Radio Frequency Sensing Systems for Artificial Intelligence of Things

by

Chao Yang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2022

Keywords: Internet of Things, Deep Learning, Indoor Localization, Vital Sign Monitoring,
RFID, Human Pose, Channel State Information

Approved by

Shiwen Mao, Chair, Samuel Ginn Endowed Professor of Electrical and Computer Engineering
Thaddeus Roppel, Associate Professor of Electrical and Computer Engineering
Mark Nelms, Professor and Chair of Electrical and Computer Engineering
Xiaowen Gong, Assistant Professor of Electrical and Computer Engineering
Yang Zhou, Assistant Professor of Computer Science and Software Engineering

Abstract

With the rapid development of artificial intelligence, the Internet of Things (IoT) has evolved into artificial intelligence (AIoT). The development of an effective and low-cost human health detection system has attracted intensive research interest from both academia and industrial areas, such as vital sign monitoring, indoor localization, and. To achieve low cost and high accuracy for smart health systems, Radio Frequency Identification (RFID) based techniques have been utilized for human vital signs measurement. In addition, the RFID system could be used for effective indoor localization and human pose estimation. Compared with a vital sign signal, a human pose signal, as a complicated 3-Dimensional signal, could be more challenging. With the foundation of developing these systems with multiple RF devices, we also propose a technology-agnostic RF sensing system for human activity recognition, which could be performed on multiple RF platforms. The dissertation includes all Radio Frequency(RF) sensing systems we have developed during Ph.D. study.

Acknowledgments

This work is supported in part by the US NSF under Grant CNS-2107190, ECCS-1923163, ECCS-1923717, CNS-1822055, and CNS-1702957 and through the Wireless Engineering Research and Education Center (WEREC) at Auburn University. I am very grateful to my advisor Prof. Shiwen Mao for his guidance. He inspired me a lot whenever I had problem with my research work. I also appreciate my other committee members, Prof. Xiaowen Gong, Prof. Thaddus Roppel, and Prof. Mark Nelms. They gave me lots of suggestions and helpful comment, so that I can modify and improve my thesis research. Besides these professors, I would like to thank my team members, particularly, Xuyu Wang, who helped me a lot in my research field. I can't make a success in my research and other projects without his help. Last but no the least, I want to thank my family members and friends, who give support with their love during my whole study in Auburn.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

With the development of Artificial Intelligence of Things (AIoT), numerous Radio Frequency sensing systems have been proposed to achieve the non-intrusive and intelligent sensing systems for human beings [1–7]. Healthcare has become an important problem [9–11]. As a key component of healthcare, detection and monitoring of vital signs are performed in traditional healthcare systems with dedicated equipment, such as capnography [13]. In addition, detecting an abnormality may require considerable efforts and experience. For example, many breathing disorders during sleep are hard to detect and diagnose because of the lacking of a suitable monitoring system. One of these disorders is obstructive sleep apnea, which can cause long-term damage on human health such as heart disease, stroke, and high blood pressure. Therefore, autonomous, unobstructive, and low-cost vital sign monitoring systems with the abnormality identification functionality are highly desirable, which can help a person to detect sleep disorders and reduce the danger of, e.g., sudden infant death syndrome (SIDS) for sleeping infants [12].

Different wireless signals have been used to detect vital signs. The idea is to capture the small signal caused by the rise and fall of the chest (or heart beats) from the received wireless signal during breathing. For example, Radar-based systems have been developed to monitor human respiration, such as ultra-wideband radar [30] and frequency modulated continuous wave (FMCW) radar [14]. However, a Radar-based system requires expensive and complicated hardware, and may operate on a wide spectrum, which may not be proper for many application

scenarios. Other WiFi based techniques have the advantage of low-cost and easy deployment, which can monitor human respiration and heartbeat by analyzing the received signal strength (RSS) [16] or channel state information (CSI) [19, 20]. Although WiFi based systems can effectively monitor human respiration, the WiFi signals are easily affected by changes in the environment, such as movements of a person nearby, thus leading to a relatively lower accuracy of breathing estimation and apnea detection. In a recent work [11], we developed SonarBeat, which is a smartphone app that exploits ultrasound signals for respiration monitoring.

RFID sensing systems have drawn increasing attention recently, which have been employed for object tracking [22], drone relays [24], orientation estimation [23], and recently, for breathing monitoring [27]. These works mainly exploit the RFID phase information, which is collected from the low level data with an RFID reader. For example, Tagoram system leverages phase information for real-time tracking of RFID tags with a differential augmented hologram technique [22]. RFly uses drones as relays for battery-free networks using RFID phase information [24]. As related to our work, Tagyro calibrates phase values from all channels to one channel for orientation estimation, where the system requires to measure the phase offset offline [23]. In addition, Tagbreathe monitors breathing signals by grouping the signals with the same channel index and using the calculated displacement in each channel [27]. However, this method does not work very well for US RFID systems, which operate in the frequency range from 902.5 MHz to 927.5 MHz with 50 channels as required by the FCC. This is because channel hoping among 50 different frequencies will cause a considerably larger latency, which make it much harder to obtain a breathing signal.

Driving fatigue detection is another primary component for smart health care. According to the NHTSA report, 795 lives have been lost due to greatly reduced if an effective driving fatigue alarm system is available. However, most drowsy driving events are hard to detect with the existing technologies in commodity vehicles. Thus, there is a compelling demand for an effective driving fatigue detection system, which can accurately detect driving fatigue and alarm drivers to avoid accidents.

Driving fatigue detection is a popular topic in the research community for quite some time, and different types of signals have been utilized to address this issue, such as electroencephalogram (EEG), video, WiFi, and ultra sound. EEG signal can achieve high fatigue detection accuracy [48], but the driver is required to wear multiple special devices, which is not suitable for long time driving. In contrast, computer vision based techniques only need to collect eyelid movements using a camera [49]. Although the required hardware, i.e., a camera, is cheaper than that in EEG based techniques, the system performance is heavily affected when the driver wears sunglasses and may require sufficient lighting in the car. Several device-free approaches have also been proposed. For example, WiFi signals can be used to detect driving fatigue by extracting driver's movements and breathing rate from channel state information (CSI) [50]. However, the WiFi signal is sensitive to the interference from surroundings, such as the movements of passengers and objects outside the vehicle. Features of drowsy driving can be detected by acoustic-based technique as well [51], but mitigating the influence of interference from passengers is still a big challenge. Since RFID is a near-field communication technique, interference from passengers or surroundings of the vehicle can hardly affect the sensing performance. Furthermore, the cost of the system is lower than other existing approaches. Therefore, RFID is highly suited for driving fatigue detection within the in-car environment. However, there are still many challenges to make RFID based driving fatigue detection work, such as the discontinuity in collected channel state data caused by frequency hopping, and effective feature extraction for driving fatigue detection.

RF based indoor localization techniques is also a big topic for the RF based sensing systems. The Radio Frequency Identification (RFID) technology has been regarded as an effective and low-cost solution for many emerging IoT applications. Existing works on RFID tag localization include received signal strength indication (RSSI) based and phase based methods. These works mainly focus on locating a single tag, i.e., one tag is located in a time slot. For RSSI based methods, numerous RFID reference tags are deployed at known locations. By comparing the RSSI data with reference tags, the position of the target tag can be determined [31]. In fact, RSSI values are raw channel information and are not stable, due to factors such as multipath propagation, tag's orientation, RFID reader power, etc. RSSI based methods usually

do not achieve high accuracy in indoor localization. On the other hand, phase based methods have been developed for estimating distance and direction of arrival (DOA) [33]. However, the measured phase is periodic, which leads to phase ambiguity and makes it less useful. Moreover, considerable measured phase errors are caused by the RFID reader antennas and the tag. To address these two issues, the synthetic aperture radar (SAR) technique is proposed for DOA estimation by moving the RFID reader antenna around [21]. The second solution is the hologram technique, which can compute the probability of each known position as the tag source in an area of interest and then choose the most likely position as the tag location [22,34]. Another solution is the hyperbolic based method for distance estimation, which can locate a static tag [35]. However, this solution does not achieve high localization accuracy due to the limited number of RFID reader antennas. In addition, RFind system can obtain higher localization accuracy with a large virtual bandwidth to estimate time-of-flight, but it requires a special hardware [91].

The sensing of human motion also contribute a lot for the development of the Artificial intelligence of things. Compared with the vital sign signal, which is a one dimensional signal, the human pose signal could be a much more complicated signal. This is because the signal for the human pose tracking is 3D signal, and each limb could be part of signal source for the human pose tracking. Actually, human pose tracking has attracted great interest in recent years, because it is useful for numerous applications such as human-computer interaction, video surveillance, and somatosensory games. The advances in human pose tracking have been mainly driven by the new developments in computer vision, from two-dimensional (2D) systems [173] to three-dimensional (3D) realtime systems [174]. However, the vision-based techniques often raise concerns of security and privacy. For example, many wireless security cameras are easily hacked by malicious users [175]. The collected video data for pose tracking could also be illegally intercepted. Several radio frequency (RF) sensing schemes have been proposed to address the privacy concern in human pose tracking, using various RF sensing techniques such as Frequency-Modulated Continuous Wave (FMCW) radar [176], millimeter wave (mmWave) radar [177], WiFi [131, 178], and RFID [157, 179, 180]. Compared with computer vision-based techniques, RF sensing-based human pose tracking does not require for sufficient lighting, does not require a line-of-sight path between the subject and camera (i.e., capable of

getting around obstacles or even through walls), and more important, the privacy of users can be better protected.

In this dissertation, we present the AutoTag system, a recurrent variational **Auto**encoder for respiration rate estimation and unsupervised detection of apnea with commercial passive UHF RFID **Tag**s. We also propose an RFID based system, to detect the nodding movements of drivers, which is a key indicator of fatigue and one of the most dangerous motions during drowsy driving. Besides, we propose a RFID based indoor localization techniques termed Sparsetag, and the RFID based pose tracking systems termed RFID pose.The highly accurate detection performance of the proposed system is validated by intensive experimental study.

1.2 Literature Review

1.2.1 RF based Vital Sign Monitoring

Nowadays, many RF based systems have been proposed to monitor vital signals of human, which are developed on different types of platforms, including Radar systems, WiFi systems, and RFID based systems. The Radar technique is a straightforward way to identify the fluctuation of human chest caused by respiration, because Radar can directly monitor the distance variation between the human chest and the device antenna. One of the representative works in this category leverages an FMCW radar to monitor respiration rate and heart rate for multiple users simultaneously [14]. Furthermore, other Radar based systems have also been proposed to measure human respiration, including Doppler radar [15] and ultra wideband (UWB) Radar [30]. These radar based systems can accurately detect vital signs, and the influence caused by surroundings is limited. However, since the special hardware is essential to such systems, the cost of such systems are usually high, hampering their wide deployment such as in homes.

To achieve low cost and ease of deployment for vital sign monitoring, WiFi based techniques have been utilized to measure both human breathing rate and heart rate. The human respiration and heartbeat can be extracted by analyzing variations in WiFi channels, for example, the RSS as in Ubibreathe [16]. However, the Ubibreathe system requires the patient to stand between the transmitter and the receiver, while some other CSI based techniques have no such strict requirements. Different from RSS based systems, CSI can provide fine-grained channel inforamtion, and can achieve higher resolution and sensitivity than RSS for monitoring human vital signs. One of the CSI based techniques can leverage amplitude of CSI to monitor breathing rate and heart rate when the patient is sleeping [17]. In addition to the amplitude, the CSI phase information can also carry human respiration signal [18]. Furthermore, the TensorBeat system can estimate respiration rates for multiple persons crowed in a small space [19], by incorporating tensor decomposition on the collected CSI phase data. Although WiFi based techniques can measure human vital signs with off-the-shelf WiFi devices, the accuracy is easily affected by the surrounding environment, because of broadcasting nature and long range of

WiFi transmissions. To address this issue, some RFID based systems like TagBreathe are developed to track human respiration by analyzing the RFID response data collected at an RFID reader [27]. Since the passive UHF RFID tags are low-cost and are easily attachable to human body, the RFID system can monitor human vital signs at a low cost and is resilient to interference from the unstable environment.

Apart from RFID based vital sign monitoring systems, various sensing systems are proposed by leveraging the data extracted from the low level protocol in RFID systems, such as RSS and phase values, which have been utilized for many applications, e.g., indoor localization. For RSS based methods, one of the representative techniques estimates the tag position by comparing the RSS from the target tag with reference tags [31]. Moreover, another technique can obtain the refined tag position by utilizing the characteristics of the coupling effect on RSS [32]. However, due to the low resolution of RSS, developing an RSS based localization scheme with high accuracy is challenging. Thus, many phase based localization techniques are proposed. One of the typical phase based methods has been developed for estimating distance with direction of arrival (DOA) [33], but the result has a relatively large ambiguity because of the periodicity of measured phase. To remove the ambiguity, some other techniques are proposed to obtain the more accurate position than the typical method by leveraging the aperture radar technique [21] and hologram technique [22, 34] [35]. Besides indoor localization, RFID tags are also widely used for other sensing techniques such as remote control of drones [24–26], object orientation estimation [23], remote temperature measurement [36], and gesture recognition [37].

Recently, passive RFID tags, as a kind of wearable sensors, have attracted increasing interest because of its low-cost and easy deployment features. RFID based sensing has been used for many applications, such as user authentication [68], material identification [69], object orientation estimation [23], vibration sensing [70], and anomaly detection [71]. For indoor localization [22,34,57] and gesture recognition [72,73], the RFID based techniques are mainly focused on the analysis of low level data collected at the reader. For example, the received signal strength (RSS) has been utilized for tag localization in [31], while the phase values have been used to recognize different kinds of gestures [55]. In addition, vital signs can also be

detected by the low level data. Specifically, TagBreathe is the first work to estimate breathing rates using RFID tags [27], while TagSheet uses RFID Tags for breathing monitoring and sleep posture recognition [74]. Even heart rate variability can be assessed with an RFID tag array attached to the human body. However, these vital sign monitoring systems are not suitable for detecting drowsiness in a driving environment, because the small vital sign signal could be easily overwhelmed by vehicle vibration and driving movements. The work presented in this paper makes a first attempt on RFID based driving fatigue detection, where commercial RFID tags are utilized for detecting the nodding movement of the driver.

RF based health sensing systems have been developed that employ Radar, WiFi, and RFID techniques. Radar based vital sign monitoring systems include frequency modulated continuous wave (FMCW) radar [80] and Doppler Radar [81]. However, they usually require customized hardware and operate over a wide spectrum. WiFi based systems mainly use received signal strength (RSS) and channel state information (CSI). For example, UbiBreathe [82] and mmVital [83] utilize WiFi RSS at 2.4 GHz and 60 GHz, respectively. To improve accuracy, CSI based systems leverage the amplitude or phase difference information of CSI for estimating single or multiple persons' breathing and heart rates [17, 19, 84, 85]. Moreover, several bimodal CSI data based systems have been proposed to tackle the weak breathing signals at some special positions [86, 87, 104].

Several RFID based breathing monitoring systems have been proposed. For example, RFID tags have been used for breathing rate estimation in [27], breathing and heart rates estimation in [88], and breathing monitoring and sleeping posture recognition in [74]. Furthermore, the RF-ECG system is proposed for heart rate variability assessment using an RFID tag array [89]. To mitigate the frequency hopping offset in FCC-compliant RFID systems, the AutoTag system is proposed for breathing monitoring and apnea detection with a variational autoencoder [8, 58]. However, these existing systems are designed for the indoor, static environment; they may not be effective in the highly dynamic, highly noisy driving environment.

Recently, WiFi based [50], acoustic based [78], and UWB based [66] systems have been developed for breath monitoring in driving environments. In fact, these existing systems are

sensitive to the environmental interference, such as the body movements of the driver himself/herself and of the passengers, due to their relatively large transmission ranges.

In addition to vital sign monitoring, RFID tags have also been applied for many other applications, such as indoor localization [22, 57], user authentication [68], material identification [69], object orientation estimation [23], vibration sensing [70], anomaly detection [71], and drone localization and navigation [53, 90]. To overcome the low accuracy when RSS values are used [31], recent works are mainly focused on the phase for indoor localization, which can be used to derive the distance and direction of arrival (DOA). To solve the phase ambiguity problem, synthetic aperture radar (SAR) [21] and the hologram techniques [22, 34] are proposed.

The RFind system estimates time-of-flight with a special hardware to achieve high localization precision [91].

1.2.2 RF based Indoor Localization

Table 1.1: Features in Different RFID Tag Localization Techniques

Localization Technique	Hardware Modification	Tag Array	Antenna Array	Dynamic Tags or Antennas
LANDMARC	No	No	Yes	No
RF-IDraw	No	No	Yes	No
Tagoram	No	No	No	Yes
RFfind	Yes	No	No	No
SparseTag	No	Yes	No	No

With the rapid development of Internet of Things, indoor localization attracts increasing attention in recent years. As an RFID-based indoor localization system, our work is closely related to the RF based localization techniques in prior work. In this section we mainly focus on WiFi based techniques and RFID based techniques.

WiFi signals are widely utilized for indoor localization because of its low-cost, wide coverage, and ubiquitous deployment. Among various techniques, Angle of Arrival (AoA) is a typical method to estimate the location of the transmitter [115], but the accurate AoA is hard to estimate because of the multipath effect on the WiFi signal. To mitigate the multipath effect,

antenna array-based systems are proposed to estimate the angle of multiple incoming paths of WiFi signal and distinguish the Line-of-sight (LoS) component [116, 117]. In addition, rather than directly calculate the AoA of the LOS path, some prior works leverage machine learning to estimate the position of the transmitter by learning the location features from collected channel state data. For example, Radar is a WiFi fingerprinting scheme using RSS [118]. Channel State Information (CSI) is regarded as fine-grained representation of the WiFi channel and can achieve more accurate localization performance [119]. However, a well-trained neural network is usually sensitive to changes in the environment, the network parameters need to be updated once the testing environment is changed. Compared with these antenna array based systems, our sparse tag array can achieve high resolution of angle estimation as well as having a low cost.

The RFID technology has been regarded as an effective and low-cost solution for many emerging IoT applications [25, 42, 54, 58, 157]. Although RFID-based systems are limited by the short communication range, the multipath effect on RFID systems is usually much smaller than that on WiFi systems. Thus, various RFID based localization schemes have been proposed to achieve higher accuracy and convenient deployment than WiFi-based systems.

Existing works on RFID tag localization can be classified into received signal strength Indicator (RSSI)-based and phase-based methods. These works mainly focus on locating a single tag, i.e., one tag is located at a time. For RSSI-based methods, a large number of reference tags are deployed at known locations. By comparing the RSSI data with reference tags, the position of the target tag can be determined [31]. In fact, RSSI values are raw channel information and are not stable, due to the factors such as multipath propagation, tag's orientation, RFID reader's transmit power, etc. RSSI based methods usually do not achieve high accuracy in indoor localization. On the other hand, phase based methods have been developed for estimating distance and direction of arrival (DOA) [33]. However, the measured phase is periodic, which leads to phase ambiguity and makes it less useful. Moreover, considerable measured phase errors are introduced by the reader antennas and the tag itself.

To address these issues, the synthetic aperture radar (SAR) technique is proposed for DOA estimation by moving the reader antenna around [21]. The second solution is the hologram

technique, which computes the probability of each known position as the tag source within an area of interest and then chooses the most likely position as the tag location [22, 34]. Another solution is the hyperbolic-based method for distance estimation, which locates a static tag [35]. However, this solution does not achieve high localization accuracy due to limited number of reader antennas. In addition, the RFind system achieves higher localization accuracy using a large virtual bandwidth to estimate time-of-flight, but it requires a special hardware [91]. The features of different RFID tag localization techniques are further summarized in Table5.2.

1.2.3 Pose Tracking Systems With RF Devices

The RFID based human pose tracking systems are closely related to prior works on RFID based sensing [10] and human pose estimation [186]. We mainly focus on these two classes of systems in the following.

Recently, passive RFID tags have attracted great interest because of their easy deployment and low-cost features [179]. The Low Level Reader Protocol used by the Reader can provide useful low-level information such as received signal strength indicator (RSSI), phase, Doppler frequency shift, timestamp, etc. [38]. As a result, many RFID-based sensing techniques have been developed for many applications, such as indoor localization [21, 22, 34, 57, 91], vital sign monitoring [8, 27, 58, 88, 89, 138, 139], user authentication [68], material identification [69], object orientation estimation [23], vibration sensing [70], anomaly detection [71], temperature sensing [52], and drone localization and navigation [53, 54, 90]. Particularly, the RF-wear system [135] and RF-Kinect system [134] utilize RFID tags attached to the human joints to estimate the movement of a particular limb, such as front arms, front legs, and thighs [134, 135]. We adopt the same approach in RFID-Pose. However, these systems may not be suitable for realtime human pose estimation, especially when multiple moving joints need to be tracked simultaneously. These RFID based sensing systems inspire us to develop an RFID based pose estimation system.

Prior works on human pose estimation are mainly based on computer vision techniques [186, 187]. For human pose estimation using video data, deep learning based method has been shown effective for 2D human pose with conventional RGB cameras [173, 188], and 3D human pose

with RGB-Depth cameras [189] and VICON systems [190]. These camera-based techniques can achieve high accuracy, but all require sufficient lighting condition and may raise privacy concerns.

These limitations motivate the development of RF based pose estimation techniques, because detecting RF signals do not require any lighting [191]. Moreover, since no video is used in the RF systems, the privacy issues are effectively addressed. However, collecting labeled pose data from RF signals is very challenging. Therefore, several RF based techniques leverage vision data as labeled pose data to train the deep learning network. This approach is also taken in the proposed RFID-Pose system. For example, RFPose is the first work to use RF signals with an FMCW radar for 2D human pose estimation, where a teacher-student deep learning model is utilized [176]. RFPose3D is the later version for 3D human pose estimation with FMCW radar [191]. Moreover, mmwave Radar is also utilized for human pose estimation with deep learning [177]. Recently, WiFi CSI has been exploited to create 2D skeletions [178] and 3D human poses [131] using cross-modal deep learning techniques. However, Radar and WiFi based human pose estimation are easily influenced by the environment noise and interference, and the FMCW radar technique is limited by the relatively higher cost (e.g., implemented with Universal Software Radio Peripherals (USRP)).

The proposed RFID-Pose system, to the best of our knowledge, is the first to apply RFID based sensing for 3D human pose estimation. The proposed system consists of a novel and effective solutions for cross-modal 3D human pose estimation using RFID and computer vision, which is much more robust compared with WiFi and Radar based methods.

A strength of the traditional camera, WiFi, and radar-based systems is that they are "markerless" methods, which are less intrusive. Video camera was first used to detect human poses in [186, 187]. With deep learning models, such systems localized the coordinates of human joints in the captured video frames, using, e.g., 2D RGB cameras [173, 188] or 3D depth cameras [189]. The most accurate 3D pose tracking performance was achieved, so far, by the Vicon system [190], which has been widely used for production of 3D movies. However, such video based schemes usually raise privacy concerns, as discussed, and their performance is usually limited by poor illumination, cluttered background, or poor camera angles.

To address the privacy concerns and mitigate the dependency on lighting and background, several RF pose tracking techniques have been proposed. Since such systems record no vision data and the RF data is not visible, user privacy can be better preserved. Furthermore, RF sensing systems perform well in poorly lighted environments and are able to detect human poses through obstacles and walls [177, 191]. FMCW Radar was first utilized to construct both 2D and 3D human poses by incorporating a vision-aided teacher-student deep learning model [176, 191]. As another type of non-intrusive sensor, WiFi channel state information (CSI) has also been analyzed to extract 2D and 3D human poses [131, 178]. Most existing RF sensing systems incorporate a deep learning model with vision data supervised training. Furthermore, due to the relatively wide transmission range of the radio signals, such systems are susceptible to interference from the operating environment. Usually radar-based techniques are more resistant to environmental interference than WiFi-based schemes, but their customized hardware, e.g., the FMCW radar implemented on the Universal Software Radio Peripherals (USRP) platform, usually incurs a higher cost.

RFID tags can serve as low-cost and light-weight wearable sensors to attach to the human body, which provides a promising solution for human pose estimation. Several RFID sensing techniques have been developed in recent years, such as human vital sign monitoring [27, 88, 89, 139], mechanical vibration sensing [70], user authentication [68], material identification [69], and temperature sensing [52], Furthermore, RFID has also been utilized for indoor localization [21, 34, 57, 91] and drone navigation [53, 54, 90].

Using RFID tags as wearable sensors, such systems are usually more robust to interference from the operating environment than other RF sensing techniques (e.g., WiFi). This feature inspires the development of several RFID based human pose tracking systems as well. For example, RF-Wear [135] and RF-Kinect [134] were developed to track the movements of a single human limb, while RFID-Pose [157] and Cycle-Pose [180] were developed to track 3D human poses in realtime. However, although the near-field RFID communications are more resilient to environmental interference, the locations of the tags and antennas still have a big impact on how the tags are sampled by the reader, and thus on the performance of the human pose tracking system.

In [192], the authors presented a domain adversarial technique to adapt to changes in the environment by utilizing a domain discriminator, which can constrain the unnecessary feature extraction from different environments. However, the proposed learning model may not be able to obtain the optimal variables when applied in a new RF environment, because all the training variables are determined by the datasets from a limited number of environments.

Inspired by the existing human pose tracking systems, we propose the Meta-Pose system, which is based on the meta-learning framework for greatly enhanced environmental adaptability. The proposed system incorporates a novel initialization algorithm to pretrain the deep learning model using a limited amount of training data, so that the system can be quickly fine-tuned with a small amount of new data when applied to a new environment, while still achieving a satisfactory performance.

