, Y. Chen, D. Liu, Y. Gu, and Z. Chen, "Feature Adaptive Online Sequential Extreme Learning Machine for lifelong indoor localization," *Neural Computing and Applications,* vol. 27, no. 1, pp. 215-225, January 01, 2016.

[179] K. Ohara, T. Maekawa, Y. Kishino, Y. Shirai, and F. Naya, "Transferring positioning model for device-free passive indoor localization," in UbiComp '15 Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 2015, pp. 885-896.

[180] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu, "See Through Walls with COTS RFID System!," in the 21st Annual International Conference on Mobile Computing and Networking, Paris, France 2015, pp. 487-499.

[181] L. Yao, Q. Z. Sheng, X. Li, T. Gu, M. Tan, X. Wang, S. Wang, and W. Ruan, "Compressive Representation for Device-Free Activity Recognition with Passive RFID Signal Strength," *IEEE Transactions on Mobile Computing,* vol. 17, no. 2, pp. 293-306, 2018.

[182] B. Mukhopadhyay, S. Sarangi, S. Srirangarajan, and S. Kar, "Indoor localization using analog output of pyroelectric infrared sensors," in 2018 IEEE Wireless Communications and Networking Conference (WCNC), 2018, pp. 1-6.

[183] L. Wu, Y. Wang, and H. Liu, "Occupancy Detection and Localization by Monitoring Nonlinear Energy Flow of a Shuttered Passive Infrared Sensor," *IEEE Sensors Journal,* vol. 18, no. 21, pp. 8656-8666, 2018.

[184] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in 2015 IEEE Wireless Communications and Networking Conference (WCNC), 2015, pp. 1666-1671.

[185] V. Rampa, M. Nicoli, C. Manno, and S. Savazzi, "EM Model-Based Device-Free Localization of Multiple Bodies," *Sensors,* vol. 21, no. 5, 2021.

[186] X. Wang, X. Wang, and S. Mao, "Deep Convolutional Neural Networks for Indoor Localization with CSI Images," *IEEE Transactions on Network Science and Engineering,* vol. 7, no. 1, pp. 316-327, 2020.

[187] M. Youssef, and A. Agrawala, "The Horus WLAN location determination system," in Proceedings of the 3rd international conference on Mobile systems, applications, and services, Seattle, Washington, 2005, pp. 205–218.

[188] F. Lemic, J. Martin, C. Yarp, D. Chan, V. Handziski, R. Brodersen, G. Fettweis, A. Wolisz, and J. Wawrzynek, "Localization as a feature of mmWave communication." pp. 1033-1038.

[189] T. Wei, and X. Zhang, "mTrack: High-Precision Passive Tracking Using Millimeter Wave Radios," in Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, Paris, France, 2015, pp. 117-129.

[190] C. Cai, H. Pu, P. Wang, Z. Chen, and J. Luo, "We Hear Your PACE: Passive Acoustic Localization of Multiple Walking Persons," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.,* vol. 5, no. 2, pp. Article 55, 2021.

[191] H. Yang, W. D. Zhong, C. Chen, A. Alphones, and P. Du, "QoS-Driven Optimized

Design-Based Integrated Visible Light Communication and Positioning for Indoor IoT Networks," *IEEE Internet of Things Journal,* vol. 7, no. 1, pp. 269-283, 2020.

[192] J. Kemper, and D. Hauschildt, "Passive infrared localization with a Probability Hypothesis Density filter," in 2010 7th Workshop on Positioning, Navigation and Communication, Dresden, Germany, 2010, pp. 68-76.

[193] C. Wei-Han, and M. Hsi-Pin, "A fall detection system based on infrared array sensors with tracking capability for the elderly at home," in 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), Boston, MA, USA, 2015, pp. 428-434.

[194] S. Narayana, V. Rao, R. V. Prasad, A. K. Kanthila, K. Managundi, L. Mottola, and T. V. Prabhakar, "LOCI: Privacy-aware, Device-free, Low-power Localization of Multiple Persons using IR Sensors." pp. 121-132.

[195] J. G. Rohra, B. Perumal, S. J. Narayanan, P. Thakur, and R. B. Bhatt, "User Localization in an Indoor Environment Using Fuzzy Hybrid of Particle Swarm Optimization & Gravitational Search Algorithm with Neural Networks," *Proceedings of Sixth International Conference on Soft Computing for Problem Solving.* pp. 286-295.

[196] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent Neural Networks for Accurate RSSI Indoor Localization," *IEEE Internet of Things Journal,* vol. 6, no. 6, pp. 10639-10651, 2019.

[197] G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "Deep Canonical Time Warping for Simultaneous Alignment and Representation Learning of Sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 5, pp. 1128-1138, 2018.

[198] Z. Luo, F. Branchaud-Charron, C. Lemaire, J. Konrad, S. Li, A. Mishra, A. Achkar, J. Eichel, and P. Jodoin, "MIO-TCD: A New Benchmark Dataset for Vehicle Classification and Localization," *IEEE Transactions on Image Processing,* vol. 27, no. 10, pp. 5129-5141, 2018.