1.3 Summary of Contributions

To our best knowledge, the proposed AuTotag system is the first apnea detection systems incorporating an enhanced recurrent variational autoencoder model. The proposed scheme is an unsupervised learning, with the desirable advantage of not requiring costly labeled medical data. In the system, we also propose a novel technique to address the frequency hopping offset, which is a real-time calibration. The proposed scheme is simple but effective in mitigating the frequency hoping offset, thus enabling many real-time sensing applications for FCC-compliant RFID readers and tags.

The Nodtrack system is the first work that leverages passive RFID tags for driving drowsiness detection under real driving settings. A specific tag deployment and several signal processing algorithms are proposed to effectively distinguish the nodding features from the strong environment noises and other types of driving related movements. Driving fatigue is detected by an unsupervised LSTM autoencoder model, which does not require labeled training data of various types of driving movements, which are hard and costly to obtain. An effective algorithm is proposed to estimate, on real-time, the phase difference between two RIFD tags that are interrogated with slotted ALOHA and under frequency hopping in commercial RFID systems.

To the best of our knowledge, we also propose the first work on respiration monitoring in driving environments using commodity RFID reader and tags. A prototype system is built with commodity RFID devices, deployed in a car, and validated in both an emulated environment and real driving environments. The experiments are conducted in various driving scenarios, where excellent performance of the proposed system is demonstrated. We propose a tensor completion technique to recover missing readings in collected phase data, and a tensor decomposition approach to extract the respiration signal of the driver from phase values sampled from multiple RFID tags. The proposed techniques are effective in combating the strong noises caused by frequency hopping, random sampling, vehicle vibration, and other movements in the driving movement.

We design the SparseTag system, which includes sparse array processing, difference coarray design, DOA estimation using a spatial smoothing based method, and a localization method. We propose a new sparse tag array design and analytically prove its superior performance over the traditional ULA design. In addition, a robust channel selection method based on the sparse tag array is proposed for mitigating the indoor multipath effect. We justify the feasibility and advantages of utilizing a sparse tag array for DOA based indoor localization through analysis and experiments. To the best of our knowledge, this is the first work to leverage sparse tag arrays for backscatter indoor localization, which does not require to move either the tags or the antenna(s).

The proposed RFID-Pose system is the first work for 3D human pose estimation using commodity RFID reader and tags, which can effectively monitor multiple human joints simultaneously in realtime. We propose a novel data preprocessing approach to mitigate the severe RFID phase distortion and compensate the large amount of missing data in sampled raw RFID data. The tensor completion technique is utilized for data imputation, so that phase data for all RFID tags can be estimated. In the system we propose a vision-aided solution for training the proposed deep kinematic neural network, to transform sensed RFID phase variations to the spatial rotation of each limb. The proposed approach effectively addresses the challenges of the low data rate in RFID systems, because rotation angle estimation requires much less data than generating a joint confidence map.

With the foundation of the RFID-pose system, we develop the Cycle-Pose system to improve the subject adaptability of the system. To the best of our knowledge, this is the first subject-adaptive 3D human pose estimation system using commodity RFID reader and tags, which can effectively track 3D human pose without vision data in the testing stage. We propose a cycle kinematic network model and train the network with self-supervision. The proposed model learns the transformation from RFID data to 3D skeleton for different subjects, to effectively achieve subject adaptability. Besides, we propose a novel Meta-Pose initialization algorithm based on meta-learning algorithms (i.e., model-agnostic meta-learning (MAML) and Reptile) to pretrain the deep learning model with a limited number of training datasets sampled from several known environments. We develop the initialization approaches based on both Reptile and MAML.

Last but not the least, we develop TARF system, which is the first technology-agnostic human activity identification system capable of performing generalized and accurate HAR using various RF sensing platforms. We investigate the challenges in technology-agnostic HAR and show that they are caused by three main factors: metric disparities, heterogeneous sensitivity distributions, and diverse motion feature translations. A universal RF data preprocessing module is proposed to reduce the disparity between different RF sensing technologies. The sensitivity diversity is addressed by remapping the signal strength measurements, and generalized tensor data is constructed using STFT. The DANN is utilized to categorize different types of human activities, which further mitigates the interference from diverse RF domains.

Chapter 2

AutoTag: Unsupervised Detection of Apnea using Commodity RFID Tags with a Recurrent Variational Autoencoder

2.1 Introduction

The population is aging in many parts of the worlds. Consequently, smart healthcare has attracted increasing concerns [8–11]. Rather than going to hospital after getting sick, people are intend to early detect and prevent diseases by monitoring their vital sings on daily basis. For example, One of the breathing disorders is obstructive sleep apnea, which can imply serious health problems in human body including high blood pressure, heart disease, and sudden infant death syndrome (SIDS) for sleeping infants [12]. However, in traditional healthcare systems, vital signs are measured by dedicated equipment like capnography [13], which is not convenient for all-day monitoring, especially when the patient is sleeping. Moreover, the breathing abnormality diagnosis may consume considerable efforts and experience from medical institution. Therefore, autonomous, low-cost, unobtrusive vital sign monitoring methods that can detect abnormality are desired for IoT based smart healthcare systems, which can benefit many people for monitoring health conditions in their daily life.

Considering the mobility and flexibility of RF devices, wireless signals are widely used on smart healthcare systems to monitor human vital signs. Since the movement of human chest and heart can slightly affect the propagation of RF signals, the signal of breathing and heartbeat can be reconstructed by analyzing the change in received RF signals. Based on this basic idea, multiple existing techniques incorporate a radar for respiration monitoring, including frequency modulated continuous wave (FMCW) radar [14] and Doppler radar [15], but at a relatively high cost due to the special hardware. To achieve low cost RF system, WiFi based techniques are

developed for heath sensing with commodity WiFi devices. Rather than directly analyzing received signals from radar, WiFi based techniques leverage either Received Signal Strength (RSS) [16] or Channel State Information (CSI) [17–20] collected from the device driver. Although the WiFi techniques are low-cost and flexible, the systems are sensitive to the noise caused by surrounding environment, such as moving objects or persons nearby. The accuracy of such systems is relatively low in unstable environments.

The low-cost and near-field features of passive RFID tags have triggered great interest on apply them for health sensing. Some RFID based systems are proposed to achieve low cost as well as reducing the influence of unstable surroundings. Multiple RFID based techniques have been developed for object tracking [22], orientation estimation [23], drones [24–26], and especially, for respiration monitoring [27]. Such existing works mainly make use of the RFID phase information collected from the RFID reader on different channels. One such typical techniques for smart healthcare is called TagBreathe, which monitors the respiration signal of a patient by grouping the RFID responses collected from the same channel and using a estimated displacement in each channel [27]. This method may not be well suited for operation with Ultra High Frequency (UHF) RFID devices in the US, which all adopt frequency hopping over 50 channels, 200 ms per channel, according to FCC requirements. When the reader and tag hop among 50 channels, it will take 10 seconds for them to return to the same channel. To collect a sufficient amount of readings from the same channel, the delay will be considerable, making it hard for real-time detection of abnormality (e.g., apnea).

To address this issue, we present the AutoTag system, an unsupervised recurrent variational autoencoder method for respiration rate estimation and abnormal breathing detection with off-the-shelf RFID Tags. To mitigate the effect caused by channel hopping, we propose a novel technique to map the RFID phase values collected from multiple different channels to a single reference channel. Since FCC requires the RFID system to hop to a different channel every 200 ms, a typical RFID based sensing system can hardly be applied for real-time monitoring of patients' vital signs. Rather than offline calibration employed in Tagyro [22], our method can enable real-time phase calibration, which is amenable to dynamic environments.

Furthermore, compared with the method used in TagBreathe, our method incurs much lower delay, because grouping data for all channels is not needed with our technique.

Furthermore, we develop an unsupervised deep learning approach for apnea detection, which can autonomously detect abnormality in human respiration. Recently, a recurrent variational autoencoder model has been successfully applied to sequence modeling [28] and human motion synthesis [29]. Inspired by these works, we develop an enhanced recurrent variational autoencoder model for detection of breathing abnormality, such as apnea. With the proposed approach, abnormality can be detected by evaluating how similar the sampled breathing signal and reconstructed signal using the deep learning network are. Our method is superior to the traditional energy-threshold based approach, since the testing environment may not be absolutely stationary. Our proposed method can easily distinguish non-periodic signals from normal periodic signal by learning the features of normal breathing signals, while the energy based method only consider apnea as a relative weak signal compared with normal cases. Since the proposed scheme is an unsupervised learning, it has the desirable advantage of not requiring labeled medical data, which is usually costly and time-consuming to obtain.

Specifically, we present the AutoTag system, a recurrent variational **Auto**encoder for respiration rate estimation and unsupervised detection of apnea with commercial passive UHF RFID **Tag**s. The AutoTag system composes of three main components, including (i) the signal extraction module, (ii) the calibration module, and (iii) the breathing monitoring module. The phase data is firstly collected from a commodity RFID reader by the signal extraction module. The calibration module is mainly used for calibration of the sampled breathing data, while the respiration monitoring module is designed for estimating the patient's respiration rate and detecting abnormalities such as apnea. We prototype the AutoTag system using a platform of commercial RFID tags and reader, and conduct extensive experiments in two different environments with four volunteers. We observe superior performance achieved by the proposed AutoTag system in these experiments. The impact of various design and environment factors are also tested in corresponding experiments.

We summarize the three main contributions of this work as follows.

- To our best knowledge, the AuTotag system is the first apnea detection systems incorporating an enhanced recurrent variational autoencoder model. The proposed scheme is an unsupervised learning, with the desirable advantage of not requiring costly labeled medical data.

- We also propose a novel technique to address the frequency hopping offset, which is a real-time calibration. The proposed scheme is simple but effective in mitigating the frequency hoping offset, thus enabling many real-time sensing applications for FCC-compliant RFID readers and tags.

- We design and prototype the AutoTag system, which is composed with (i) signal extraction, (ii) data calibration, and (iii) respiration monitoring modules, and evaluate the system in two different representative healthcare environments. We present our experimental results that validate the efficacy of the proposed AutoTag system.

The mainder of this paper is organized as follows. The preliminaries of RFID based sensing is discussed in Section 2.3. The AutoTag system design is presented in detail and analyzed in Section 2.4. The experimental performance evaluation of the proposed system is provided in Section 2.5. After reviewing related work in Section 2.2, we conclude this paper in Section 2.6.

2.2 Related Research

The AutoTag system is mostly related to RF signal based vital sign monitoring systems and RFID based sensing. We briefly introduce these two classes of related work in this section.

Nowadays, many RF based systems have been proposed to monitor vital signals of human, which are developed on different types of platforms, including Radar systems, WiFi systems, and RFID based systems. The Radar technique is a straightforward way to identify the fluctuation of human chest caused by respiration, because Radar can directly monitor the distance variation between the human chest and the device antenna. One of the representative works in this category leverages an FMCW radar to monitor respiration rate and heart rate for multiple users simultaneously [14]. Furthermore, other Radar based systems have also been proposed to measure human respiration, including Doppler radar [15] and ultra wideband (UWB) Radar [30].

These radar based systems can accurately detect vital signs, and the influence caused by surroundings is limited. However, since the special hardware is essential to such systems, the cost of such systems are usually high, hampering their wide deployment such as in homes.

To achieve low cost and ease of deployment for vital sign monitoring, WiFi based techniques have been utilized to measure both human breathing rate and heart rate. The human respiration and heartbeat can be extracted by analyzing variations in WiFi channels, for example, the RSS as in Ubibreathe [16]. However, the Ubibreathe system requires the patient to stand between the transmitter and the receiver, while some other CSI based techniques have no such strict requirements. Different from RSS based systems, CSI can provide fine-grained channel inforamtion, and can achieve higher resolution and sensitivity than RSS for monitoring human vital signs. One of the CSI based techniques can leverage amplitude of CSI to monitor breathing rate and heart rate when the patient is sleeping [17]. In addition to the amplitude, the CSI phase information can also carry human respiration signal [18]. Furthermore, the TensorBeat system can estimate respiration rates for multiple persons crowed in a small space [19], by incorporating tensor decomposition on the collected CSI phase data. Although WiFi based techniques can measure human vital signs with off-the-shelf WiFi devices, the accuracy is easily affected by the surrounding environment, because of broadcasting nature and long range of WiFi transmissions. To address this issue, some RFID based systems like TagBreathe are developed to track human respiration by analyzing the RFID response data collected at an RFID reader [27]. Since the passive UHF RFID tags are low-cost and are easily attachable to human body, the RFID system can monitor human vital signs at a low cost and is resilient to interference from the unstable environment.

Apart from RFID based vital sign monitoring systems, various sensing systems are proposed by leveraging the data extracted from the low level protocol in RFID systems, such as RSS and phase values, which have been utilized for many applications, e.g., indoor localization. For RSS based methods, one of the representative techniques estimates the tag position by comparing the RSS from the target tag with reference tags [31]. Moreover, another technique can obtain the refined tag position by utilizing the characteristics of the coupling effect on RSS [32]. However, due to the low resolution of RSS, developing an RSS based localization

scheme with high accuracy is challenging. Thus, many phase based localization techniques are proposed. One of the typical phase based methods has been developed for estimating distance with direction of arrival (DOA) [33], but the result has a relatively large ambiguity because of the periodicity of measured phase. To remove the ambiguity, some other techniques are proposed to obtain the more accurate position than the typical method by leveraging the aperture radar technique [21] and hologram technique [22, 34] [35]. Besides indoor localization, RFID tags are also widely used for other sensing techniques such as remote control of drones [24–26], object orientation estimation [23], remote temperature measurement [36], and gesture recognition [37].

2.3 Preliminaries of RFID Sensing

According to FCC regulations, UHF RFID readers should use channel hopping to avoid co-channel interference. The spectrum from 902.5 MHz to 927.5 MHz is partitioned into 50 non-overlapping channels, and the reader remains on each channel for 200 ms. Usually such frequency hopping introduces an additional phase offset in the RFID response signal, causing large errors in RFID based sensing.

According to the manual of RFID reader, e.g., [38], the phase ϕ of the received RFID response signal can be expressed as

$$\phi(f_i, d) = \text{mod}\left(\frac{2\pi f_i d}{c} + \delta_T + \delta_R + \delta_{Tag}, 2\pi\right), \tag{2.1}$$

where d is the total distance from the reader's antenna to the tag and then back to the reader antenna, f_i is the frequency of channel i, c is a constant representing the speed of light, and δ_T, δ_R, and δ_{Tag} are the phase offsets caused by the transmitter circuit, the receiver circuit, and the tag's reflection characteristics, respectively.

For Impinj R420, a commodity RFID reader, the phase offset between two adjacent channels that it hops to is not a constant, even though the distance d rmains the same, as found in our experimental studies. Since the three offsets in (2.1) are irrelevant to the distance d, we can lump the three offsets into a single variable δ_i for each channel i. The phase $\phi(f_i, d)$ for

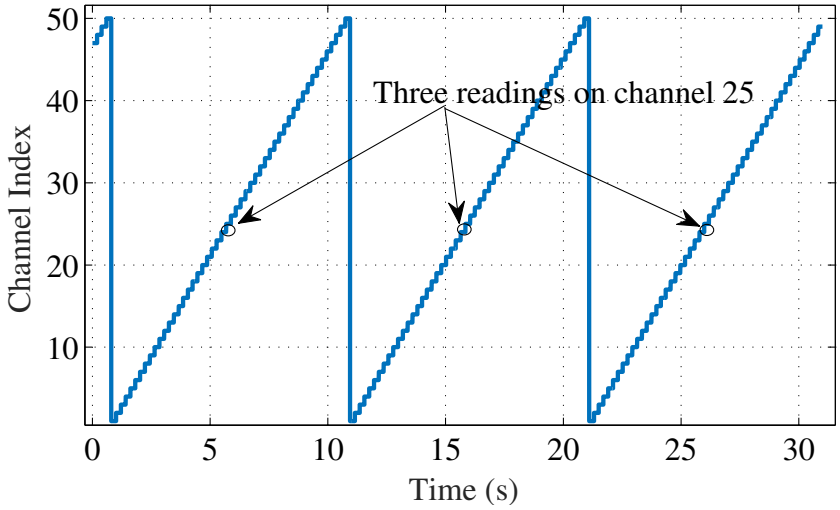


Figure 2.1: The channel indexes used by an FCC-compliant RFID reader during a period of 30 seconds.

channel f_i under round-trip distance d can thus be expressed as

$$\phi(f_i, d) = \text{mod}\left(\frac{2\pi f_i d}{c} + \delta_i, 2\pi\right). \tag{2.2}$$

The main challenge for extracting the breathing signal from the RFID phase measurements is how to mitigate the discontinuity in phase data, which is caused by channel hopping. One way to eliminate the channel hopping influence, as proposed for the TagBreathe system [27], is to group the signals collected from the same channel and to use the estimated displacement in each channel to track the breathing signal. As discussed earlier, this method may not work well for RFID systems in the US, since the reader must hop among 50 different frequencies, following the FCC requirement. Fig. 2.1 plots the change of channel index in a period of 30 seconds. We can see that it takes about 10 seconds for the antenna to hop through all the 50 channels. Thus, the TagBreathe method will take a very long time to collect and group multiple phase readings on the same channel, leading to extremely long delay in respiration measurement with FCC-compliant readers.

To address the extremely long delay caused by channel hopping among 50 different frequencies, the Tagyro system calibrates phase values collected from all channels based on one reference channel [23]. Specifically, the Tagyro technique first measures the phase offset δ_i for all the 50 channels. Then, the phase offset introduced by channel hopping can be removed by

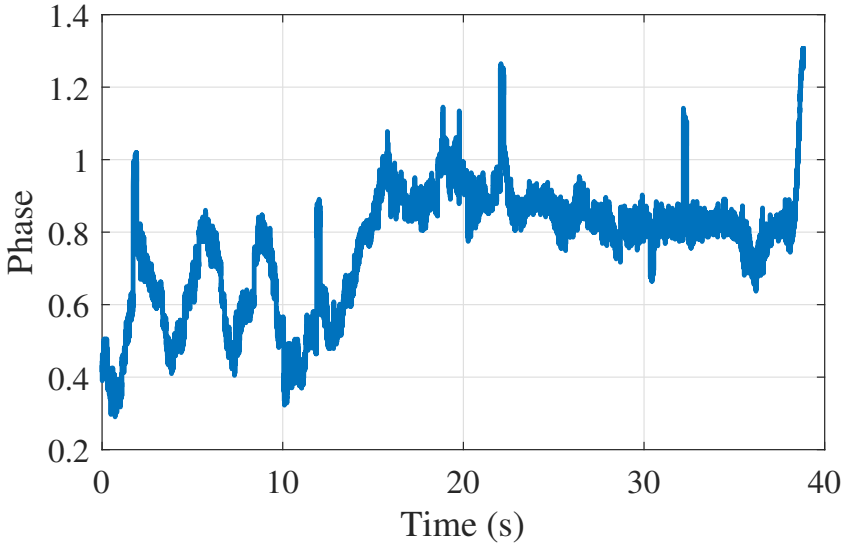


Figure 2.2: Calibrated phase obtained using the Tagyro method [23].

subtracting the phase offset δ_i in each channel. This method is suitable for a static environment; but it may not be effective for tracking human breathing signal and apnea, where the tags are mounted on the human body and moves as the patient breaths. This is because the wireless channel will change if the patient moves (even slightly). The movement causes an additional offset in δ_i, so that the estimated phase offset $\hat{\delta}_i$ does not match the real-time δ_i after the small movement.

Fig. 2.2 plots the calibrated phase data obtained with the Tagyro method. It can be observed that the breathing signal can be detected in the beginning (i.e., the first 15 seconds), because the initial phase offset is correct. After the first 15 seconds, the breathing signal cannot be detected, because the channel hopping effect cannot be perfectly mitigated. To continuously eliminate the frequency hopping effect, we propose a new method in the proposed AutoTag system, to update and remove the phase offset δ_i in real-time for breathing and apnea detection.

2.4 Design and Analysis of the AutoTag System

2.4.1 Design of the AutoTag System

The AutoTag system aims to measure human respiration and detect breathing abnormalities, such as apnea, with multiple RFID tags attached to the patient's body (i.e., clothes). As given in (2.1), the collected phase information is indicative of the round-trip distance d between the

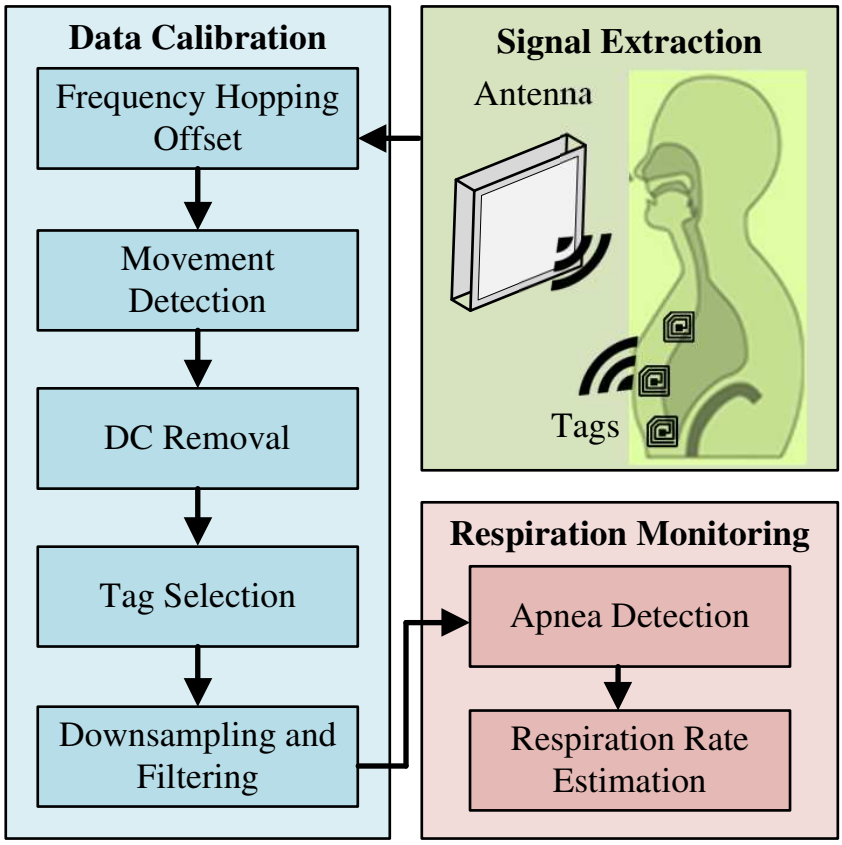


Figure 2.3: The AutoTag system architecture, which includes signal extraction, data calibration, and respiration monitoring.

reader antenna and the corresponding tag. When the patient breaths, the distance d changes slightly with the chest movements. Thus, by analyzing the phase variations collected from the tags attached to the human body, we can obtain the periodic signals caused by chest movements. However, for accurately measuring the human respiration rate and precisely detecting apnea, several challenges should be addressed, such as mitigating the channel hopping effect, the sensitivity divergence for different tags, and dealing with the interference from surroundings. To address these issues, we incorporate three modules in the AutoTag system, including (i) signal extraction, (ii) data calibration, and (iii) respiration monitoring, as illustrated in Fig. 2.3.

In the *signal extraction* module, we extract the phase data from the received responses from three tags attached to the human body, using a directional antenna at the reader. In the following *data calibration* module, we firstly remove the influence caused by channel hopping of the RFID reader. Then we detect whether the monitored patient is moving or not based on a threshold based method, i.e., movement detection. After that, we remove the DC component

from the selected signal to eliminate the impact of small movements of the patient. Then, tag selection is implemented to choose the tag with the strongest signal strength. Finally, we apply downsampling and filtering to obtain the final respiration signal. In the *respiration monitoring* module, we adopt a recurrent variational autoencoder for detection of abnormalities such as apnea. Our approach is an unsupervised learning, which has the great advantage of not requiring labeled medical data, which is extremely costly to collect. The respiration rate can also be estimated with a peak detection method when the patient is breathing normally. We will introduce the detailed design of each module in the remainder of this section.

2.4.2 Signal Extraction

As shown in Fig. 2.3, the first module is used for extracting low-level phase readings from received tag responses. The movements of the patient's chest and abdomen induced by breathing, cause the tag-reader antenna distance to vary with human respiration. The time-varying distance translates to the time-varying phase in the tag response signal, which is indicative of the respiration signal. To increase the sysetm's robustness, three passive UHF RFID tags are attached to the upper body of the patient. The RFID reader uses a directional antenna to transmit RF interrogating signals to the tags and to read low-level data from the backscattered signals from the tags, which includes time stamp, phase, received signal strength indicator (RSSI), and Doppler shift.

For most RFID systems, the collected RSSI data usually has a very low resolution, and the signal-to-noise ratio (SNR) of Doppler shift is usually low. Thus these two types of information are not very helpful for detecting the respiration signal. In AutoTag, we focus on the collected phase information from RFID responses for respiration rate estimation and apnea detection.

2.4.3 Data Calibration

The captured phase information cannot be directly used for detecting the respiration signal. In Fig. 2.4, we plot the uncalibrated phase data received from one of the reader antennas for a duration of 28 s. It can be observed that when the reader hops around various channels (200 ms on each channel), the measured phase data exhibits a wide range of variations. Furthermore,

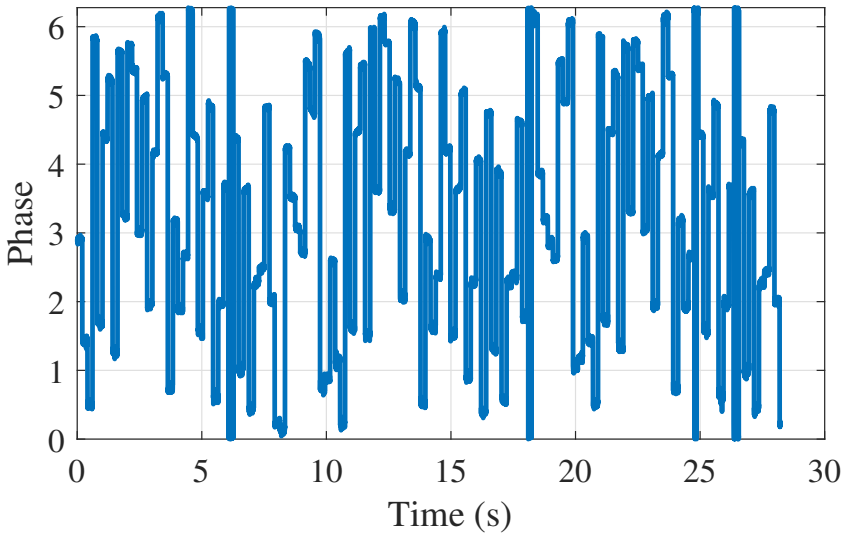


Figure 2.4: Uncalibrated phase data collected from a tag for a duration of 28 s.

there is a large offset incurred in the phase data when the frequency hopping happens. It is thus highly challenging to extract the weak respiration signal from such uncalibrated data. The raw phase data should be calibrated first to facility the extraction of the respiration signal.

Mitigating the Frequency Hopping Offset

We unwrap the captured phase signal to remove the offset introduced by the modulo operation in (2.2). With the modulo operation, a slight change in the real phase may lead to a large jump in the received phase signal. For instance, a small change of the real phase from 0.1π to -0.1π will cause the received phase to change from 0.1π to 1.9π. Since the sampling rate of the reader is usually higher than 100 Hz for each tag in our system, the interarrival time of phase samples is usually smaller than 0.01 s. Assuming that the phase value change of two back-to-back readings is smaller than π under such a small interarrival time, $\pm 2\pi$ can be added to recover the original phase value when the change exceeds π. Note that frequency hopping can also cause big variations between two back-to-back phase samples, and unwrapping will only be used for consecutive phase samples collected from the same channel. After the unwrapping operation, the phase samples from each channel is smoothed.

The second step is to splice the smoothed phase signals from all the 50 channels into a single phase signal, by mitigating the frequency hopping offsets. The key is to translate the

phase signal from the next channel that the reader hops to, to a transformed phase signal using the previous channel as a reference with real-time calibration. As illustrated in Fig. 2.5, the phase signal from the previous channel, e.g., channel i, can be written as

$$\phi(f_i, d) = \text{mod}\left(\frac{2\pi f_i d}{c} + \delta_i, 2\pi\right). \tag{2.3}$$

Then the reader hops from channel i to channel $(i+1)$. Similarly, for the next channel that the reader hops to, we have

$$\phi(f_{i+1}, d) = \text{mod}\left(\frac{2\pi f_{i+1} d}{c} + \delta_{i+1}, 2\pi\right). \tag{2.4}$$

For simplifying notation, ignore the modulo operation for now. We then multiply the channel $(i+1)$ phase signal by a factor of (f_i/f_{i+1}), to have

$$\begin{aligned}\hat{\phi}(f_{i+1}, d) &= \phi(f_{i+1}, d) \times \frac{f_i}{f_{i+1}} \\ &= \frac{2\pi f_i d}{c} + \delta_{i+1} \times \frac{f_i}{f_{i+1}} \\ &\doteq \frac{2\pi f_i d}{c} + \delta_i + \Delta\delta_i, i = 1, 2, ..., 49.\end{aligned} \tag{2.5}$$

Therefore, we have

$$\hat{\phi}(f_{i+1}, d) = \phi(f_i, d) + \Delta\delta_i,\ i = 1, 2, ..., 49, \tag{2.6}$$

where

$$\Delta\delta_i \doteq \delta_{i+1} \times \frac{f_i}{f_{i+1}} - \delta_i, \tag{2.7}$$

represents the transformed phase offset, as illustrated in Fig. 2.5, which can be easily estimated as follows.

Note that we already know the frequency for each channel. So we first multiply the phase $\phi(f_{i+1}, d)$ collected from channel $(i+1)$ with a ratio f_i/f_{i+1}. The next value we need to

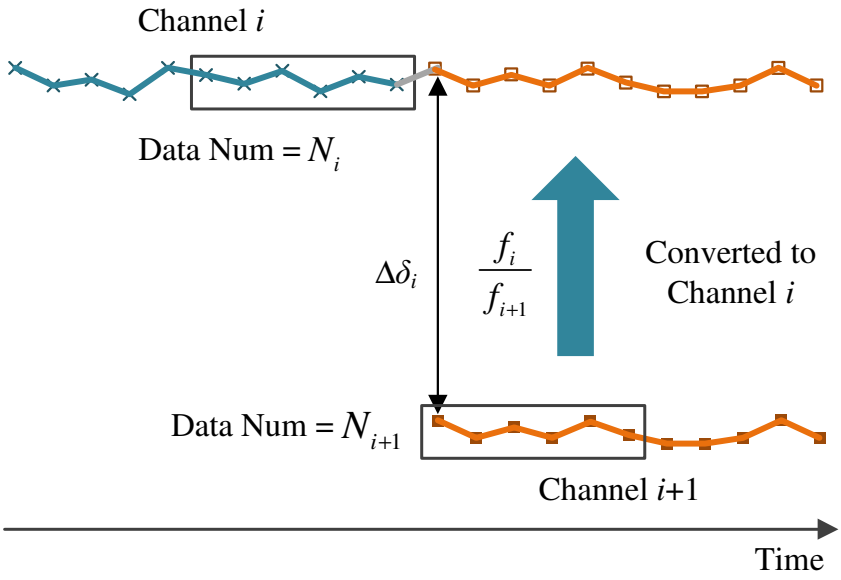


Figure 2.5: Illustration of the proposed frequency hopping offset mitigation scheme.

calibrate is $\Delta\delta_i$ as in (2.7). In AutoTag, only three tags are attached to the human body. The overall sampling frequency of the reader is 600 Hz, and the sampling rate for each tag is more than 100 Hz. It takes only less than 1 ms for the reader to hop from one channel to another. Due to the high sampling rate and the small channel hopping time, it is reasonable to assume that the antenna-to-tag distance and the surrounding environment remain the same during channel hopping. Under this assumption, the difference between the phase data before channel hopping and the transformed phase data after channel hoping, is all caused by $\Delta\delta_i$ along with thermal noise.

To mitigate the influence of thermal noise, we apply a Hampel filter to the signal read from each channel. We choose a sliding window of 20 samples and a threshold of 0.01 for the Hampel filter. Next, we compute the difference between (i) the average of the last 6 phase readings in the previous channel i (denoted as $\phi(f_i, d, k)$), and (ii) the average of the first 6 transformed phase readings in the present channel $(i+1)$ (denoted as $\hat{\phi}(f_{i+1}, d, k)$), as an estimate for the transformed phase offset $\Delta\delta_i$. As shown in Fig. 2.5, and also from (2.6), we have

$$\hat{\Delta}\delta_i \approx \frac{1}{6}\left(\sum_{k=N_i-5}^{N_i}\phi(f_i, d, k) - \sum_{k=0}^{5}\hat{\phi}(f_{i+1}, d, k)\right), \tag{2.8}$$

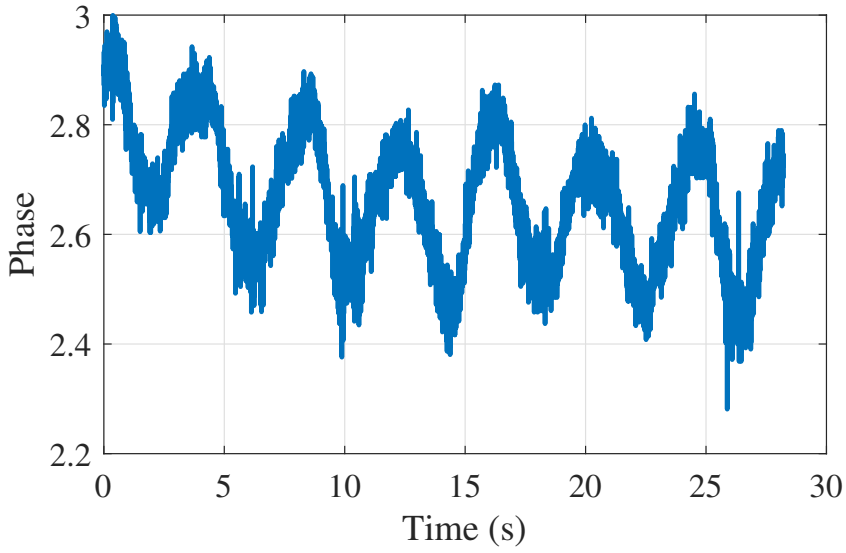


Figure 2.6: The resulting phase data after removing the frequency hopping offset.

where N_i is the total number of phase readings collected from channel i.

After compensating for the frequency hopping offset, the phase samples on the present channel $(i+1)$ can be approximated as in (2.6). Fig. 2.6 plots the calibrated phase samples after removing the frequency hopping offset. It can be seen that after calibration, the collected phase data shows an visible periodic respiration signal, which is completely missing in the uncalibrated phase data in Fig. 2.4 (for the same period of 28 seconds).

Movement Detection

After mitigating the frequency hopping offset, the next step is to detect whether the patient is moving. Note that the breathing signal is very weak comparing to other types of body movement. Therefore, we should only use the signals collected while patient if in a stationary state, to avoid the interference introduced by large movements of human body.

Movement detection is accomplished with a threshold based method. In particular, a sliding window is applied to the collected phase data, as showed in Fig. 2.7. For each window, we calculate the total mean absolute deviation of the phase samples from all tags, denoted by T, as

$$T = \frac{1}{|W|} \sum_{j=1}^{N} \sum_{k \in W} |\phi^j(k) - \mathbb{E}(\phi^j(k))|, \tag{2.9}$$

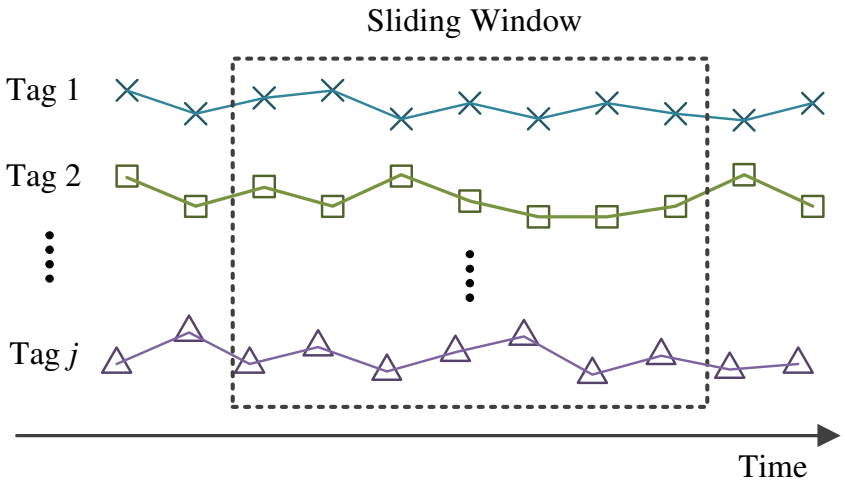


Figure 2.7: A sliding window on phase readings from the tags.

where W represents the set of all the phase readings in the sliding window, $|\cdot|$ is the cardinality, N is the number of tags, and $\phi^j(k)$ is the kth phase sample obtained from tag j. If the patient is not stationary, the phase values will exhibit big changes (due to the large changes in d caused by body movements). Thus, by setting a threshold on T, we can detect considerable movements of the patient. In AutoTag, we use a threshold value of 0.9 and a window size larger than 6 seconds for movement detection, based on our extensive experiments with various body movements.

DC Removal

Although the frequency hopping offset can be effective mitigated using a reference channel, the initial offset on the reference channel is still a random value that introduces a random DC component in the phase data. To remove the random DC component, as well as interference from other movements from the environment, we pass the phase signal through a detrending operation. Specifically, another Hampel filter is used, whose window size is 2000 and threshold is 0.001, to obtain an estimate of the DC component (i.e., the trend). Finally, the calibrated phase signal is obtained by subtracting the trend from the filtered signal. The calibrated signal is ploted in Fig. 2.8 for the same period of 28 s. It can be seen that the calibrated signal is now centered at zero, like a periodic AC signal.

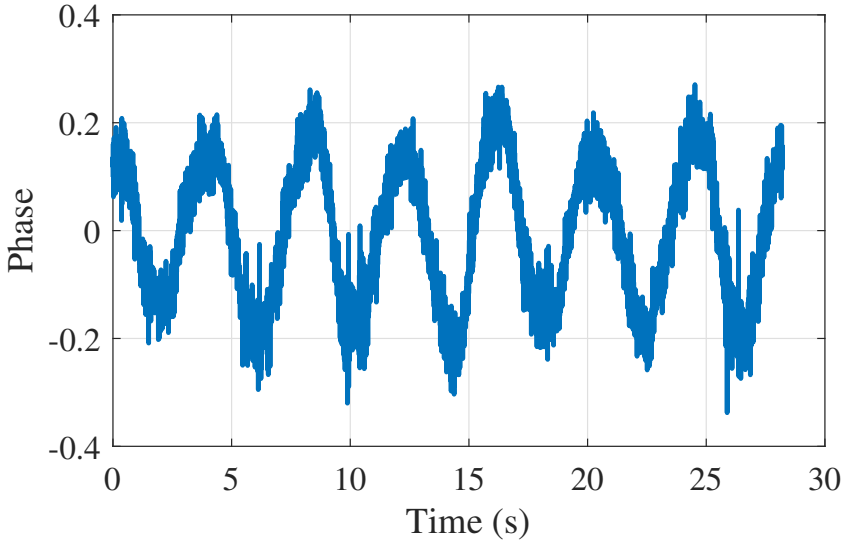


Figure 2.8: The calibrated phase signal after removing the DC component with a Hampel filter.

Tag Selection

When analyzing the experimental results, we find that each tag has a different sensitivity to the human breathing signal from other tags, which depends on the slightly different tag parameters and the different propagation environment (e.g., distance or angle) of each tag. In particular, the angle between the tag and the reader antenna is a factor that causes the difference in tag's sensitivity. To obtain the most sensitive signal, we measure the signal strength using the average absolute deviation within a certain window size of 6 seconds (see (2.9)). The tag with the largest signal strength will be chosen for the remaining processing steps.

Downsampling and Filtering

Due to the high sampling frequency, i.e., about 600 Hz with one reader antenna, the signal need to be downsampled to reduce the computational complexity. In AutoTag, the signal is downsampled with a factor of 10 before feeding to the respiration monitoring module. In addition, there are still some false peaks introduced by thermal noises, which affect the accuracy of peak detection for respiration rate estimation. Note that the normal human respiration rate is usually much lower than 0.5 Hz. Thus, we apply a low-pass filter and choose a cutoff frequency of 0.5 Hz to the downsampled signal, in order to remove the remaining high frequency noise. In Fig. 2.9, we plot the downsample and filtered phase signal, which is quite smooth now. It

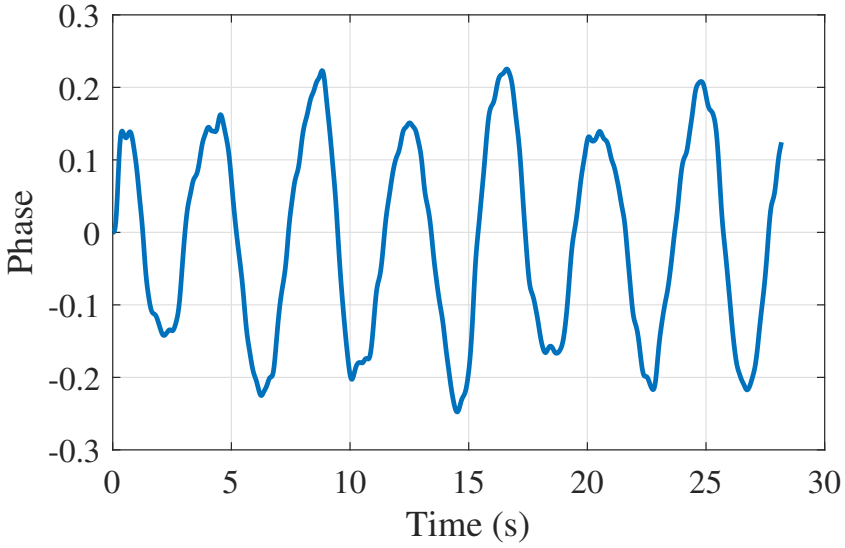


Figure 2.9: The finally recovered respiration signal.

is then used as input to the next module for the following apnea detection and respiration rate estimation tasks.

2.4.4 Apnea Detection and Respiration Rate Estimation

Recurrent Variational Autoencoder for Apnea Detection

For the purpose of accurate respiration detection, we introduce a recognition model as an approximation to the intractable true posterior, where the parameters are not computed from some closed-form expectation, but are learned from the calibrated data. To detect apnea, the main idea of AutoTag is to incorporate a *variational autoencoder* to compute the difference between the sampled signal and the reconstructed signal within a time window. Note that this is an unsupervised learning [39, 40], with the desirable advantage of not requiring labeled medical data that is hard or costly to obtain. If the computed difference is smaller than a given threshold, the sampled signal is regarded as a regular breathing signal, from which the respiration rate can be estimated; otherwise, the signal sequence is regarded to be abnormal, and apnea is detected. Since the energy based threshold has been applied for movement detection in the earlier stage (see Section 2.4.3), small breathing signals can be detected now at this stage.

The proposed recurrent variational autoencoder for unsupervised respiration abnormality detection is illustrated in Fig. 2.10. We first apply the variational autoencoder model to obtain

a reconstructed version of the input signal. This model is to maximize the marginal likelihood given below.

$$p_\theta(x) = \int p_\theta(x|z)p(z)dz, \tag{2.10}$$

where x, z, and θ are the observed variables, the latent random variables, and the set of parameters, respectively; $p(z)$ is the prior over the latent random variables z; and $p_\theta(x|z)$ is the conditional probability, representing an observation model under the parameter set. Usually $p_\theta(x)$ is intractable due to the integral operation. Although the Monte Carlo sampling method can be used to solve the problem, it usually incurs a considerable computational cost even for small-sized datasets. The variational autoencoder model utilizes the variational approximation $q_\phi(z|x)$ instead of the true posterior $p_\theta(z|x)$. Specifically, the variational autoencoder model has $q_\phi(z|x)$ with parameter set ϕ as encoder and $p_\theta(x|z)$ with parameter set θ as decoder. According to Jensen's inequality, the variational autoencoder model can achieve the optimal values for sets ϕ and θ. This is achieved by maximizing a lower bound on the log-likelihood, given by [39]

$$\max \mathcal{L} = -D_{KL}(q_\phi(z|x)||p(z)) + \mathbb{E}_{z\in q_\phi(z|x)}\left[p_\theta(x|z)\right], \tag{2.11}$$

where D_{KL} represents the KL divergence. In (2.11), $-D_{KL}(q_\phi(z|x)||p(z))$ represents the regularization over the latent variables z, while $\mathbb{E}_{z\in q_\phi(z|x)}\left[p_\theta(x|z)\right]$ is the autoencoder. The latent variables z are sampled from $q_\phi(z|x)$, and the reconstructed signal $\hat{x}$ can be sampled from $p_\theta(x|z)$.

To reduce the training overhead, the variational autoencoder model utilizes the reparametrization technique. With this technique, the latent vector z is computed from the mean vector $\mu_\phi(x)$ and the variance vector $\sigma^2_\phi(x)$, as

$$z = \mu_\phi(x) + \sigma_\phi(x) \odot \epsilon, \tag{2.12}$$

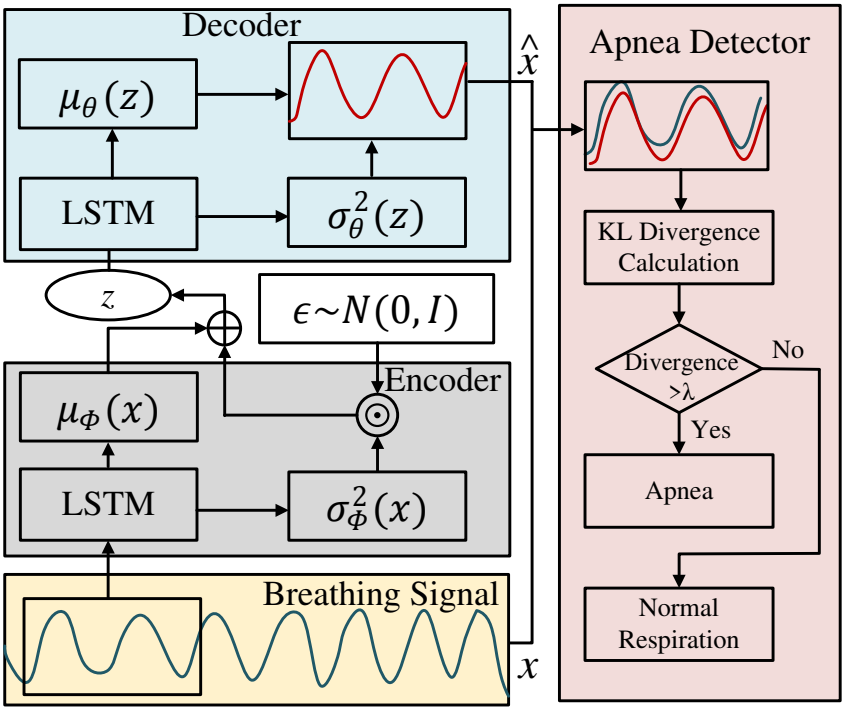


Figure 2.10: Architecture of the proposed recurrent variational autoencoder for unsupervised apnea detection.

where $\epsilon \in \mathcal{N}(0, 1)$ (i.e., a standard Gaussian distribution), and $\odot$ represents the element-wise product operation. The lower bound on the log-likelihood, i.e., $\mathcal{L}$, can then be approximated as follows.

$$\mathcal{L} \approx \frac{1}{2} \sum_{j=1}^{J} \left(1 + \log(\sigma_j^2(x)) - \mu_j^2(x) - \sigma_j^2(x)\right) + \frac{1}{M} \sum_{l=1}^{M} \log p_\theta(x|z_l),$$

where M is the number of samples in z, and J is the cardinality of z.

Next, we consider the data samples within a time window, to be processed by a long short-term memory (LSTM) network. LSTM belongs to the class of recurrent neural networks (RNN) that can effectively handle time series data. It can also deal with the problem of vanishing gradient problem exists in RNN. Besides, consider that the major difference between normal breathing signal and apnea signal can be reflected by both periodicity and the shape of signal, which can be considered as long range dependency and short range dependency, respectively. These long range dependencies can be effectually captured by LSTM, because the data in a time series can be stored or deleted by a non-linear gate in each unit. Thus, in the proposed AutoTag

system, the LSTM network is utilized to encode the respiration signal sequence within a time window. Then its outputs are used to obtain estimations for the mean vector $\mu_\phi(x)$ and the variance vector $\sigma^2_\phi(x)$ using two linear modules, respectively. The sampled z is fed to another LSTM network for decoding the estimated mean and variance vectors. Eventually we obtain a reconstructed respiration signal for the same time window.

Once the reconstructed respiration signal is obtained, we propose a KL divergence based method for apnea detection. Specifically, we first normalize both the original signal and the reconstructed signal in the same time window, in order to ensure that both the apnea signal and the respiration signal are within the same amplitude range. The similarity between the original signal and the reconstructed signal is then calculated in the form of KL divergence. Since the proposed neural network is well trained by a large amount of normal breathing signal, the KL divergence between input signal and reconstructed signal is very small when the signal is sampled during normal breathing. In contrast, the KL divergence is very large when the input signal includes abnormal respiration (apnea), because the network didn't know the features of these abnormalities. Finally, we collected the values of KL divergence calculated for normal breathing and apnea respectively, and a threshold λ is properly chosen to determine whether the signal in this time window is for apnea or normal respiration. Specifically, if the KL divergence is greater than the threshold, apnea is detected in this time window; otherwise, the signal is for normal respiration.

The proposed deep learning based approach has two desirable features. First, the recurrent variational autoencoder is an unsupervised learning. Therefore, there is no need to collect labeled data for regular and abnormal respiration signals, which could be costly and time consuming. Second, the proposed method can learn the periodic features of respiration signal in offline training, rather than simply detecting the strength of the breathing signal. Therefore, this method is superior to the energy threshold based method, especially when the patient moves.

Peak Detection for Breathing Rate Estimation

Finally, the peak detection technique [18] is used to estimate the interval between two neighboring peaks, when a regular respiration signal is detected. Although most of the noise coming

from the environment have been removed at this stage, some false peaks may still exist. False peaks are usually relatively smaller than the peaks of the respiration signal, but they still could affect the accuracy of peaks detection.

To mitigate the influence of such false peaks, we execute the peak detection algorithm with a small sliding window of data points, instead of on the entire sampled signal sequence. In AutoTag, a window size of 11 data points are used. Only if the center data point of the sliding window is the maximum among all the data points within the window, the center data point will be identified as a peak. Then the mean value of all the peak intervals is calculated to approximate the period of the respiration signal τ. Finally, the respiration rate can be computed as $60/\tau$ breaths per minute (bpm).