[199] E. Maiorana, "Deep learning for EEG-based biometric recognition," *Neurocomputing,* vol. 410, pp. 374-386, 2020/10/14/, 2020.

[200] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery,* vol. 33, no. 4, pp. 917-963, 2019/07/01, 2019.

[201] X. Huang, G. C. Fox, S. Serebryakov, A. Mohan, P. Morkisz, and D. Dutta, "Benchmarking Deep Learning for Time Series: Challenges and Directions," in 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 5679-5682.

[202] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *CoRR,* vol. abs/1803.01271, /, 2018.

[203] Y. Wang, F. Liu, and L. Yang, "EEG-Based Depression Recognition Using Intrinsic Time-scale Decomposition and Temporal Convolution Network," in The Fifth International Conference on Biological Information and Biomedical Engineering, Hangzhou, China, 2021, pp. Article 5.

[204] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," *ArXiv,* vol. abs/1412.3555, 2014.

[205] R. Mutegeki, and D. S. Han, "A CNN-LSTM Approach to Human Activity Recognition,"

in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2020, pp. 362-366.

[206] S. Zhang, "Language Processing Model Construction and Simulation Based on Hybrid CNN and LSTM," *Computational Intelligence and Neuroscience,* vol. 2021, pp. 2578422, 2021/07/07, 2021.

[207] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomedical Signal Processing and Control,* vol. 47, pp. 312-323, 2019/01/01/, 2019.

[208] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR,* vol. abs/1412.6980, 2015.

[209] Z. Wang, and T. Oates, "Imaging time-series to improve classification and imputation," in Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015, pp. 3939–3945.

[210] P. Barsocchi, A. Calabrò, A. Crivello, S. Daoudagh, F. Furfari, M. Girolami, and E. Marchetti, "COVID-19 & privacy: Enhancing of indoor localization architectures towards effective social distancing," *Array,* vol. 9, pp. 100051, 2021/03/01/, 2021.

[211] N. Faulkner, F. Alam, M. Legg, and S. Demidenko, "Device-Free Localization Using Privacy-Preserving Infrared Signatures Acquired From Thermopiles and Machine Learning," *IEEE Access,* vol. 9, pp. 81786-81797, 2021.

[212] Y. S. Lin, P. H. Huang, and Y. Y. Chen, "Deep Learning-Based Hepatocellular Carcinoma Histopathology Image Classification: Accuracy Versus Training Dataset Size," *IEEE Access,* vol. 9, pp. 33144-33157, 2021.

[213] A. Bailly, C. Blanc, É. Francis, T. Guillotin, F. Jamal, B. Wakim, and P. Roy, "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models," *Computer Methods and Programs in Biomedicine,* vol. 213, pp. 106504, 2022/01/01/, 2022.

[214] A. Nessa, B. Adhikari, F. Hussain, and X. N. Fernando, "A Survey of Machine Learning for Indoor Positioning," *IEEE Access,* vol. 8, pp. 214945-214965, 2020.

[215] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology,* vol. 49, no. 4, pp. 764-766, 2013/07/01/, 2013.

[216] O. Kaltiokallio, R. Hostettler, and N. Patwari, "A Novel Bayesian Filter for RSS-Based Device-Free Localization and Tracking," *IEEE Transactions on Mobile Computing,* vol. 20, no. 3, pp. 780-795, 2021.

[217] P. Bahl, and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064), 2000, pp. 775-784 vol.2.

[218] M. Moussa, and M. Youssef, "Smart Devices for smart environments: Device-free passive detection in real environments," in 2009 IEEE International Conference on Pervasive Computing and Communications, 2009, pp. 1-6.

[219] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-Grained Indoor Fingerprinting System," in 2012 21st International Conference on Computer Communications and Networks (ICCCN), 2012, pp. 1-7.

[220] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, "MonoPHY: Mono-stream-based

device-free WLAN localization via physical layer information," in 2013 IEEE Wireless Communications and Networking Conference (WCNC), 2013, pp. 4546-4551.

[221] X. Wang, L. Gao, and S. Mao, "PhaseFi: Phase Fingerprinting for Indoor Localization with a Deep Learning Approach," in 2015 IEEE Global Communications Conference (GLOBECOM), 2015, pp. 1-6.

[222] B. Chen, S.-l. Zhao, and P.-y. Li, "Application of Hilbert-Huang Transform in Structural Health Monitoring: A State-of-the-Art Review," *Mathematical Problems in Engineering,* vol. 2014, pp. 317954, 2014/06/02, 2014.

[223] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," *ArXiv,* vol. abs/1607.08022, 2016.

[224] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using Convolutional Networks," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 648-656.

[225] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in Proceedings of the 14th international conference on artificial intelligence and statistics, Proceedings of Machine Learning Research, 2011, pp. 315-323.

[226] N. E. Huang, Z. Shen, S. Long, M. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences,* vol. 454, pp. 903 - 995, 1998.

[227] T. Yang, P. Guo, W. Liu, and X. Liu, "DeepPIRATES: Enabling Deployment-Independent Supervised PIR-Based Localization," in 2020 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA), Shanghai, China, 2020, pp. 151-156.

[228] L. Perez, and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.