2.5 Prototyping, Experiments, and Discussions

2.5.1 Prototyping and Experimental Environments

To evaluate the proposed AutoTag system, we adopt Impinj R420 as the RFID reader to collect phase information from ALN-9740 tags. To be FCC-compliant, the circular polarized antenna equipped in our system hops among 50 channels ranging from 902.5 MHz to 927.5 MHz, and remains on each channel for 200 ms. The user interface and the signal processing is implemented with an MSI laptop equipped with a Nvidia GTX1080 GPU and an Intel(R) Core(TM) i7-6820HK CPU. A software is also implemented in our system to collect data from the reader using the Low-level Reader Protocol (LLRP), which can extract useful low-level data from received tag responses, including time stamp, Doppler shift, RSSI, and phase value.

We carry out extensive experiments that involve four volunteers. The experimental results in two different environments are presented in this section. The test settings include a 5.6 m $\times$ 7.5 m lab, which is cluttered with tables, chairs, and computers, and a 2.4 m $\times$ 20.0 m empty corridor with no moving persons in Broun Hall in the Auburn University Campus. In the lab setting, the multipath effect is obviously larger than that in the corridor setting. All the volunteers are tested in three cases: (i) sitting in a chair and breathing normally, (ii) sitting in the chair and holding breath, and (iii) moving randomly while breathing normally.

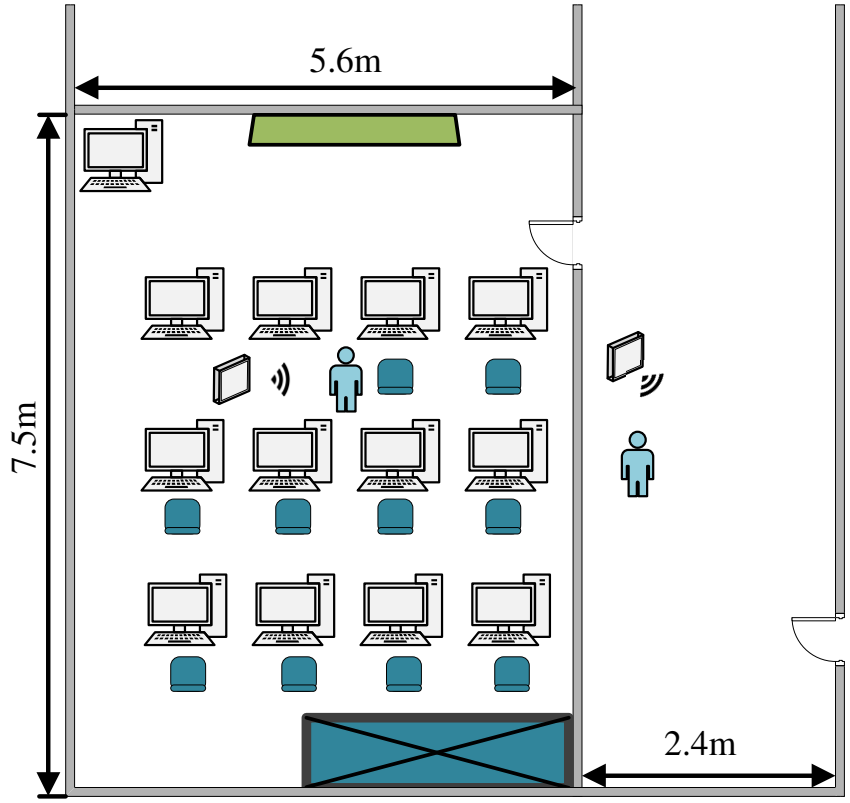


Figure 2.11: Experimental environments for validating the performance of AutoTag: (i) a cluttered computer laboratory; (ii) an empty corridor.

For respiration rate estimation, we obtain the cumulative distribution function (CDF) of estimation errors for performance validation. For apnea detection, the following two performance metrics are used:

- True Negative (TN) rate: this is the success rate when a regular respiration signal is successfully detected;
- True Positive (TP) rate: this is the success rate when apnea is correctly detected.

2.5.2 Experimental Results and Discussions

For normal breathing scenario, we evaluate the accuracy of respiration rate estimation under the two settings. We also use the NEULOG Respiration Belt sensor wrapped on the volunteer's chest to measure the ground truth. The CDFs of estimation errors are plotted in Fig. 2.12 for the lab and corridor experiments. We find the maximum errors are 0.462 bpm and 0.326 bpm, while the median errors are 0.105 bpm and 0.093 bpm, for the lab and corridor settings, respectively.

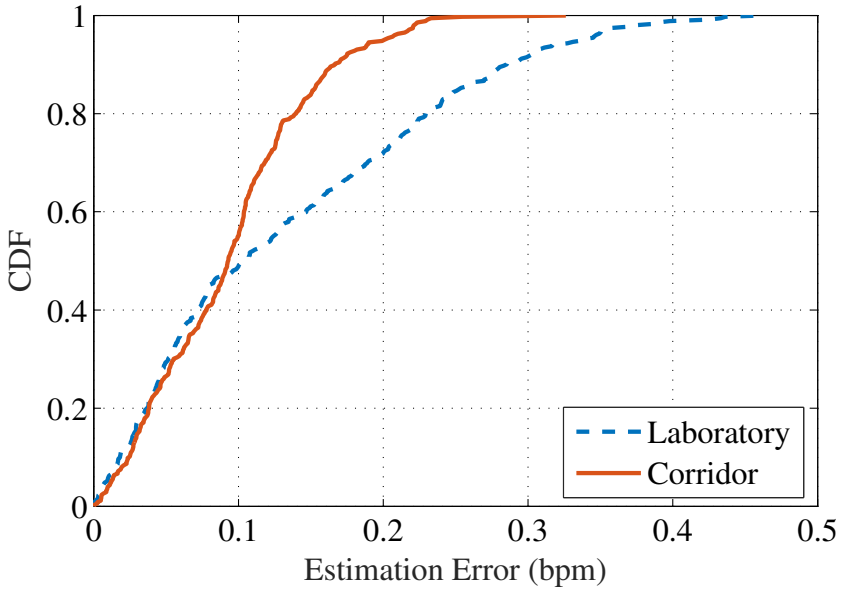


Figure 2.12: CDFs of estimated breathing rates in the computer lab and corridor scenarios.

The maximum and median errors in the lab setting are both larger than the corresponding error in the corridor setting. In addition, more than 50% estimation errors obtained in the corridor setting are smaller than that obtained in the lab setting. These indicate that the multipath effect does affect the accuracy of respiration rate estimation. Furthermore, it can be seen that the median error in the lab and corridor tests are 0.104 bpm and 0.0925 bpm, respectively. The close median errors indicate that the effect of multipath is not substantial. Since a directional antenna is used by the reader, the backscattered tag response in the line-of-sight (LOS) path is the dominant component. Thus, the respiration rate estimation in the two settings are both accurate.

We also study the influence of various factors on the estimation precision of our system. Fig. 2.13 plots the impact of the distance between two adjacent tags. We test the system with a tag array attached to the human chest with different distances between adjacent tags. The figure shows that when the distance between two tags is 1 cm, the estimation error is 0.265 bpm, which is relatively large. This is because when tags are too close to each other, they will suffer from stronger mutual coupling effect. The measured phase will be distorted by mutual coupling, and thus the estimated respiration rate will also be affected. Fortunately, the figure also reveal that when the tags are more than 2 cm apart, the error will become smaller than 0.124 bpm, which

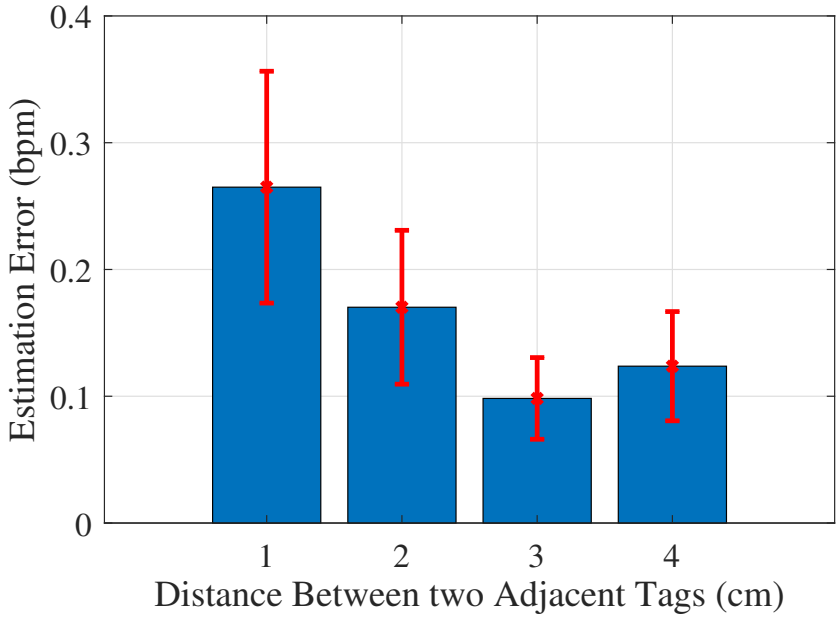


Figure 2.13: Evaluating the effect of the distance between two neighboring RFID tags.

indicates the influence of mutual coupling is negligible in these cases. Therefore, we deployed the tags with a 3 cm tag-interval in the AutoTag system.

Fig. 2.14 presents the influence of the number of tags used in AutoTag. We test the system with an increased number of tags from 1 to 5, while the tag distance is set to 3 cm. As can be seen in the figure, the estimation error is higher than 0.307 bpm when only one tag is used. When 3 or more tags are used, the error becomes smaller than 0.134 bpm. This is because the sensitivity of the single tag is more susceptible to the propagation environment, such as different surroundings and human postures. When multiple tags are used, we select the most sensitive tag to measure the human respiration, which increases the sensitivity of the system. Based on the results showed in Fig. 2.14, we adopt 3 tags on the volunteers to increase the accuracy of our system.

We also experiment by placing tags on different parts of the human body. Fig. 2.15 shows that, the estimation error is large when the tags are placed on the arms and neck of the volunteers, which are 0.637 bpm and 0.304 bpm, respectively. When the tags are placed on the chest and abdomen, the error becomes lower than 0.119 bpm. This is because respiration directly causes the chest and abdomen to move. When the tags are placed on the neck and arms, the

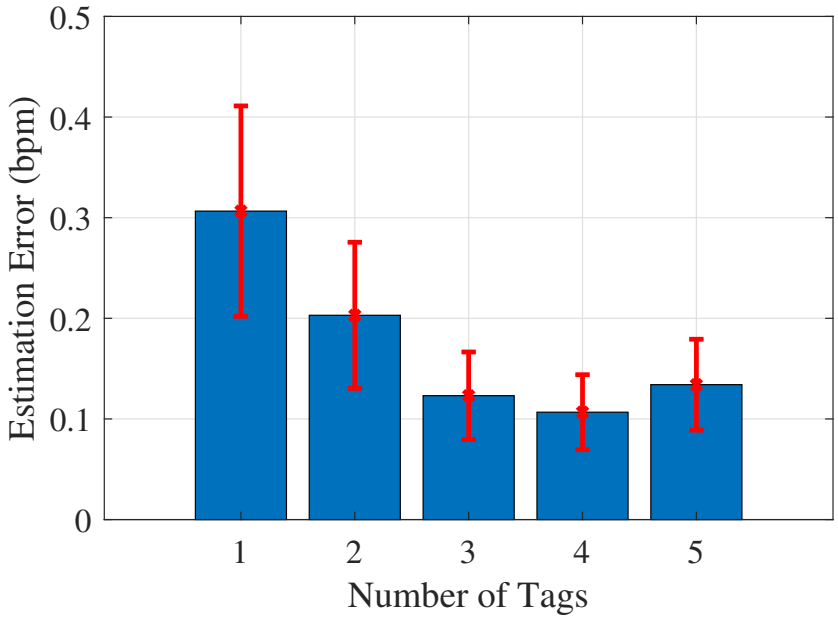


Figure 2.14: Evaluating the effect of the number of attached RFID tags.

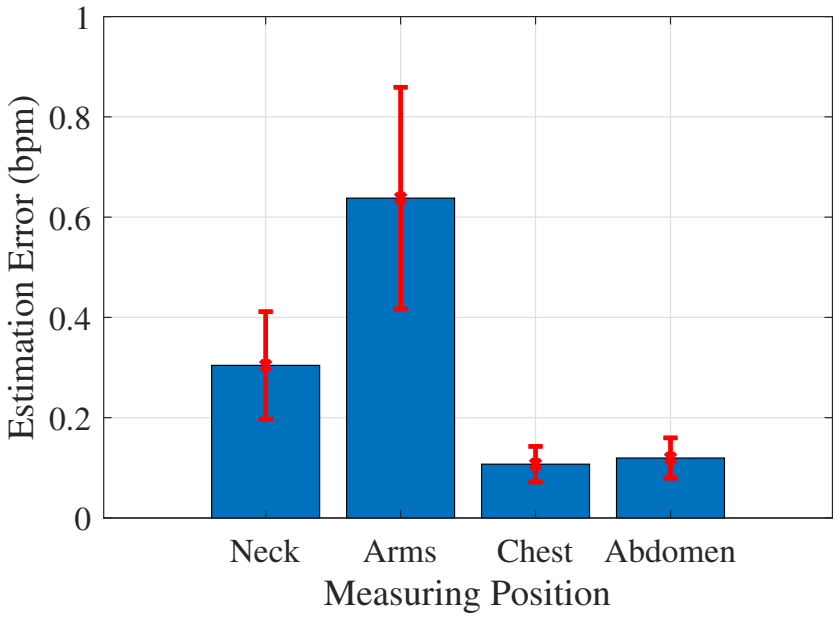


Figure 2.15: Evaluating the effect of different measuring positions.

strength of the movement, and thus the phase signal is not strong enough for effectively extraction of respiration signal, although the backscattered signal is also reflected from the chest. In addition, the arm movements generate a large noise, which also affect the accuracy of the system. Therefore, all the tags are attached to the human chest or abdomen in the AutoTag system.

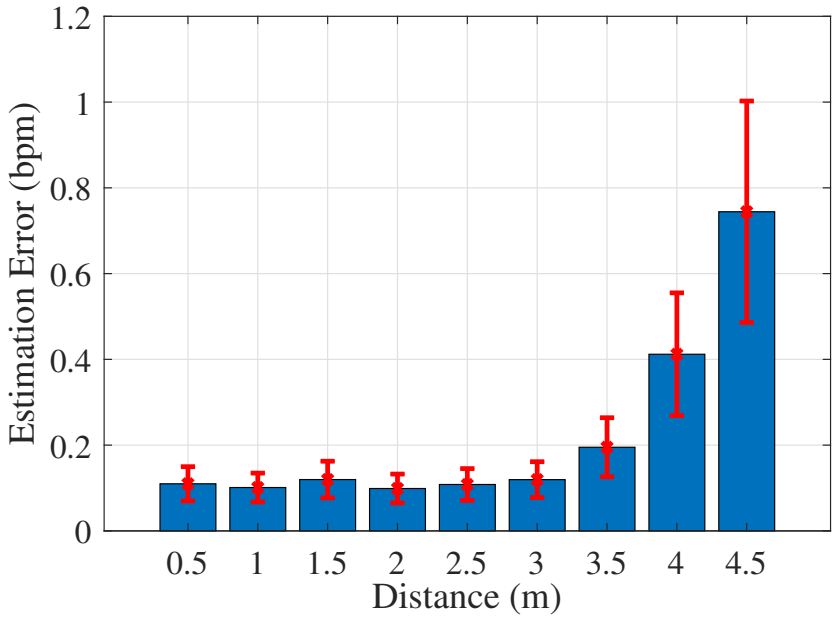


Figure 2.16: Evaluating the effect of the distance between the patient chest and reader antenna.

Next, the effective rage of our system is evaluated, and the results are presented in Fig. 2.16 and Fig. 2.17. Fig. 2.16 presents the estimation errors at different distances between the patient and the antenna. We find that the respiration rate error is less than 0.112 bpm when the distance is shorter than 3.5 m. The error increases drastically when the distance becomes larger than 4 m. To evaluate the impact of the angle of the directional antenna, we measure the respiration rate at a fixed distance with various relative orientation angles between the patient and the directional antenna: where $0°$ means the antenna directly faces the patient. As show in Fig. 2.17, the estimation error is smaller than 0.127 bpm when the angle is between -$20°$ and $20°$. The error increases to be larger than 0.678 bpm when the angle is beyond $\pm 60°$. This is because the received power of the backscattered signal is different at different radiation angles for the directional antenna. According to the results presented in Figs. 2.16 and 2.17, we conclude that the distance between the patient and the antenna should be less than 3 m and the orientation angle of the antenna should be within $\pm 20°$.

We also study the proposed AutoTag system by comparing with an energy based baseline method [41] under two different settings. In the first setting, the patient sits quietly with no other movements than breathing; in the second setting, the patient is allowed to move slightly, such as moving legs and hands, and twisting neck. Fig. 2.18 provides the TN and TP rates

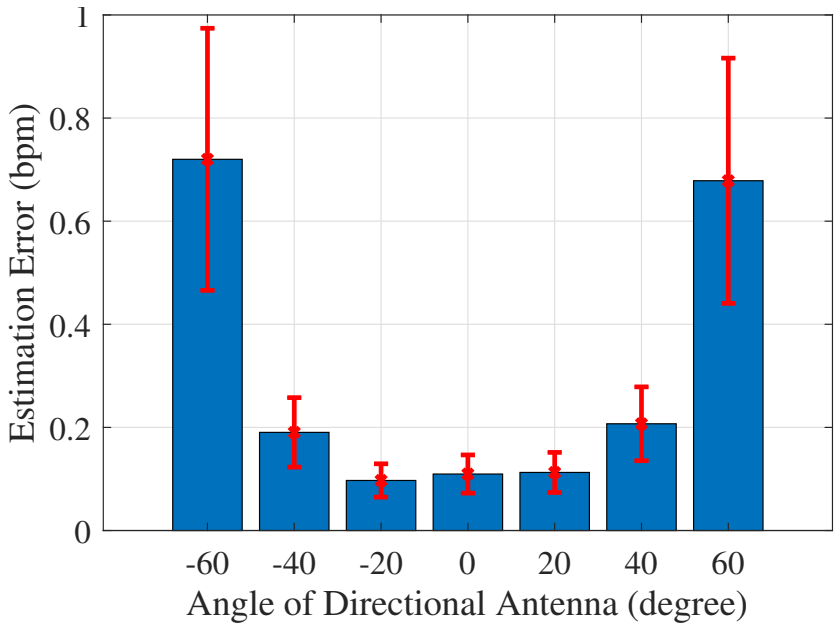


Figure 2.17: Evaluating the effect of the angle of the directional reader antenna.

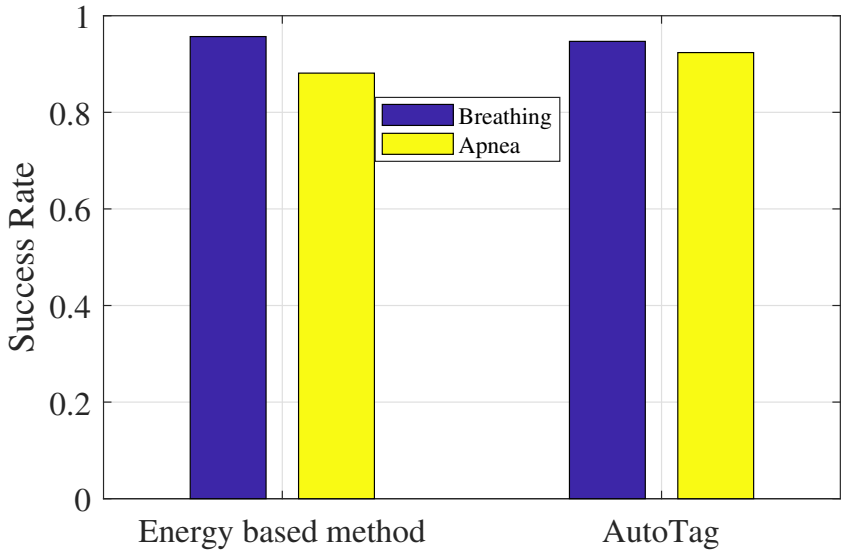


Figure 2.18: True Negative rate and True Positive rate obtained in a stable setting.

obtained with the proposed method and the baseline method in the stationary setting. When there are no body movements, the TN rates are both over 94% for the proposed and baseline schemes. Furthermore, the TP rates are 88% and 92% for the proposed and baseline schemes, respectively. These results indicate that both AutoTag and the energy based baseline method can achieve considerably high TN and TP rates when the patient does not make slight moves.

When the patient moves slightly, however, the situation becomes quite different. The results under slight body movements are given in Fig. 2.19. We find the TP rate of the energy

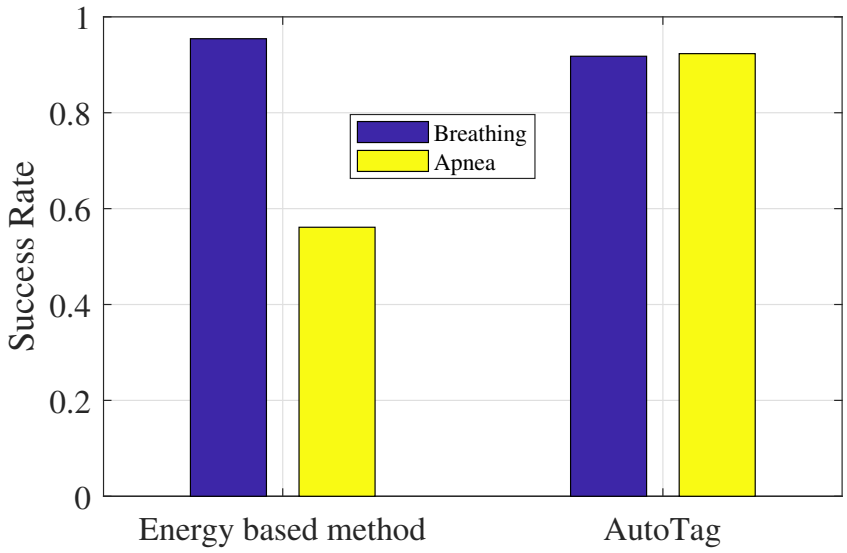


Figure 2.19: True Negative rate and True Positive rate obtained when there are small body movements.

based baseline scheme drops to 56%, but the TP rate of AutoTag is still as high as 92%. Under the disturbance of small body movements, the energy of the apnea signal can be greatly increased. It is therefore challenging to choose a suitable threshold for apnea detection in the baseline scheme. In contrast, AutoTag detects respiration abnormality by matching the shapes of the original and reconstructed signals; it does not rely on the signal's energy level. Consequently, AutoTag is resilient to the energy disturbance caused by small body movements in this setting. Furthermore, as shown in Fig. 2.19, the TN rates achieved by AutoTag and the baseline scheme are both above 91%, which are only slightly less than the TN rates achieved in the stationary setting (although the body movements also distort the shape of the respiration signal). We conclude that AutoTag system can accurately detect apnea and normal respiration signals in the two settings.

2.6 Conclusions

This paper presented the AutoTag system for unsupervised respiration rate estimation and detection of apnea in real-time with commodity RFID Tags. The AutoTag system incorporated a novel technique to effectively address the effect of frequency hopping offset for RFID systems

that comply to FCC regulations, and thus can be used for many RFID applications with real-time requirements. The proposed system also incorporated an unsupervised learning, thus has the desirable advantage of not requiring labeled medical data, making it low-cost and easy to deploy. The superior performance of the AutoTag system is demonstrated by extensive experiments in typical healthcare environments.

Chapter 3

Unsupervised Drowsy Driving Detection with RFID

3.1 Introduction

Driving fatigue is now considered as a primary cause of traffic accidents. It is reported by the National Highway Traffic Safety Administration (NHTSA) that, over 72,000 reported crashes involved drowsy driving from 2009 to 2013, and 16.5% of fatal crashes are caused by driving fatigue [43]. Human lives are at high risk in such accidents caused by drowsy driving. The situation is even worse with the increasing popularity of autonomous driving [44]. Such risks and losses can be greatly reduced if an effective driving fatigue alarm system is in place. However, most drowsy driving events are hard to detect with existing technologies in commodity vehicles. Thus, there is a compelling demand for an effective driving fatigue detection system, which can accurately detect driving fatigue and alarm drivers to avoid accidents [45–47].

Driving fatigue detection is a popular topic in the research community in recent years, and various types of signals have been exploited in prior works, such as electroencephalogram (EEG) [48], video camera [49], WiFi [50], and ultra sound [51]. As a straightforward signal from human brain, EEG signal can achieve an excellent performance on fatigue detection [48]. However, because of the complex equipment required, the EEG system is currently not suitable for practical use in cars. As a non-intrusive approach, vision based techniques can detect the driving fatigue by recognizing eyelid movements [49]. The required hardware of vision based system, i.e., a camera, is much cheaper than that in EEG based techniques, but the system requires sufficient lighting inside the vehicle, and the performance could be heavily affected if the driver wears sunglasses. Without the lighting requirements, radio frequency (RF) and acoustic

signals are also leveraged to detect driving drowsiness. For example, channel state information (CSI) of WiFi signals can be used to detect driving fatigue by detecting the respiration rate and movements of the driver [50]. Unfortunately, due to the large range, the WiFi signal is sensitive to the interference from surroundings, such as the movements of the driver and passengers, and of objects outside the vehicle. The same challenge also exits for the current acoustic-based techniques [51], which detects drowsy-driving with the embedded microphone and speaker in smartphones.

RFID sensing has drawn considerable attention recently, with interesting new applications for remote temperature sensing [52], drone navigation [24,53,54], gesture recognition [55,56], localization [57], and breathing monitoring [8,27,58]. The passive RFID tags can be directly attached to the target object. Due to the near-field nature, the interference of surrounding noises can be effectively mitigated. Furthermore, the cost of tags is low and the performance can hardly be affected by the lighting condition inside the vehicle. However, there have only been very limited work on application of RFID in vehicles, which are mostly focused on localization [59–61]. There are many challenges to build a highly accurate RFID based driving fatigue detection system, such as effectively extracting driving drowsiness features and the discontinuity in collected phase data as caused by frequency hopping.

In this paper, a driving drowsiness detection system is proposed to fully exploit advanced machine learning and RFID based sensing [62]. We firstly introduce the collected phase model in commodity RFID systems, as well as the challenges caused by channel hopping, such as the discontinuity of the sampled phase and the cumulative error caused by the frequency hopping offset. We then introduce the design of the proposed system, which is composed of four main components, including data sensing, movement feature extraction, offline training, and online drowsiness detection. Specifically, to effectively detect the nodding features with RFID tags in a driving environment, we deploy two RFID tags on a hat worn by the driver, and employ the phase difference between two RFID tags to mitigate the noise caused by vehicle vibration. According to the FCC policy, phase data of both tags are sampled sequentially, while the reader hops among various channels. A novel algorithm is proposed to estimate the phase difference between two RFID tags. With an analysis of the collected phase data, the cumulative error

caused by the channel hopping offset will also be eliminated by a novel differentiation process of collected phase difference. Finally, to avoid the high cost of collecting labeled data from driving environments, an unsupervised LSTM autoencoder model is proposed to distinguish the nodding movement from other driving movements. This is achieved by measuring the divergence between the input and reconstructed signal from the well trained autoencoder model. We have implemented the proposed system using commodity RFID tags and readers, and carried out extensive emulations and experiments in real driving settings, e.g., parked, city street driving, and high way driving, to validate the performance of the proposed system.

The main contributions made in this paper can be summarized as follows.

- To the best of our knowledge, this is the first work that leverages passive RFID tags for driving drowsiness detection under real driving settings.
- A specific tag deployment and several signal processing algorithms are proposed to effectively distinguish the nodding features from the strong environment noises and other types of driving related movements. An effective algorithm is proposed to estimate, on real-time, the phase difference between two RIFD tags that are interrogated with slotted ALOHA and under frequency hopping in commercial RFID systems.
- We analyze the cumulative error caused by the frequency hopping offset in FCC-compliant UHF RFID systems, and propose a differentiation based method to mitigate the influence of cumulative error.
- Driving fatigue is detected by an unsupervised LSTM autoencoder model, which does not require labeled training data of various types of driving movements, which are hard and costly to obtain.
- A prototype system is built with commodity RFID devices, deployed in a car, and validated in both an emulated environment and real driving environments. The experiments are conducted in various driving scenarios, where excellent performance of the proposed system is demonstrated.

In the rest of this paper, the related works are discussed in Section 3.2. The preliminaries are introduced in Section 3.3. We present the system design in Section 3.4 and performance evaluation in Section 3.5. Section 3.6 summaries this paper.

3.2 Related work

This work is highly relevant to the prior work on driving fatigue detection and RFID based sensing systems. Driving safety is a hot topic for recent years, and several techniques are proposed to detect drowsy driving to prevent drivers from falling asleep when driving [46]. For example, physiological signals, such as electrooculograms (EOG) and cephalography (EEG), are used to detect driving fatigue [48, 63]. Compared with EEG, EOG is more robust to noise because of its higher amplitude values. Drowsiness can be effectively detected when the physiological signal is labeled correctly. These systems usually have the highest sensitivity to drowsiness, because of they directly monitor the human brain. However, the EOG and EEG equipment are expensive and not suitable for deployment in vehicles.

Other types of systems are also proposed to achieve higher flexibility and reduce cost. Various types sensors, such as video camera, smartphone, and RF devices, are used. Different from physiological sensors, such sensors detect driving drowsiness by analyzing the movement of drivers, such as blinking, yawning, and nodding. Camera based systems could detect the eye location or eyelid movement [49, 64]. However, the accuracy of the system is highly dependent on the lighting condition inside the vehicle. It may also raise concerns of violating the privacy of drivers.

In addition, vital signs can also be effective indicators of drowsiness, which has been used in RF based techniques. Since drowsiness is closely related to respiration rate [65], respiration rate monitoring in driving environment becomes a promising approach. To this end, ultra-wideband (UWB) radar has been adopted to detect the breathing rate of drivers [66]. WiFi has also been utilized for this purpose [50]. Movements of the driver's chest could be captured by the chest reflected WiFi signal, and by examining the Channel State Information (CSI), the respiration rate can be estimated. One shortcoming of this approach is that the RF signals are sensitive to environment interference, such as the movements of the driver and passengers, as

well as the movements outside the car. It is a big challenge to mitigate the impact of such strong interference.

Smartphones are considered as a type of multifunctional platform for sensing because of its embedded sensors, such as video camera, microphone, and gyroscope, which enable numerous smartphone based sensing system designs. With the embedded video camera, smartphone can also be used to detect driving fatigue by capturing eye movements [67]. By incorporating the microphone and speaker, acoustic systems have also been developed to detect the movements of the driver, such as yawning, steering, and nodding [51]. High flexibility and low-cost are the two key benefits of smartphone based acoustic systems. However, acoustic signals are also very sensitive to the movements of the driver and passengers, as well as vehicle vibration, which may hurt the performance of such systems.

Recently, passive RFID tags, as a kind of wearable sensors, have attracted increasing interest because of its low-cost and easy deployment features. RFID based sensing has been used for many applications, such as user authentication [68], material identification [69], object orientation estimation [23], vibration sensing [70], and anomaly detection [71]. For indoor localization [22,34,57] and gesture recognition [72,73], the RFID based techniques are mainly focused on the analysis of low level data collected at the reader. For example, the received signal strength (RSS) has been utilized for tag localization in [31], while the phase values have been used to recognize different kinds of gestures [55]. In addition, vital signs can also be detected by the low level data. Specifically, TagBreathe is the first work to estimate breathing rates using RFID tags [27], while TagSheet uses RFID Tags for breathing monitoring and sleep posture recognition [74]. Even heart rate variability can be assessed with an RFID tag array attached to the human body. However, these vital sign monitoring systems are not suitable for detecting drowsiness in a driving environment, because the small vital sign signal could be easily overwhelmed by vehicle vibration and driving movements. The work presented in this paper makes a first attempt on RFID based driving fatigue detection, where commercial RFID tags are utilized for detecting the nodding movement of the driver. The proposed system consists of several novel techniques to effectively deal with the strong noisy driving environment, as will be elaborated in Section 3.4

3.3 Preliminaries and Challenges

3.3.1 Measured Phase at an RFID Reader

To distinguish different types of head movements, we need to detect and analyze the variation of the distance between the reader and the tags attached to the driver's hat. Such changes are captured by the phase values collected by the RFID reader. According to the low level reader protocol (LLRP), the reader can provide low level data, such as Received Signal Strength Indicator (RSSI), RF phase, and Doppler Shift, for each received tag response [38].

The received phase value can be written as [8]

$$\Phi = \text{mod}\left(\frac{2\pi(2L)}{\lambda} + \Phi_R + \Phi_T + \Phi_{tag}, 2\pi\right), \tag{3.1}$$

where L is the distance between the reader antenna and the target tag, λ is the wavelength of the signal, Φ_R and Φ_T represent the phase offsets caused by the receiver and transmitter, respectively, and Φ_{tag} is the phase shift caused by the refection circuit of the target tag. Since λ, Φ_R, Φ_T, and Φ_{tag} are constant when the reader operates on a given channel, the collected phase value Φ varies along with the change in the tag-to-reader distance L (i.e., chest movements).

3.3.2 Frequency Hopping Offset and Cumulative Error

According to FCC regulations, Ultra High Frequency (UHF) RFID readers should adopt frequency hopping to benefit from the maximum reader transmitted power allowances. When interrogating tags, the reader periodically hops among 50 different channels from 902 MHz to 928 MHz. Since the values of λ, Φ_R, Φ_T, and Φ_{tag} in (3.1) are all related to the operation frequency, the measured phase is affected by both the tag-to-reader distance and the current occupied channel.

The measured phase from a channel k can be written as

$$\Phi(f_k, L) = \text{mod}\left(\frac{4\pi L f_k}{c} + \Phi_k, 2\pi\right), \tag{3.2}$$

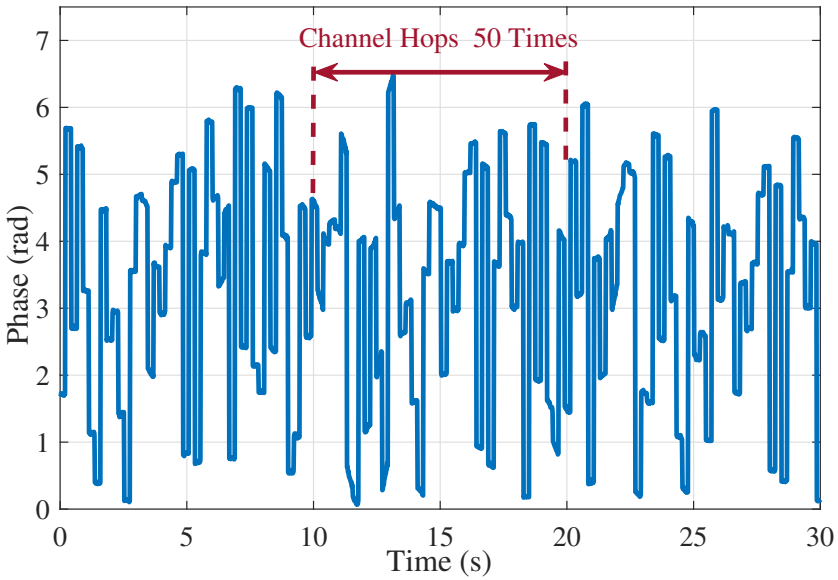


Figure 3.1: Raw phase data collected by the RFID reader.

where c is the speed of light, f_k is carrier frequency of channel k, and Φ_k represents the initial phase offset on channel k due to Φ_R, Φ_T, and Φ_{tag}.

Fig. 3.1 shows the raw phase data collected by the reader. It can be seen that the reader hops 50 times in a period of 10 seconds and the frequency hopping offset causes considerable discontinuity in the measured phase data. Thus, the variation of L, which represents the useful signal, is hard to be detected from the raw phase data. To address this issue, two solutions have been proposed in recent works. The Tagyro system adopts a calibration process of 10 seconds to estimate the initial phase offset for each channel, and then subtract it from the measured phase data [23]. This method works well in a static environment; but it is not suitable for RFID systems in a noisy driving environment. This is because the movements of the driver and vehicle vibrations could hurt the accuracy of the calibration process. In the respiration monitoring system Autotag [8, 58], a real-time method is proposed to mitigate the frequency hopping effect. Rather than estimating the initial phases on all channels with a calibration phase, the Autotag system focuses on mapping the phase data sampled in the current channel to the previous channel, by removing the frequency hopping offset between two adjacent channels.

The proposed method in [8,58] can eliminate most of the frequency hopping offset for realtime sensing applications. However, there is still some residual error remains each time when the reader hops to another channel, and the error will accumulate to become larger and larger

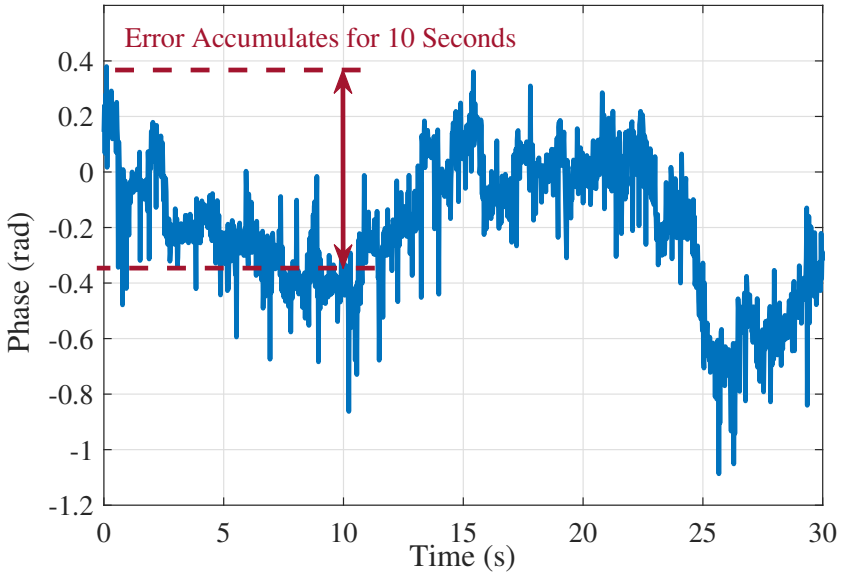


Figure 3.2: An example of the cumulative error in calibrated phase signal.

as the reader hops among more and more channels. For the respiration rate monitoring problem considered in [8, 58], such cumulative error can be effectively removed with a detrending process, because breathing rate detection only concerns the periodicity of the signal. However, for the driving fatigue detection problem considered in this paper, the information of head movements is also embedded in the low frequency components of the signal. If a detrending process is applied as in [8, 58], the useful nodding signal will also be removed.

In Fig. 3.2, we plot the phase data collected from a tag attached to an object (i.e., a book) in a stationary state. The sampled phase data are calibrated by the proposed method in [8, 58]. It can be seen that although the object is static, the calibrated phase still exhibits large variations. For the first 10 seconds, the error accumulates to 0.76 rad, which will greatly affect the accuracy of phase measurements. This is because the residual error in estimating the initial phase offset for each channel happens every 0.2 second (when the reader hops to a new channel), and the error starts to accumulate over time. On the other hand, if we only use the phase data collected from the same channel, it will take 10 seconds for the reader to return to the same channel, making it unsuitable for realtime sensing applications. In order to detect head movement features from the calibrated phase data from all channels, accurate phase data should be firstly estimated. Thus, extracting movement features from the calibrated phase signal with cumulative error is a big challenge.

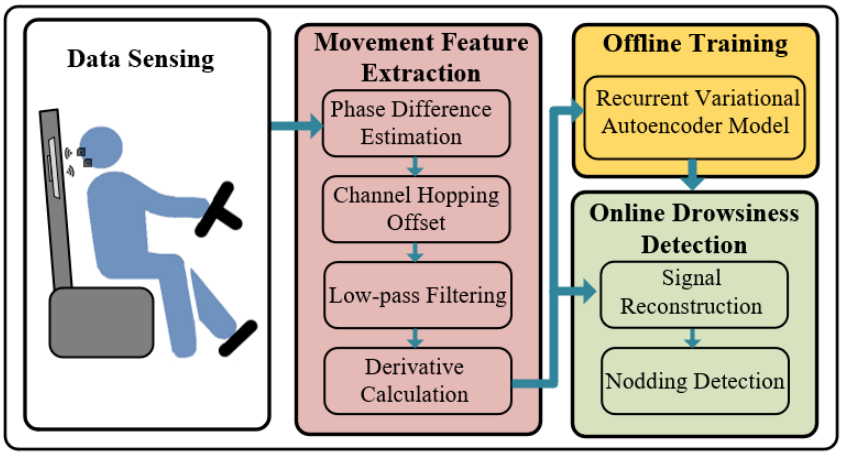


Figure 3.3: Architecture of the proposed system.

3.4 System Design for Drowsy Driving Detection

3.4.1 System Overview

The proposed unsupervised driving fatigue detection system is illustrated in Fig. 3.3. Our system is composed of four main modules, including data sensing, movement feature extraction, offline training, and online drowsiness detection. In the data sensing module, head movement could be captured by received phase values from the tags attached to the driver's hat. Then, the nodding features are distinguished from other head movements in the movement feature extraction module. The phase difference between two RFID tags are estimated to mitigate the influence of vehicle vibration. Derivative calculation is proposed to remove the cumulative error caused by the realtime frequency hopping. Finally, an unsupervised learning model is proposed to learn the nodding features, and the online nodding detection is executed with the well-trained model. Nodding will be detected by calculating the divergence between the input and output signals of the autoencoder model. The detailed design of the proposed system will be elaborated in the remainder of this section.

3.4.2 Nodding Feature Extraction

In many machine learning systems, offline training requires a large amount of featured data. However, it's a big challenge to learn the features of normal driving based on head movements, because drivers may randomly rotate their head to the left or right to check side-view mirrors

or traffic conditions in different lanes. Such head movements during normal driving are usually unpredictable. In contrast, nodding is a typical symptom of driving fatigue, which can be easily labeled simply from collected data. Therefore, we use the features of nodding from collected data for training the model. There are still some challenges remaining. *First*, drivers may change their posture or move their head forward or backward during driving, and thus the head movements include both 3-D rotations and position shifts. It is difficult to separate the head shifting signal from the collected signal, because phase value is affected by both types of movements. *Second*, nodding features are hard to distinguish from head rotation. This is because both movements can be considered as a round-trip rotation of the head, which makes the resulting phase variations very similar. *Finally*, as discussed in Section 3.3, the cumulative error caused by the channel hopping offset is still a big problem. In the following subsections, we address all these challenges and show how to effectively extract the nodding features.

Phase Difference Calculation

To mitigate the impact of driver's body movements and vehicle vibration on collected data, we calculate the phase difference between two tags rather than directly utilizing the calibrated phase data. Since the driver is buckled up, the body movement is usually constrained, and the typical head movements include shifting, rotation, and nodding.

Fig. 3.4 illustrates the three types of typical head movements when driving. All types of movements generate phase variation of RFID tags. It is hard to differentiate them with a single RFID tag; so we leverage two tags to sense the head movement. We find that, although head shifting and vibration affect the phase value of each tag, the influence on the phase difference between the two tags could be negligible [92]. This is because both head vibration and head shifting generate the same alteration of tag-to-reader distance for both tags, resulting in similar phase shifts that are canceled when calculating phase difference. In contrast, both nodding and head rotation could cause different alterations in the tag-to-reader distances of the two tags, resulting in a large phase difference change. Thus, in order to mitigate the influence of head shifting and vibration, phase difference is more suited for extracting nodding features than phase values collected from a single tag.

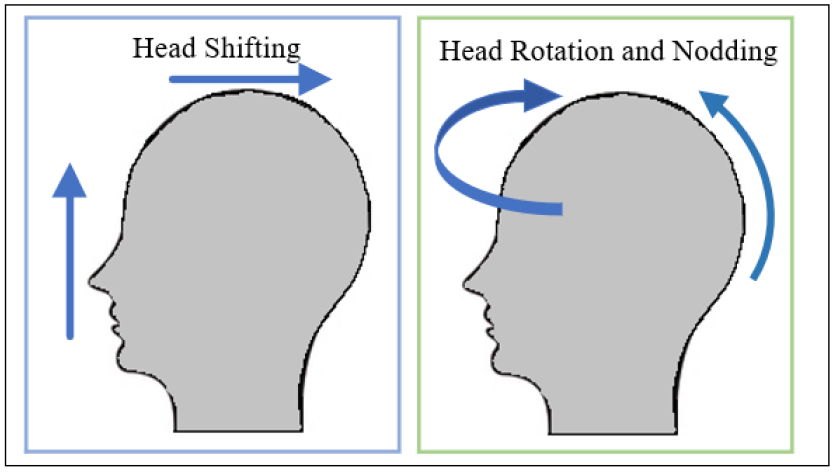


Figure 3.4: Three types of head movements.

Algorithm 1: Phase Difference Calculation Algorithm

Input: Phase collected from two RFID tags (Tag A and Tag B) from the same channel, denoted by P_n^a and P_n^b, $n = 1, 2, ..., N$, and the timestamp for each data frame, denoted by T_n^a and T_n^b, $n = 1, 2, ..., N$;
Output: Phase difference between two tags P_n^d, $n = 1, 2, ..., N$;
//Search for the nearest sample ;
for $n = 1 : N$ **do**
    Set $T_{previous}^{ab} = |T_n^a - T_1^b|$;
    for $m = 2 : M$ **do**
        $T_{current}^{ab} = |T_n^a - T_m^b|$;
        if $T_{current}^{ab} > T_{previous}^{ab}$ **then**
            $P_n^d = P_n^a - P_{m-1}^b$;
            break ;
        else
            $T_{previous}^{ab} = T_{current}^{ab}$;
        end
    end
end

Unfortunately, following the RFID anti-collision protocol, the communications between the tags and reader are based on slotted ALOHA protocol, which means only one tag can send its EPC and low level data to the reader in every time slot. The slotted ALOHA based transmission determines that the phase values are sampled sequentially. Therefore, it is impossible to obtain the phase values from both tags at the same time to calculate the phase difference. To address this issue, we propose an effective algorithm to estimate the phase difference on each individual channel, as shown in Algorithm 1.

First, we collect the phase sequences sampled from two tags on the same channel, which are denoted by P_n^a and P_m^b, respectively, together with their corresponding time stamps T_n^a and T_n^b. Second, for each phase sample in P_n^a, we search for the *nearest* Tag b phase sample by calculating the difference between two time stamps as $|T_n^a - T_m^b|$. In the algorithm, we scan the phase sequence P_m^b from 1 to M following the sampling order. Since the phase data is sampled continuously, the timestamp value for each tag is always increasing. Once the current time difference $|T_n^a - T_m^b|$ is larger than the previous one, the following calculated difference will keep on increasing. Thus, the previous sample P_{m-1}^b right before $T_{current}^{ab} > T_{previous}^{ab}$ will be the nearest sample with the minimum time difference. Finally, the phase difference sample sequence P_n^d will be calculated by subtracting each of the sampled phase data in P_n^a from the selected, nearest phase sample in P_m^b.

Tag Deployment and Data Collection

After the influence from head shifting is successfully mitigated, we next distinguish nodding from head rotation. Fig. 3.5 shows the calibrated phase data from a single tag after frequency hopping offset mitigation. The data is sampled when the driver nods and looks around (i.e., head rotation) sequentially, as marked in the figure. However, it is hard to differentiate nodding from head rotation based on the calibrated phase data, because both movements generate sharp peaks in phase values. For the purpose of extracting unique nodding features, we adopt *a simple solution with a specific tag deployment*. As shown in Fig. 3.6, the tags are attached to the back side of the head *horizontally* (i.e., on a hat). When the driver looks around to check traffic, the head movement can be approximately considered as a horizontal rotation. Such a head rotation causes similar changes in the tag-to-reader distance for the two tags, so that the change in phase difference is negligible. In contrast, during nodding, one tag moves closer to the reader while the other tag moves away from the reader. Hence the phase difference between the tags will increase sharply.

Fig. 3.7 shows the calibrated phase difference between two tags placed horizontally on the back side of the head. The data is sampled when the driver nods and rotates the head sequentially, as marked in the figure. We find that the data sampled during nodding is sufficiently

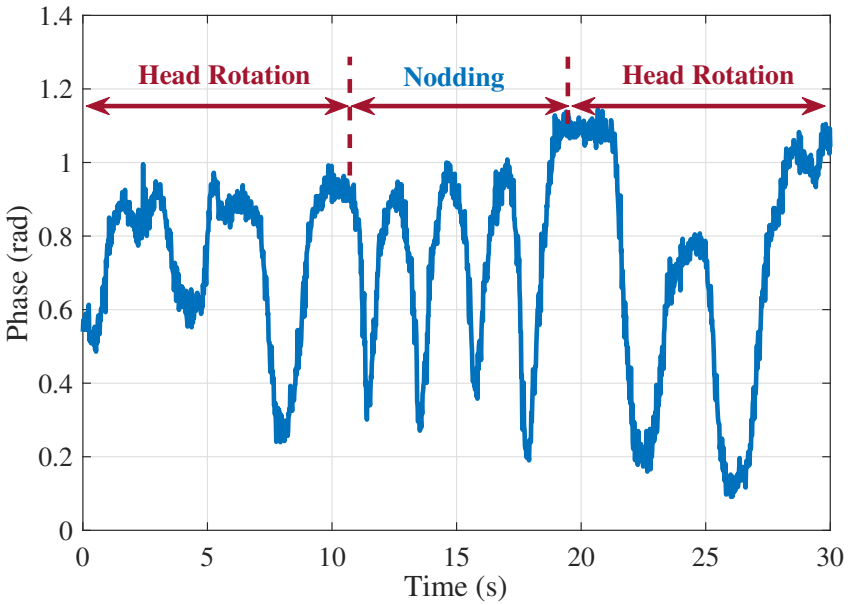


Figure 3.5: Measured phase data from a single RFID tag when the driver looks around and nods sequentially.

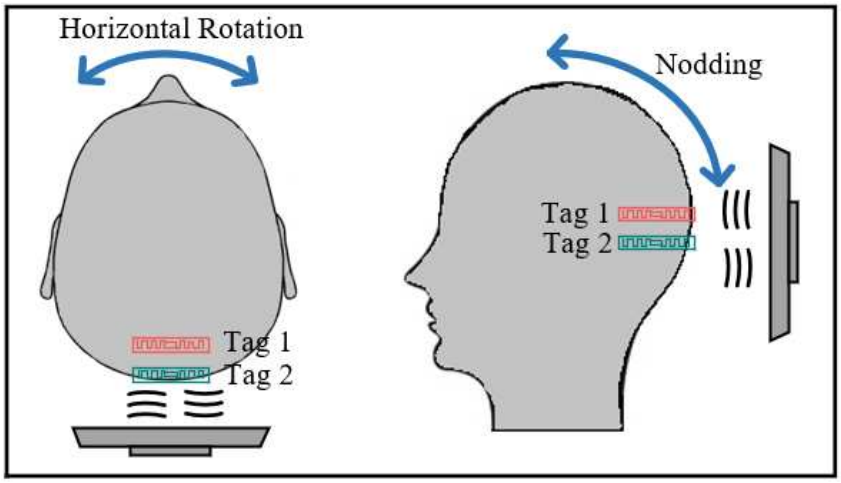


Figure 3.6: A special tag deployment scheme with two tags horizontally attached to the back of head (e.g., on a hat).

different from that sampled in the rotation period; head rotation does not generate sharp peaks on the calibrated phase difference. Thus nodding features can be effectively extracted from the phase difference between the two tags deployed as shown in Fig. 3.6.

Mitigating the Cumulative Error

To further improve the feature extraction performance, the cumulative error due to frequency hopping should be addressed, because the calibrated phase difference may be significantly

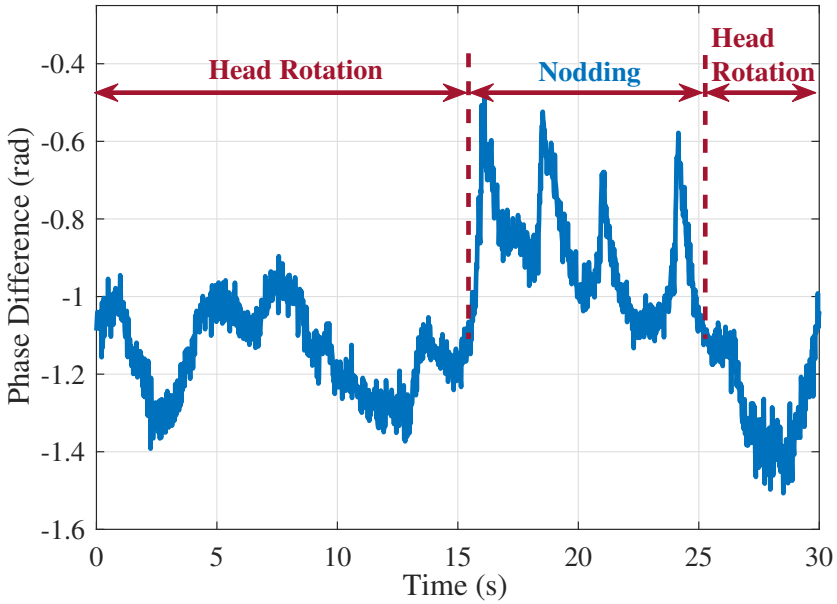


Figure 3.7: Calibrated phase difference between two horizontally attached tags when the driver looks around and nods sequentially.

distorted after a long period of driving. To address this issue, we first provide an analysis of the cumulative error. According to (3.2), the raw collected *phase difference* can be simply obtained by subtracting the phase value from one tag (with distance L_b) from the phase value from another tag (with distance L_a), as

$$\Delta\Phi(L_a, L_b) = \text{mod}\left(\frac{4\pi(L_a - L_b)f_k}{c} + \Delta\Phi_k, 2\pi\right), \tag{3.3}$$

where L_a and L_b are the tag-to-reader distance from Tag 1 and Tag 2, respectively, $\Delta\Phi_k$ is the initial phase offset difference between the two tags. Note that our analysis is mainly focused on the influence of the initial phase offset. So we neglect the multipath effect and mutual coupling between the two tags, and assume the phase difference is only affected by the tag-to-reader distances (i.e., L_a and L_b) and frequency hopping (i.e., $\Delta\Phi_k$).

We calibrate the data by mapping all phase difference data on the current channel to the previous reference channel [8], such that all the calibrated data can be considered as sampled from the same reference channel (i.e., to the first channel f_1 used when the measurement starts, as a reference channel). With the translation, all $\Delta\Phi_k$'s will be converted to $\Delta\Phi_{k-1} + \delta_1$, where δ_1 is the estimation error caused by each conversion. Although the estimation error δ_i is negligible for each ith conversion, it will accumulate overtime. After i times of frequency

hopping, $\Delta\Phi_k$ will be converted to $\Delta\Phi_1 + \sum_i \delta_i$. Thus, the calibrated phase difference after i hops is given by

$$\Delta\Phi(L_a, L_b) = \text{mod}\left(\frac{4\pi(L_a - L_b)f_1}{c} + \Delta\Phi_1 + \textstyle\sum_i \delta_i, 2\pi\right) \tag{3.4}$$

where f_1 and $\Delta\Phi_1$ are the frequency and initial phase offset difference on the first (i.e., reference) channel, respectively, δ_i represents the estimation error generated by the ith frequency hopping.

In (3.4), $\sum_i \delta_i$ is hard to estimate from collected data, because we do not know the accurate values of L_a, L_b, and $\Delta\Phi_1$. However, if we differentiate both sides of (3.4), the constant $\Delta\Phi_1$ will be removed from the equation. The derivative of the estimation error, δ', only remains in the first sample for each channel, so the error accumulation over time is effectively stopped. Suppose channel hopping starts from channel 1. When the system hops to channel k, it collects n_k samples (i.e., calibrated phase difference data) on channel k. The derivative of the channel k samples at time n, $n \in \{1, 2, ..., n_k\}$, can be derived as:

$$\begin{aligned} &\Delta\Phi'_n(L_a, L_b) \\ &= \begin{cases} \frac{4\pi f_1}{c}(L'_a - L'_b) + \delta'_{k-1}, & n = 1 \\ \frac{4\pi f_1}{c}(L'_a - L'_b), & n = 2, 3, ..., n_k, \end{cases} \end{aligned} \tag{3.5}$$

where L'_a and L'_b are the derivative of the tag-to-reader distances of Tag 1 and Tag 2, respectively, and δ'_k is the derivative of δ_k with $\delta'_0 = 0$. We can see from (3.5) that although estimation error still remains in the derivative of the first sample when the system hops to a new channel, it has been removed from the derivatives of the remaining samples on the new channel.

Unfortunately, we find that the derivative of phase difference cannot be directly used to extract nodding features because of the large noise. To demonstrate this observation, in Fig. 3.8, we plot the derivative of the signal plotted in Fig. 3.7. We find that the nodding features, which are quite obvious in Fig. 3.7, however, are completely overwhelmed by the white noise. This is because the differentiation operation can be considered as a high pass filter applied to the

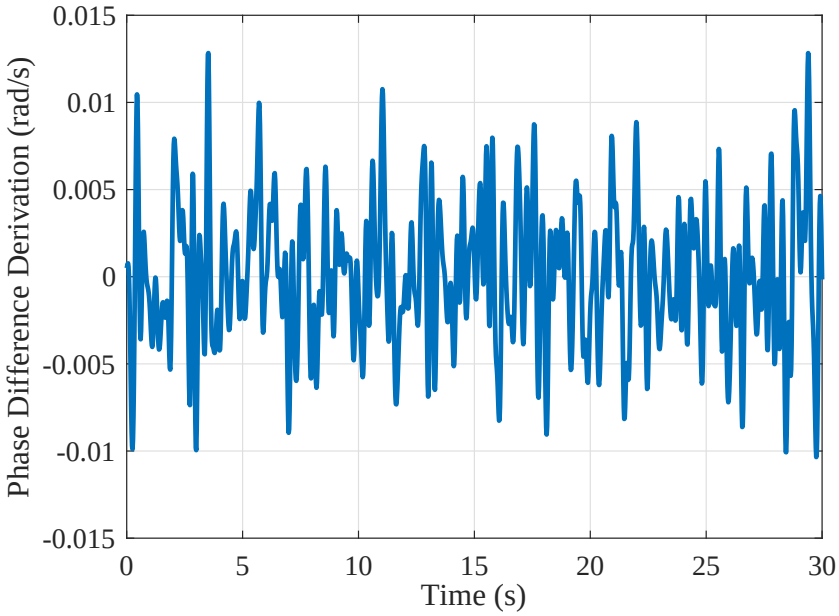


Figure 3.8: Derivative of the calibrated phase difference data given in Fig. 3.7.

calibrated signal. For convenience, we first assume that all phase differences are sampled at the same sampling rate of 55Hz (as tested in our experiments). Then the differentiation operation can be transformed into a convolution between the input signal and a vector $[F, -F]$, where F is the sampling frequency. To get more data in the frequency domain, we zero padding the vector by adding $N-2$ zeros after $-F$. Thus we could obtain a vector with length N, and the Discrete Fourier transform (DFT) result of the vector can be expressed as

$$\Gamma_k = \sum_{n=0}^{N-1} f_n \cdot e^{\frac{i2kn\pi}{N}}, n \in \{0, 1, ..., N-1\}, k \in \{0, 1, ..., N-1\},$$

where f_n represent the nth sample in the vector before DFT. Since we have $f_0 = -F$ and $f_1 = F$, while $f_n = 0, n = 2, 3, ..., N-1$, Γ_k can be written as

$$\begin{aligned} \Gamma_k &= -Fe^{\frac{i2\pi}{N}k\cdot 0} + Fe^{\frac{i2\pi}{N}k\cdot 1} \\ &= F\cos\left(\frac{i2\pi k}{N}\right) - F + i\sin\left(\frac{i2\pi k}{N}\right). \end{aligned} \tag{3.6}$$

The system gain in the frequency domain can be represented as the modulo of Γ_k. From (3.6), we find that the gain equals to 0 when $k = 0$, which means the 0 Hz component is removed by the differentiation operation. In contrast, when $k = N/2$, which represents $F/2$ Hz, the

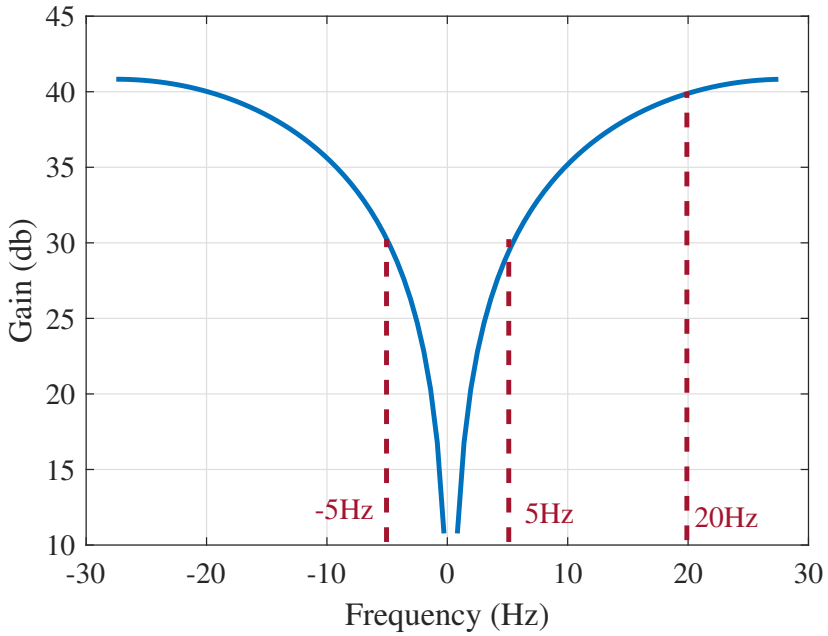


Figure 3.9: The differentiation effect in the frequency domain.

gain reaches its maximum value of $2F$. With $F = 55$ Hz used in the system, we can map Γ_k for different frequencies ranging from -27.5Hz to 27.5Hz. The result is shown in Fig. 3.9, which shows that differentiation leads to extremely large gains at high frequencies (higher than 10 Hz), while significantly suppressing the signals below 5 Hz. However, when we analyze the frequency domain response of calibrated phase difference signal plotted in Fig. 3.10, we can see that the power of the signal mostly concentrate in the low frequency region, ranging from $-5Hz$ to $5Hz$. Thus, we can conclude that both nodding movements and other driving movements are mostly composed of low frequency signals, which is highly attenuated by the differentiation operation. Besides, the white noise existing in the high frequency region will be considerably amplified. Consequently, only the high frequency noise remain after the differentiation operation, as shown in Fig. 3.8.

To mitigate such negative influence caused by the differentiation process, we firstly incorporate a low-pass filter with a 5 Hz cutoff frequency to filter the calibrated phase signal before applying differentiation. After filtering, the low frequency component will be amplified while the high frequency noise will be greatly suppressed, so that the movement related components could remain after differentiation. The final results after filtering and differentiation are plotted in Fig. 3.11. The figure shows that the high frequency noise is effectively removed, and there

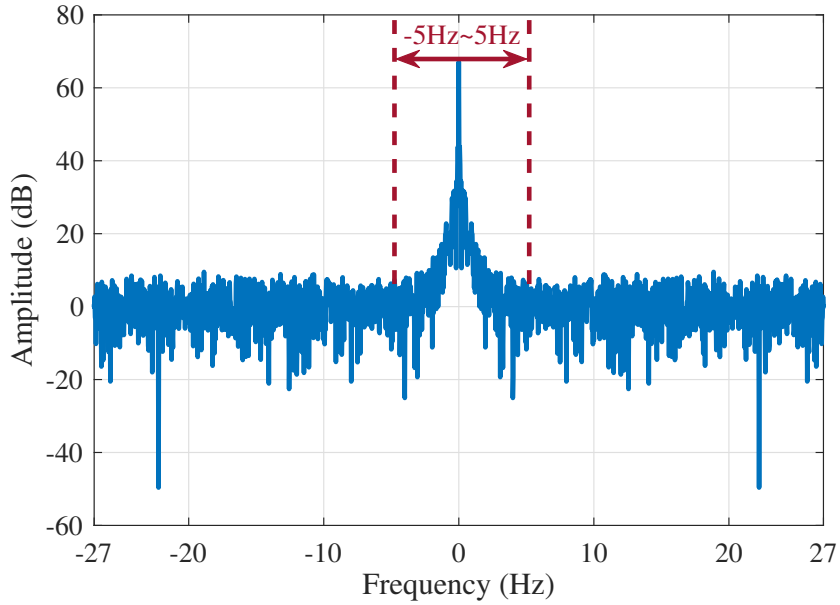


Figure 3.10: Phase difference in the frequency domain.

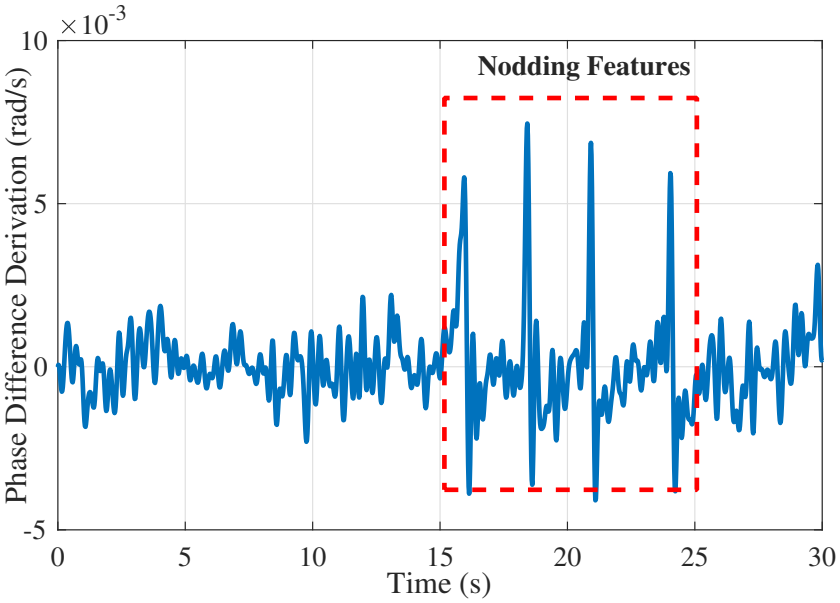


Figure 3.11: Derivative of the filtered phase difference.

is no cumulative error remaining in the signal anymore. The nodding features can be clearly distinguished from other remaining noises.

3.4.3 Driving Fatigue Detection

We utilize an unsupervised LSTM variational autoencoder to learn the nodding features from sampled, calibrated data during driving. After the model is well trained, the input signal can

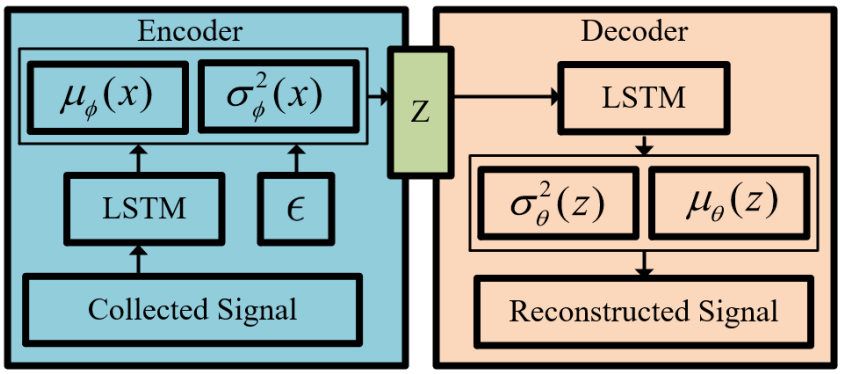


Figure 3.12: The recurrent variational autoencoder for driving fatigue detection.

be well reconstructed by the autoencoder if it is sampled during nodding. Otherwise, the reconstructed signal will contain high distortion. Thus, we can detect nodding by calculating the divergence between the input signal and the reconstructed signal. The details of the training model and divergence calculation are presented in the following.

The Learning Model and Training

The learning model adopted for offline training is composed of an LSTM-based variational autoencoder [8], which is an unsupervised learning model as shown in Fig. 3.12. As we known, drivers could have numerous types of driving movements, which introduce two challenges. First, it is hard to distinguish the nodding movement from all types of other driving movements with a simple threshold-based method. Thus, a learning-based method could be a better choice for nodding detection. Second, training the network with labeled data for all movement types (which could be many) is challenging and costly. To address the problems, we intend to collect and learn the features of nodding instead of learning all kinds of driving movements. In this case, autoencoder is a good choice because, as an effective unsupervised learning algorithm, no labeled data is required for the training process. Furthermore, compared with deep learning models, autoencoder has a simpler model structure and lower complexity, which translate to shorter training time. Thus, we propose the unsupervised LSTM variational autoencoder model for nodding detection, which can effectively reduce the cost of collecting labeled data for various driving movements.

Consider that all the data are sampled as a time sequence, and nodding causes an obvious change of calibrated phase difference, as shown in Fig. 3.11, LSTM is an effective model for capturing the nodding features, because LSTM can better learn the long-range dependency in data than traditional recurrent neural networks. Then, the variational autoencoder model is applied to reconstruct the input signal. The goal is to maximize the marginal likelihood given below.

$$p_\theta(x) = \int p_\theta(x|Z)p(Z)dz, \tag{3.7}$$

where x, θ, and z are the observed variables, the set of parameters, and the latent random variables, respectively; $p(Z)$ is the prior over the latent random variables Z; and $p_\theta(x|Z)$ is the posterior conditional probability, representing an observation model under the parameter set.

Usually $p_\theta(x)$ is hard to estimate because of the integral operation. The computation usually introduces considerable complexity, even though the size of the dataset is small. To reduce the computational cost for training, the autoencoder leverages the variational approximation $q_\phi(Z|x)$, rather than calculating the true conditional probability $p_\theta(x|Z)$. Thus, the autoencoder model has ϕ as encoder to approximate $q_\phi(Z|x)$, and set θ as the parameter for $p_\theta(x|Z)$ in the decoder. The reparametrization technique is implemented in the autoencoder model to mitigate the training overhead. The latent vector Z is computed by the mean vector $\mu_\phi(x)$ and the variance vector $\sigma^2_\phi(x)$ generated by the two linear modules from the LSTM outputs as

$$Z = \mu_\phi(x) + \sigma_\phi(x) \odot \epsilon, \tag{3.8}$$

where ϵ represents a Gaussian noise and $\odot$ represents the element-wise product operation. Based on the latent vector, the variance vector $\sigma^2_\phi(Z)$ and mean vector $\mu_\phi(Z)$ for the reconstructed signal can be decoded from the LSTM network. Eventually, the input signal can be reconstructed as the output of the decoder.

In the proposed LSTM autoencoder network, the dimension of the LSTM layer is set to 20, and 10 units are used in the latent Z layer. The offline training process aims to learn the

features of nodding, so all training data is sampled when the volunteers are nodding their heads. Specifically, the volunteers are sitting in a parked car and nodding their heads randomly when the reader is interrogating the RFID tags attached to the hat. To achieve a high success rate for nodding detection, we collected 3000 nodding samples from three volunteers. We use 2400 samples from the collected data for training, and the other 600 samples for testing.

Online Drowsiness Detection

After offline training, the newly collected signal in realtime can be fed into the autoencoder, and the autoencoder will generate the reconstructed signal. Fig. 3.13 and Fig. 3.14 show the reconstructed signals when the input is a nodding signal and a normal driving signal, respectively. We can see that the nodding signal is well reconstructed by the autoencoder, while the reconstructed normal driving signal is quite different from the input signal. The input signal is flat, while the reconstructed signal has large variations similar to nodding features. This is because our LSTM-autoencoder model has been trained by nodding features. Therefore, the new signals sampled during nodding can be better reconstructed than the signals sampled during other types of head movements (as well as when there are no head movements). Thus, we can detect if the driver is nodding or not, by calculating the divergence between the input signal and the reconstructed signal.

We adopt a sliding window with 2 second duration to extract the input signal from calibrated phase difference, in order to guarantee that all nodding movement can be captured in the window. The divergence is calculated in the form of Mean Absolute Error (MAE), given by

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^r|, \tag{3.9}$$

where n is the total number of samples in the sliding window, y_i is the ith sample of the input signal, and y_i^r is the ith sample of the reconstructed signal. Then we group the MAEs from nodding and normal driving, respectively, and plot all the errors in the form of cumulative distribution function (CDF) in Fig. 3.15. The figure shows that 91.26% MAEs of the nodding signal is lower than the minimum error of the reconstructed normal driving signal, which is

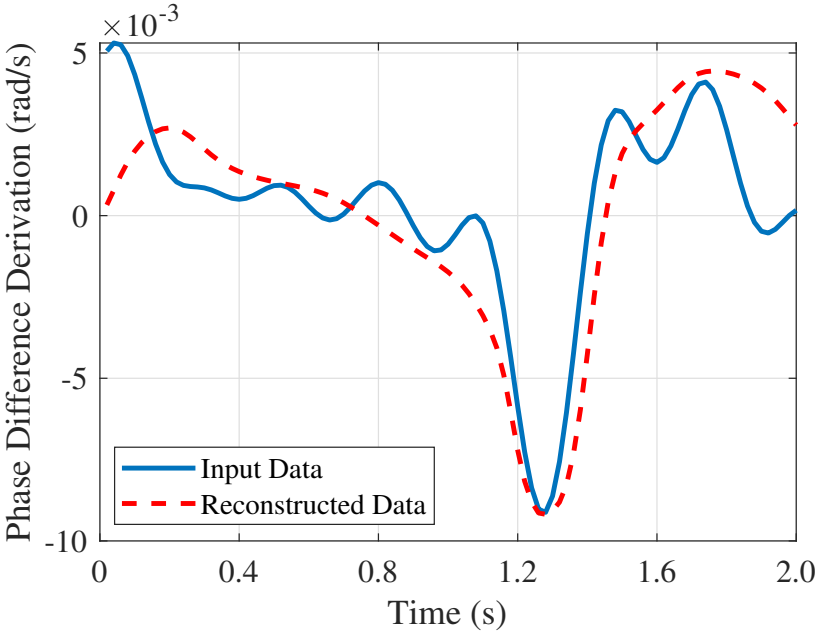


Figure 3.13: The reconstructed signal when the input to the autoencoder is a nodding signal

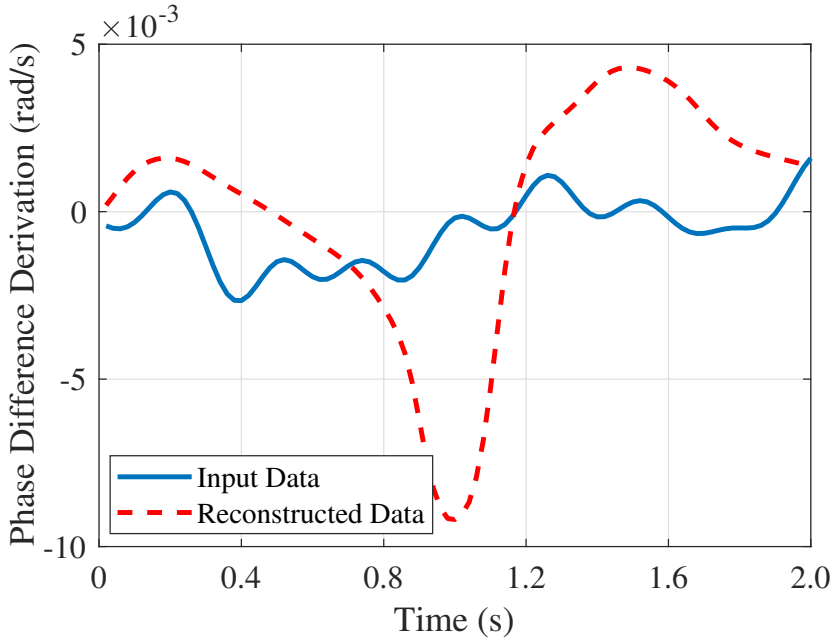


Figure 3.14: The reconstructed signal when the input to the autoencoder is a normal driving signal.

0.21. Thus, we conclude that nodding movement can be effectively distinguished by MAE from other types of head movements.

To further investigate the most suitable threshold of MAE, we test the system accuracy with different MAE thresholds. The True Positive (TP) and True Negative (TR) rates are computed, where the TP rate indicates the accuracy of nodding detection, and the TN rate represents the accuracy of normal driving recognition. Fig. 3.16 shows that the TN rate is 98.8% when

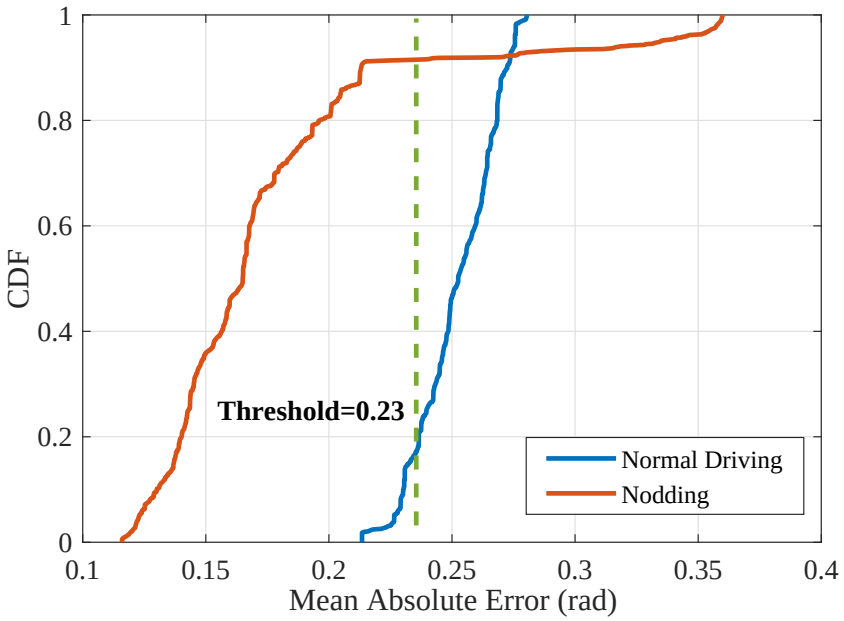


Figure 3.15: CDFs of the mean absolute errors for normal driving and nodding.

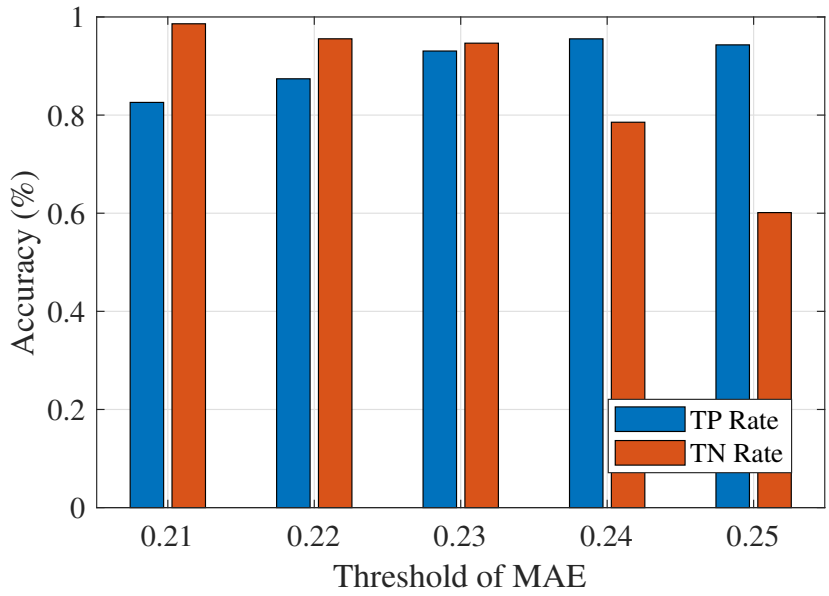


Figure 3.16: Detection accuracy with different MAE thresholds.

the threshold is set to 0.21, but the TP rate is low, which is 82.58%, in this setting. When the threshold is set to 0.25, although the TP rate is as high as 94.31%, the TN rate decreases significantly to 60.03%. This is because, a smaller threshold makes the signal more likely to be considered as normal diving, which makes it harder to recognize the nodding movements. Thus, an appropriate threshold should be set to achieve a tradeoff. Finally, we set the threshold of MAE to 0.23 for the proposed drowsiness detection system.

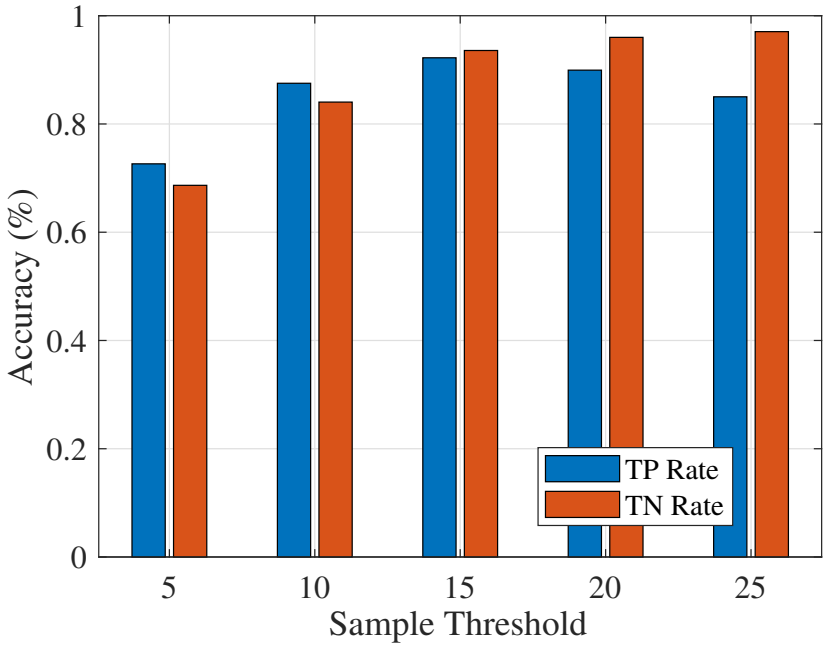


Figure 3.17: Detection accuracy for different values of sampling threshold α.

However, if an alarm is sent whenever a divergence lower than the threshold is detected, there could be many false alarms triggered. This is because the calculated divergence could have fake peaks, which could cause a false alarm. To avoid the effect of the sharp peaks and detect the nodding pattern accurately, we leverage a simple counting algorithm. Only when the divergence is larger the threshold and remains there for over α samples, will the current driver movement be considered as nodding. To investigate the appropriate number of samples for nodding detection, we test the system with different sample threshold values. The results are shown in Fig. 3.17. It can be seen that both the TP rate and TN rate are lower than 80% when the threshold is set to 5 samples. However, the TP rate becomes 85.02% when the threshold is 25. This is because the duration of nodding movement is usually very short, and 25 samples literally means 0.45 second given the 55 Hz sampling frequency, which is too long for nodding detection. Finally, the sample threshold is set to 15 to achieve highest accuracy.

3.5 Experimental Study

3.5.1 Experiment Configuration

To evaluate the proposed driving fatigue detection system, we build a prototype system with commercial RFID devices, and test it in both an emulated environment and real driving environments. The volunteers are required to wear a hat, with two passive RFID tags of the ALE-9470 type are attached to the back side. All tags are scanned by a commodity RFID reader of the Impinj R420 model, which is equipped with a polarized S9028PCR antenna. Following the FCC rules, the reader hops every 0.2 second among 50 channels from 902MHz to 928MHz. Low-level data, such as RSSI, phase, and timestamp will be sampled by the reader and processed in an MSI laptop with an Intel Core i7-6820HK CPU and a Navidia GTX 1080 GPU. One possible limitation of the current prototype system is the relatively higher cost compared with other driving fatigue detection systems, such as smartphone-based system, WiFi-based system, and camera-based system. Fortunately, the overall system cost could be reduced by using cheaper readers. For example, since only one antenna is required in our system, one port reader like Impinj R120 can be used. Further, medium range readers, such as Feig MRU102-PoE, will be another low-cost option, because the interrogate range for car environment monitoring is not demanding. Finally, the cost of the future commercial system could be further reduced, if customized readers are used and mass produced.

The system is firstly tested in an emulated environment, which is in a 8.8m $\times$ 4.5m laboratory. The volunteer is seated on a chair and nod or rotate his/her head naturally when the reader is scanning the RFID tags. The antenna is placed on a shelf behind the volunteer. All data sampled will be transmitted to the laptop for data calibration and nodding detection. The system is also evaluated in real driving environments. Specifically, the system is deployed in a BMW 328i vehicle made in 2014, as shown in Fig. 3.18. The polarized antenna is placed on the back of the driver chair to continuously interrogate the two tags, which are attached to the back side of the hat. The driver is required to drive naturally in different scenarios, such as on a highway, in a parking lot, and in city streets, all the nodding movements happens during driving are recorded as the ground truth.

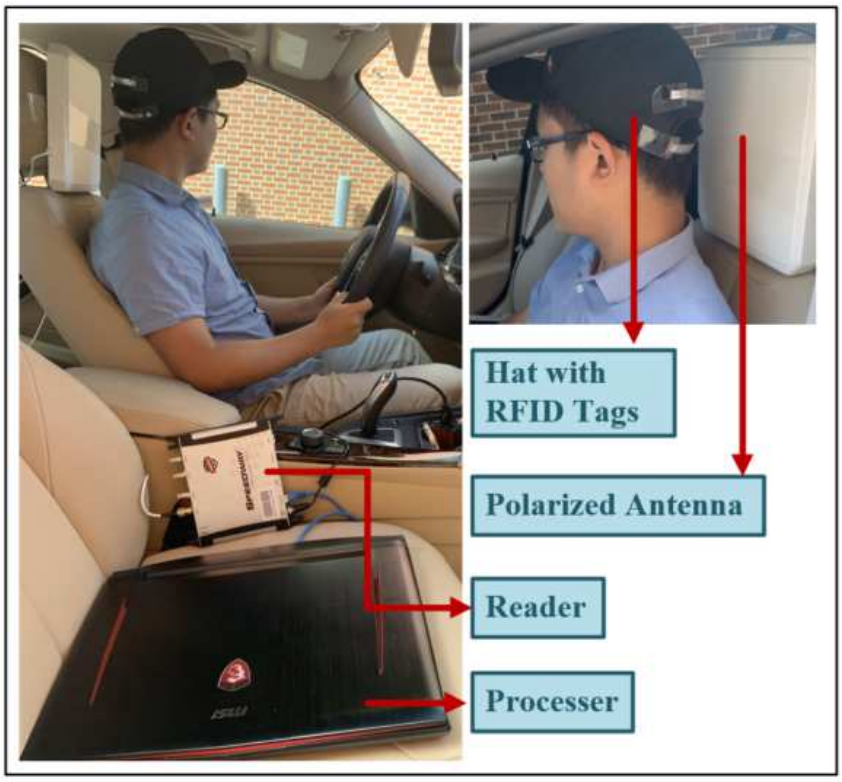


Figure 3.18: The system setup in a BMW 328i car in our experiments.

3.5.2 Results and Discussions

Overall Performance

Experimental results of our drowsiness detection system are presented in Fig. 3.19. The figure shows the TP and TN rates in two different scenarios in the emulated environment. Recall that the TP rate means the accuracy of nodding detection, and the TN rate is the accuracy of normal driving recognition. In the first scenario, the volunteer is asked not to rotate his/her head, but only nod occasionally during the test. In the other scenario, the volunteer can move his/her head and body casually (i.e., to generate large interference). The results show that our system can achieve a 97.23% TP rate and a 96.72% TN rate when no other head movements present. The achieved TP rate and TN rate are 91.48% and 95.38%, respectively, even though the drivers rotate or shift their heads during the experiment. The high detection accuracy in different scenarios proves that our system can effectively mitigate the influence of head rotation and shifting during driving, as well as the other large noises in the driving environment.

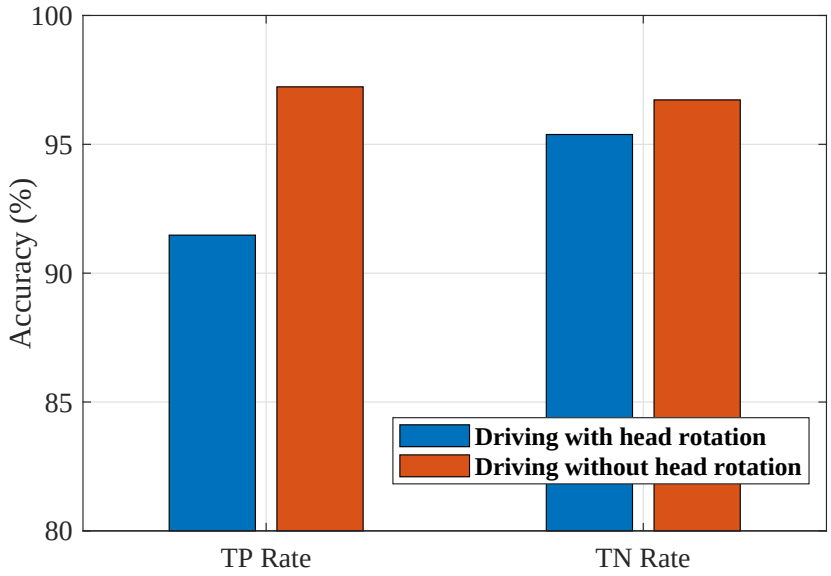


Figure 3.19: Detection accuracy in two scenarios: TP rates and TN rates.

Fig. 3.20 shows the accuracy of our system for all the volunteers involved in the experiments. The TP rate is 95.08% and the TN rate is 92.61% for the the 5th driver, which is the highest among them. In contrast, the accuracy of nodding detection for the second driver is the lowest. The difference in detection accuracy is mainly due to the different driving and nodding movements of different drivers. In addition, the height of the driver also affects the relative location between the antenna and the tags, and thus the sensitivity of nodding feature extraction. Fortunately, we find that all the TP rates and TR rates are higher than 90.33% and 89.01%. respectively. We can thus conclude that the accuracy of nodding detection is sufficiently high and robust for different drivers.

To compare the system accuracy with different existing driving fatigue systems, we summarize the average driving fatigue detection accuracy values that are provided in the related papers [49–51] in Table 3.1. From the table, we can observe that the average detection accuracy of the proposed RFID based system is 92.8%, which is sufficiently higher than the vision-based and WiFi-based systems. The average accuracy of our system is only 0.5% lower than the acoustic-based system, which is implemented with a smartphone. However, as a broadcast signal, both WiFi signal and acoustic signal are sensitive to the movement of passengers, especially for the passenger next to the driver. The acoustic approach could also be interfered

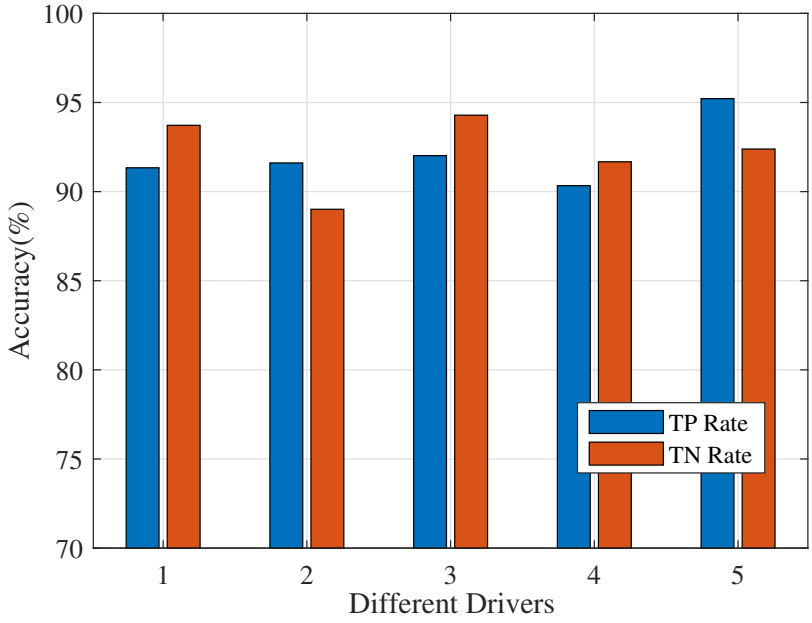


Figure 3.20: Detection accuracy for different drivers.

Table 3.1: Average Detection Accuracy Comparison of Different Driving Fatigue Detection Systems

System	*Average Detection Accuracy*
Video Camera [49]	88.9%
WiFi Device [50]	89.6%
Smartphone (Acoustic-based) [51]	93.3%
RFID Tags and Reader	92.8%

surrounding noise in the same frequency when driving in traffic. Thus, we conclude that the RFID based system is more suitable for the noisy driving environment.

Impact of Model Parameter

Since the nodding features are carried in a data sequence rather than single samples of data, a sliding window is utilized to extract the data sequence. To investigate the suitable size for the sliding window, we test the system performance under different window sizes ranging from 0.5 s to 5 s. Fig. 3.21 represents the TP and TN rates when the data is trained with different window sizes. It can be observed that, when the window size is larger than 3.5 s, the TN rate and TP rate are lower than 85.24% and 74.64%, respectively. When the window size is smaller than 1.5 s, the TN rate is lower than 83.51% and the TP rate is lower than 79.67%. The observations show that the system performance is sensitive to the sliding window size. The highest detection

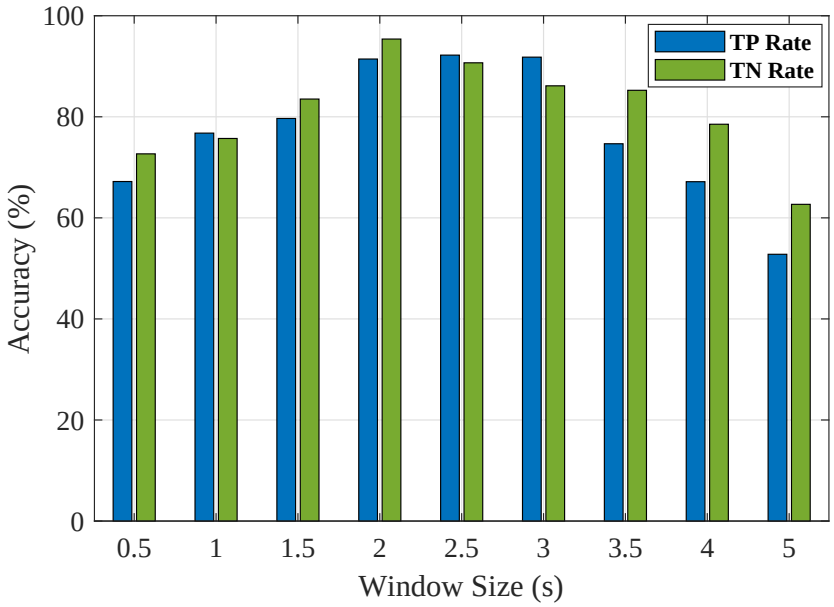


Figure 3.21: Impact of the window size used for training.

accuracy requires a suitably set sliding window size for learning and detection. Based on the result shown in the figure, the sliding window of our system is set to 2 s for the highest accuracy.

Different Driving Scenarios

To investigate the influence of vehicle vibration on the detection accuracy, we evaluate the system in three different scenarios, including (i) driving on a highway, (ii) driving in city streets, and (iii) parked. Fig. 3.22 shows that the system can achieve the highest accuracy of a 96.78% TP rate and a 95.78% TN rate when the vehicle is parked. For highway driving, the TP rate and TN rate are 93.67% and 94.66%, receptively, which means the influence of the vehicle vibration is negligible in this case. This is because the vibration generates similar variation on the phase data for both tags, which can be effectively mitigated by using the phase difference.

However, for the city street driving scenario, the accuracy decreases obviously to a 87.47%TN rate and a 84.75%TP rate. We also present the corresponding false alarm rate in different driving scenarios in Fig. 3.23. The figure shows that the false alarm ratio is lower than 5.34% when driving on the highway and when parked; but the false alarm ratio increases to 12.53% in the in-town driving scenario. It can be concluded that the detection accuracy is considerably degraded by the more extensive driving movements for in-town driving (e.g., checking for traffic conditions, finding directions, and turning the steering wheel for turning at street corners). In

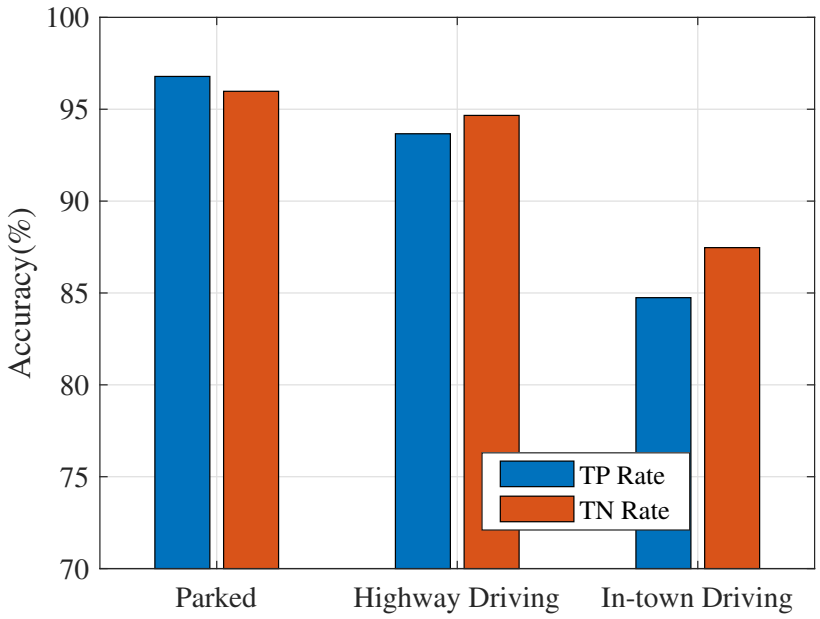


Figure 3.22: Accuracy in different driving scenarios.

addition, when the vehicles stops and restarts at traffic light or stop signs, the driver may also nod his/her head because of inertia and acceleration. These movements generate large interference and cannot be effectively distinguished from the normal nodding due to drowsiness. Fortunately, drowsy driving usually does not happen in the city street driving case, where the drivers turning wheels or stop/restart the vehicle frequently. The high accuracy in the different real driving scenarios has proved that the system can effectively detect driving drowsiness, especially when driving on highways.

Impact of Passengers

Finally, Fig. 3.24 shows the impact of various numbers of passengers in the car, whose movements also generate interference to driver nodding detection. The experiments are conducted to evaluate if the system can still achieve high accuracy under the interference from an unstable testing environment. As an in-car RF-based sensing system, the performance could be affected by the surroundings because of the multipath effect. A noisy testing environment could introduce considerable variation to the wireless channel, which affects the sampled phase value at the RFID reader. Since the movements of other passengers mainly contribute to the environmental noise in the driving environment, more passengers could generate larger interference in

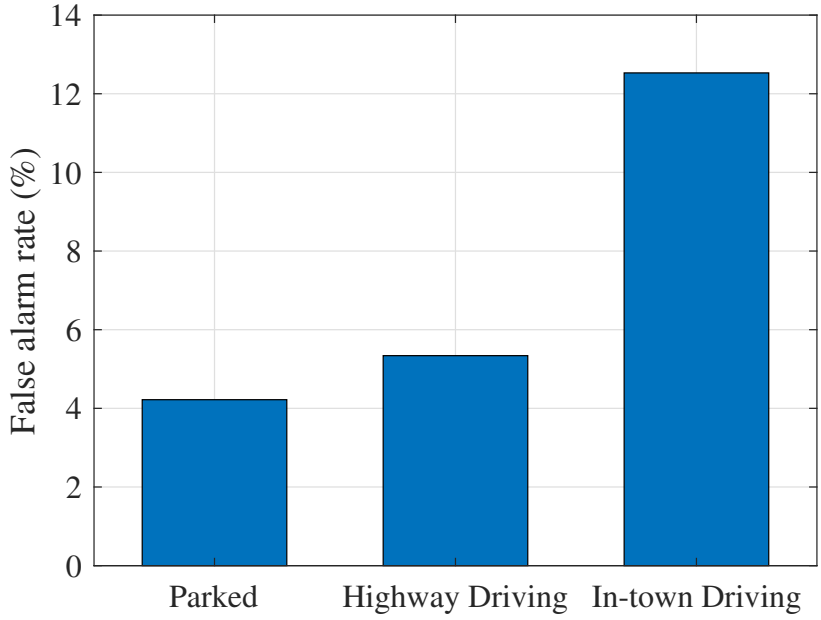


Figure 3.23: False alarm rate in different driving scenarios.

the system. Thus, we test the system with different number of passengers, increased from 1 to 4.

In the experiments, the passengers move naturally in the vehicle. We can see that the system achieves the highest TN rate when there is only one passenger in the car. With more passengers, the system can still achieve high accuracy such as a 91.75% TP rate and a 91.99% TN rate. The false alarm ratios shown in Fig. 3.25 are all lower than 8.01% even when four passengers are in the vehicle. The results verify that the influence of passengers can be effectively mitigated by the proposed system. This is because the range of the polarized antenna is limited, so the multipath effect is limited for transmissions between the tags and reader. Thus, the movements of passengers can hardly affect the phase information sampled by the reader, and drowsiness detection is robust even when multiple passengers are loaded in the vehicle.

3.6 Conclusions

In this paper, we proposed a driving drowsiness detection system by detecting the nodding movements of drivers. The nodding movements were detected by using the received phase values in RFID tag responses. We proposed a tag deployment scheme to effective deal with the high noisy driving environment. To mitigate the influence of vehicle vibration during driving,

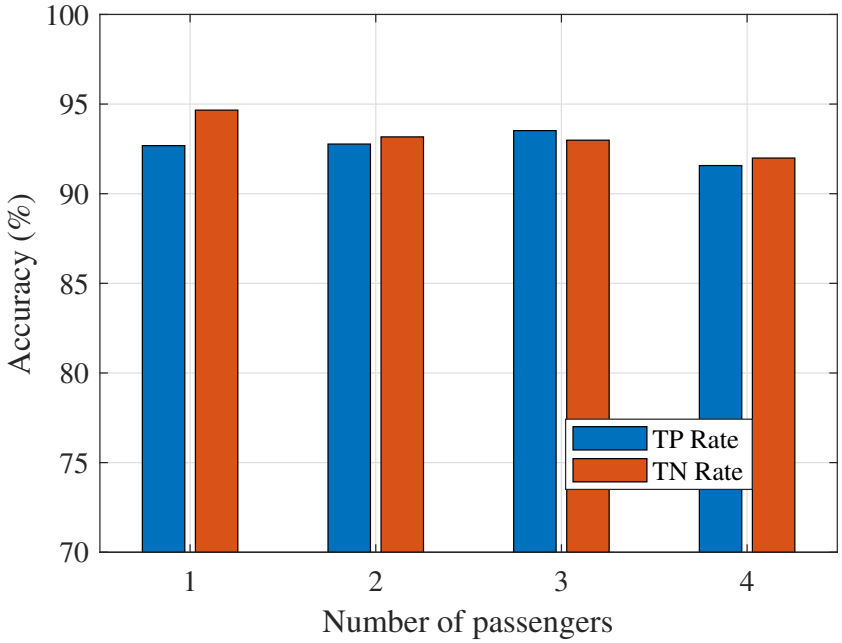


Figure 3.24: Accuracy with different number of passengers in the vehicle.

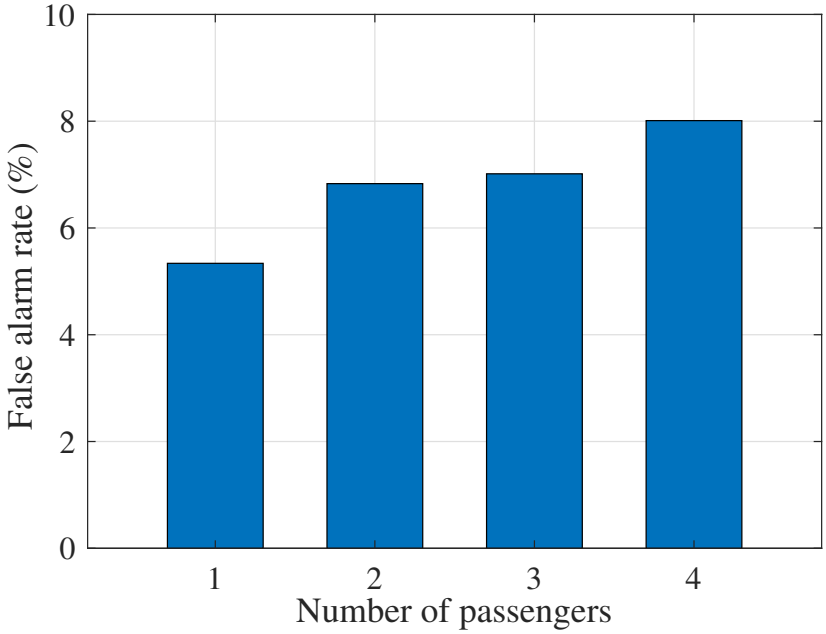


Figure 3.25: False alarm rate with different number of passengers in the vehicle.

the phase difference of the two tags were estimated by a proposed algorithm. An unsupervised LSTM autoencoder model was incorporated to learn the nodding features, which can effectively reduce the cost of collecting labeled data for various types of driving movements. The high detection accuracy of the proposed system was demonstrated by experiments in both an emulated environment and real driving scenarios with commercial tags and readers.

Chapter 4

Respiration Monitoring with RFID in Driving Environments

4.1 Introduction

Vehicles play a critical role in our society. With the drastic increase of the number of vehicles as well as driving time, driving safety has become ever more important than before. Driving fatigue is one of the primary causes of car accidents, which takes many lives every year [75]. It has been reported that there are on average 6 million car accidents in the U.S. every year [76], and more than 90 lives are lost every day. Such car accidents could be effectively avoided and human lives could be saved, if drivers are warned when they become sleepy.

Drowsy driving can be indicated by multiple features of the driver, such as eyelid movements, driving movements, and human vital signs. Among various types of vital signs, respiration rate is a useful indicator of driver's fatigue state. It has been shown that the breathing rate usually decreases notably (i.e., for about 3 breaths per minute (bpm)) before the driver falls asleep [65]. Thus, accurately monitoring the respiration of the driver is a promising way, and the first step, to prevent drowsy driving.

However, respiration monitoring for drivers is challenging, due to the highly strong noises in the driving environment, such as vehicle vibration and movements of the driver and passengers. Several techniques have been proposed for respiration monitoring using different types of signals, including video, ultra sound, and RF signals. Vision based methods detect respiration by analyzing the chest movements captured by a video camera [77], but it may not work well when lighting is poor (e.g., driving at night) and may raise privacy concerns. RF signals, such as WiFi [50] and UWB radar [66], have also been exploited, with the advantage of not requiring

sufficient lighting inside the vehicle. However, due to the mutipath effect, such non-invasive RF sensors may be easily affected by the movements of driver and passengers. Recently, ultra sound signals generated by smartphones have been considered in [78]. Acoustic signals are shown to be effective to capture the human breathing signal [11], but as RF signals, they are also sensitive to the strong noises in the driving environment.

To this end, radio frequency identification (RFID) provides a promising alternative solution. RFID based sensing has become a hot problem area recently. Unlike other contact-free sensors, RFID tags are much cheaper and can be easily attached to the target object. As a near-field communication technology, it is more robust to surrounding noises. However, to exploit commodity RFID (rather than customized hardware) for respiration monitoring, many challenges should be addressed. For example, frequency hopping, as required by the Federal Communications Commission (FCC), causes large phase offsets, making received phase data useless. With the Slotted ALOHA medium access control protocol, the tags are sampled randomly, and it is common that the readings from the same tag are sparse (many are missing). There have been several recent works employing RFID for respiration monitoring [8, 27, 58], but *all the existing schemes work in a static indoor environment*. The strong interference from the driving environment prevents their application for driver respiration monitoring.

In this paper, we address the above challenges with novel, effective solutions, and propose an RFID based respiration monitoring system for the driving environment. Specifically, we propose to attach multiple tags to the driver's seat belt, which allow us to exploit the *tag diversity*. Although the same respiration signal is sampled by all the tags, the reader collected data is of high diversity as resulted from the multiple independent sampling. We also develop an effective tensor completion technique to mitigate the effect of frequency hopping and random sampling. The recovered phase difference by tensor completion helps to combat the effect of vehicle vibration and driver/passenger body movements. Finally we apply tensor Canonical Polyadic Decomposition (CPD) to separate the small respiration signal from strong noises. The idea of increasing the dimension of data is essential to detect the weak respiration signal from noisy and sparse samples.

Specifically, we first provide an analysis of the sampled phase data that is used in the proposed system, and investigate the impact of frequency hopping offset as well as the challenges of respiration monitoring in driving environments. The proposed system is designed with several novel components to address these challenges, including data collection, data preprocessing, CPD, and respiration signal estimation. With multiple tags attached to the seat belt, the breathing signal can be effectively embedded in the phase data captured by the reader. However, the sampled phases are sparse due to random sampling, and are greatly distorted by channel hopping, vehicle vibration, and body movements. To combat such noises, we propose a High Accuracy Low Rank Tensor Completion (HaLRTC) based technique to estimate the phase difference between each pair of tags in each time slot, and to eliminate the frequency hopping effect simultaneously [146]. To extract a clean breathing signal, we leverage a CPD based technique to decompose the tensor data constructed by the previous estimated phase difference. Finally, the breathing signal is recovered from decomposed tensor data, and the breathing rate can be estimated with a peak detection algorithm.

The main contributions of this paper are summarized below.

- To the best of our knowledge, this is the first work on respiration monitoring in driving environments using commodity RFID reader and tags.
- We propose a tensor completion technique to recover missing readings in collected phase data, and a tensor decomposition approach to extract the respiration signal of the driver from phase values sampled from multiple RFID tags. The proposed techniques are effective in combating the strong noises caused by frequency hopping, random sampling, vehicle vibration, and other movements in the driving movement.
- We develop a prototype system with commodity RFID devices and test the system in real driving environments. Extensive experiments are conducted to evaluate the system performance in various driving scenarios, such as parked, in city streets, and on a highway, and system configurations, where a highly accurate and robust performance is demonstrated.

In the following, we review related work in Section 4.2 and present the preliminaries and system overview in Section 4.3. The detailed system design is introduced and analyzed in Section 4.4. We present our experimental performance evaluation in Section 4.5 and conclude this paper in Section 4.6.

4.2 Related Work

This work is closely related to RF based vital sign monitoring and RFID based sensing. We mainly review these two classes of systems in this section.

RF based health sensing systems have been developed that employ Radar, WiFi, and RFID techniques. Radar based vital sign monitoring systems include frequency modulated continuous wave (FMCW) radar [80] and Doppler Radar [81]. However, they usually require customized hardware and operate over a wide spectrum. WiFi based systems mainly use received signal strength (RSS) and channel state information (CSI). For example, UbiBreathe [82] and mmVital [83] utilize WiFi RSS at 2.4 GHz and 60 GHz, respectively. To improve accuracy, CSI based systems leverage the amplitude or phase difference information of CSI for estimating single or multiple persons' breathing and heart rates [17, 19, 84, 85]. Moreover, several bimodal CSI data based systems have been proposed to tackle the weak breathing signals at some special positions [86, 87, 104].

Recently, several RFID based breathing monitoring systems have been proposed. For example, RFID tags have been used for breathing rate estimation in [27], breathing and heart rates estimation in [88], and breathing monitoring and sleeping posture recognition in [74]. Furthermore, the RF-ECG system is proposed for heart rate variability assessment using an RFID tag array [89]. To mitigate the frequency hopping offset in FCC-compliant RFID systems, the AutoTag system is proposed for breathing monitoring and apnea detection with a variational autoencoder [8, 58]. However, these existing systems are designed for the indoor, static environment; they may not be effective in the highly dynamic, highly noisy driving environment.

Recently, WiFi based [50], acoustic based [78], and UWB based [66] systems have been developed for breath monitoring in driving environments. In fact, these existing systems are

sensitive to the environmental interference, such as the body movements of the driver himself/herself and of the passengers, due to their relatively large transmission ranges.

In addition to vital sign monitoring, RFID tags have also been applied for many other applications, such as indoor localization [22, 57], user authentication [68], material identification [69], object orientation estimation [23], vibration sensing [70], anomaly detection [71], and drone localization and navigation [53, 90]. To overcome the low accuracy when RSS values are used [31], recent works are mainly focused on the phase for indoor localization, which can be used to derive the distance and direction of arrival (DOA). To solve the phase ambiguity problem, synthetic aperture radar (SAR) [21] and the hologram techniques [22, 34] are proposed. The RFind system estimates time-of-flight with a special hardware to achieve high localization precision [91].

This work, to the best of our knowledge, is the first to apply RFID based sensing for monitoring breathing signals in a driving environment. The proposed system consists of novel and effective solutions for noise and movement interference removal and breathing signal separation. The tensor based approach in this work has been analyzed and proven to be effective for the noisy driving environment.

4.3 Challenges and System Overview

4.3.1 Phase of the RFID Signal

To detect the breathing signal from received phase values from multiple RFID tags, we should firstly know how these phase samples are collected by the RFID reader. In RFID systems, phase information is one type of low level data, which is collected when the RFID reader receives the Electronic Product Code (EPC) from interrogated tags. When the multipath effect is negligible, the measured phase sample can be written as [38]

$$\varphi = \mathrm{mod}\left(2\pi(2D)\cdot f/c + \varphi_{tag} + \varphi_T + \varphi_R, 2\pi\right), \tag{4.1}$$

where D is the distance between the sampled RFID tag and the reader antenna, f is the frequency of the currently occupied channel, and c represents the speed of light. Moreover,

φ_{tag}, φ_T, and φ_R are the additional phase rotation for the tag, the transmitter, and the receiver, respectively. These additional phase rotations are mainly caused by the circuits in the reader and tag hardware. Furthermore, different reflection characteristics of these devices also contribute to the phase distortion.

This model implies that we could use phase values to detect the changes in the distance between the antenna and the tag. When the tag is attached to the human chest, the phase changes are indicative of the breathing signal. However, in addition to D, the other parameters, i.e., f, φ_{tag}, φ_T, and φ_R, are all susceptible to the current channel used by the reader. Thus, the measured phase value will have a different phase offset when the system hops to a new channel. Following FCC regulations, the Ultra-High Frequency (HUF) RFID system should hop among 50 channels within 10 seconds, so the sampled phase data will be heavily corrupted by frequency hopping. *The frequency hopping effect poses a big challenge for extracting the respiration signal*.

4.3.2 Respiration Monitoring in Driving Environments

To capture the respiration of a driver, we attach RFID tags on the seat belt, as illustrated in Fig. 4.1. Since the seat belt is bonded on the driver's body, it (and the tags) moves along with the rise and fall of the chest. Such movements are carried in the sampled phase values of each tag (see (4.1)), which will be captured by the RFID reader placed on the ceiling above the driver seat. *A big challenge is that, the seat belt/tag movements are no only caused by breathing, but also affected by other environmental factors in the car.*

Fig. 4.1 presents the sampled phase signal from a single RFID tag collected in different scenarios. For better illustration, the plotted phase data has already been preprocessed by removing the frequency hopping effect (to be discussed in Section 4.4.1). Theoretically, without the influence of channel hopping, the phase samples should exhibit the periodicity of human respiration. In fact, the collected phase signal, when the car is parked, exhibits strong periodicity. However, the sampled phase values in a moving vehicle, as shown in the second subplot in Fig. 4.1, are highly random, i.e., far from a periodic signal. This is because that respiration

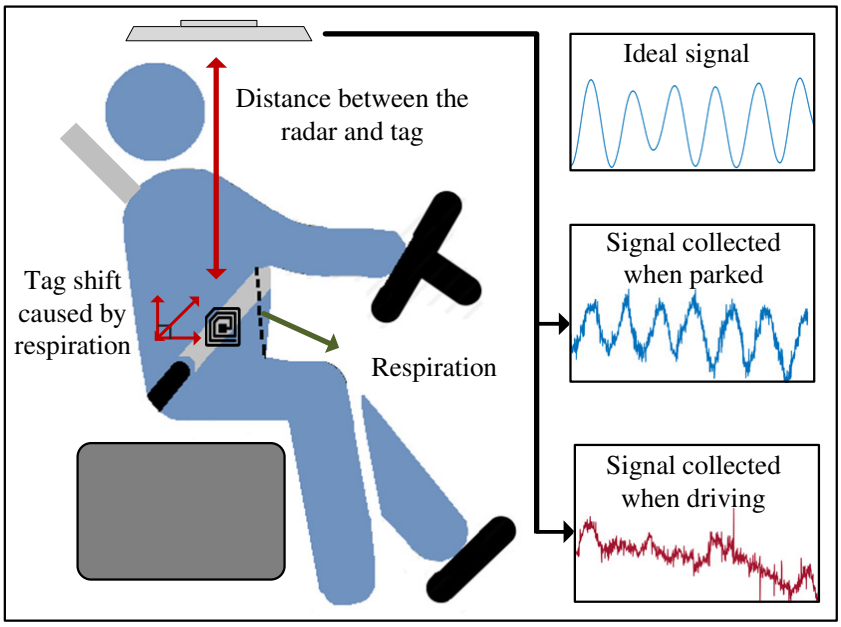


Figure 4.1: Illustration of the respiration monitoring mechanism.

monitoring in a driving environment is very different from other respiration monitoring scenarios, in which the user is usually in a relative stable state, such as sleeping or sitting [8,18,19,58]. This assumption does not hold true in driving environments. First, the vehicle vibrates when moving fast on the road, which makes the seat belt vibrate along with the car. Second, drivers do not remain completely still. For example, the arms move when the driver turns the steering wheel, and the head moves when the drive looks around to check traffic conditions, which will cause additional movements to the tags. Such environmental movements also cause time-varying multipath interference by reflecting the RFID signal. Accordingly, *how to mitigate the impact of vehicle vibration and body movements poses another challenge (in addition to frequency hopping), to be addressed in our system.*

4.3.3 System Architecture Overview

To overcome the above challenges, we develop a tensor based breath monitoring system with an architecture shown in Fig. 4.2. The proposed system consists of four modules, including data collection, breathing data preprocessing, tensor decomposition, and respiration signal estimation.

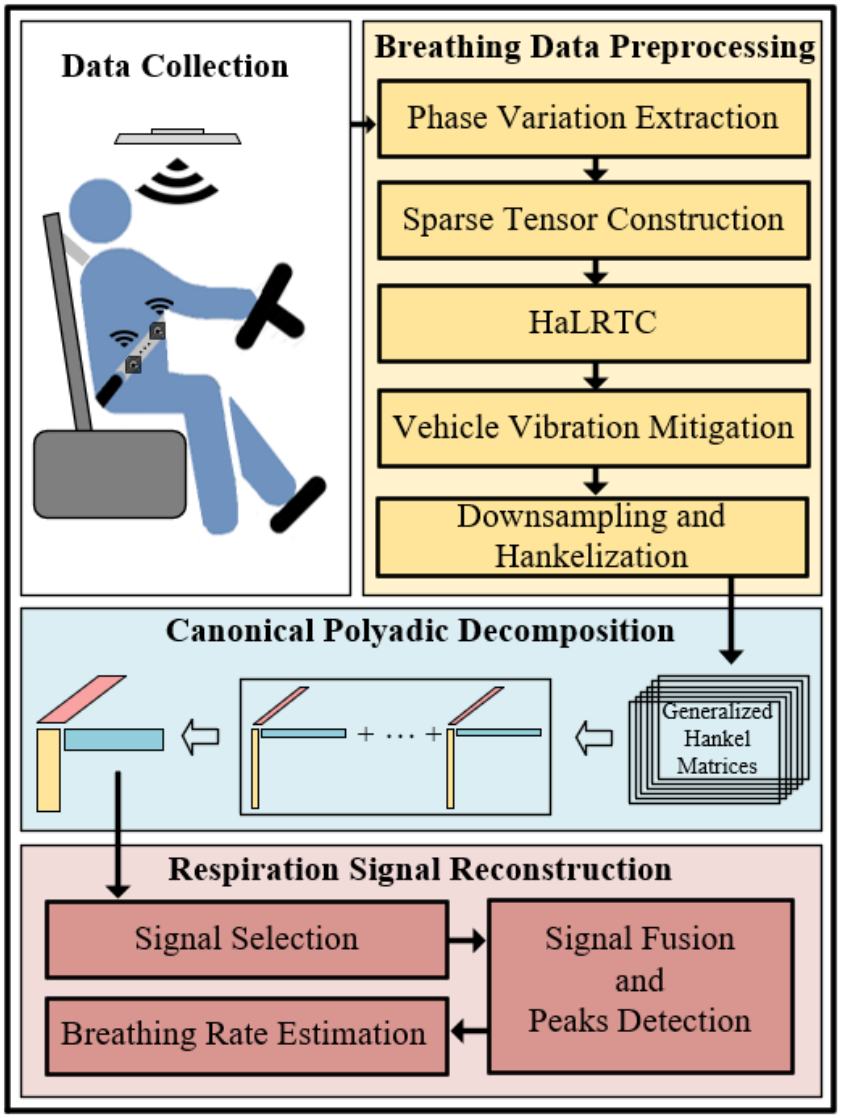


Figure 4.2: Architecture of the proposed system.

In the *data collection module*, the reader keeps on interrogating the RFID tags attached to the seat belt and collecting phase samples from each tag. The *data preprocessing module* is to remove the channel hopping offset and to mitigate the noise from the driving environment. Since the raw phase data is corrupted by frequency hopping, it cannot be directly employed for breathing signal extraction. Fortunately, the phase variation within each individual channel is not affected by channel hopping; this fact is leveraged in our system to recover the phase data for a certain channel (see next section). Vehicle vibration mitigation is accomplished by calculating the recovered phase difference between a pair of tags. Since the vibration has a similar impact on the two tags, subtracting the phase data from the two tags can effectively

mitigate the vehicle vibration noise. However, the challenge is that, the phase data is not sampled simultaneously from all the tags; there is only one phase reading from one of the tags in each time slot. To calculate the phase difference between two tags in each time slot, the missing phase sample(s) in the same time slot should be accurately estimated. To this end, we leverage a compressed sensing technique named HaLRTC to estimate the missing phase data [146], so that we can calculate accurate phase difference for each pair of tags.

Next, we hankelize all calculated phase difference values to construct a tensor for the following tensor decomposition module. After the phase difference tensor is constructed, we leverage CPD to extract the components related to driver's respiration. Finally, we recover the respiration signal of the driver by fusing all breathing related components and estimate the breathing rate with a peak detection algorithm. The system design will be elaborated and analyzed in the next section.

4.4 System Design and Analysis

4.4.1 Combating Frequency Hopping Offset

The FCC regulation requires frequency hopping for UHF RFID systems. The phase offset generated by frequency hopping should be firstly removed in signal preprocessing. We rewrite (4.1) for the phases sampled from the 50 channels as

$$\varphi = \text{mod}\left(4\pi D f_K/c + \varphi_K, 2\pi\right), \quad K = 1, 2, ..., 50, \tag{4.2}$$

where K is the channel index, f_K represents the frequency of channel K, and φ_K is the sum of φ_{tag}, φ_T and φ_R, because all these values are irrelevant to D but are affected by frequency hopping. From (4.2), we can see that the phase offset due to channel hopping is actually caused by f_K and φ_K.

Several techniques have been proposed to remove the effect of these two factors. For example, in AutoTag [8], the phase offset between two adjacent channels is estimated by the difference between the mean value near the end of the previous channel and that at the beginning of the next channel. This technique achieves a very good performance when the number

of interrogated RFID tags is no more than 3. However, as the number of tags is increased, it would be hard to guarantee that enough samples can be collected from all the tags, which are needed for calculating the mean values. In addition, Tagyro leverages a full channel scan and calibration to measure φ_K in each channel [23]. The technique is suitable for RFID systems with more tags. In a driving environment, however, the estimated φ_K could be greatly affected by vehicle vibration and driver's body movement.

Since our system needs to interrogate at least 4 tags for tag diversity (see Section 4.4.2), we propose a novel approach to remove the channel hopping effect. If we subtract the sampled phase φ_n from the previous sampled phase data φ_{n-1}, as given in (4.2), we can obtain the nth *phase variation* on channel K as

$$v_n = 4\pi f_K(D_n - D_{n-1})/c, \tag{4.3}$$

where v_n represents the phase variation between the current and the previous phase sample, D_n is the antenna-tag distance in the nth sample. We find that the phase variations on the same channel are not related to the initial phase offset φ_K. With (4.3), we can easily translate the variations from all channels to a reference channel R by multiplying a factor f_R/f_K. In fact, the reference channel can be any one of the 50 channels. In our system, we set $R = 1$ to make channel 1 the reference channel for convenience.

However, when the nth sample and the $(n-1)$th sample are not from the same channel, Eq. (4.3) should be updated as

$$v_n = 4\pi(f_K D_n - f_{K-1} D_{n-1})/c + \varphi_K - \varphi_{K-1}. \tag{4.4}$$

Eq. (4.4) shows that when two adjacent phase samples are not collected from the same channel, the phase variation is still affected by the initial phase offsets of the two different channels. To mitigate the frequency hopping offset, we should drop these distorted phase variations. Fortunately, since only a single phase variation is affected once the system hops to a new channel, most of the remaining data are still usable. Such results are illustrated in Fig. 4.3.

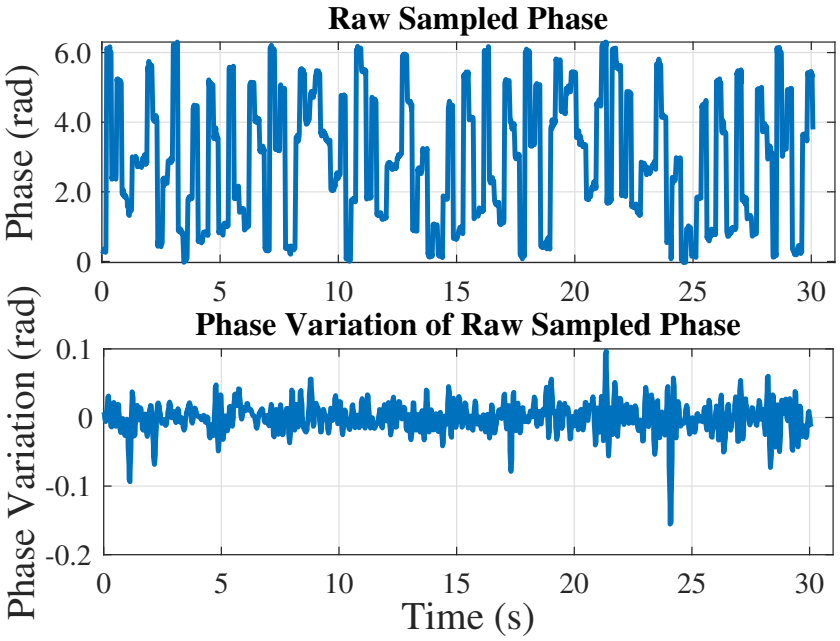


Figure 4.3: Raw phase and the filtered phase variation signal.

We calculate the phase variation from the raw, sampled phase data, and delete the distorted phase variations when hopping to a new channel. It can be observed that although the phase is corrupted by frequency hopping (the upper plot), the phase variation is confined within [-0.15 rad, 0.1 rad]. This result shows that most of the phase variations are not affected by channel hopping.

4.4.2 Recovering Phase for Each Time Slot

To combat vehicle vibration, we propose to leverage the phase difference between two tags. However, following the Gen2 protocol [38], when one tag is sending its EPC to the reader, all other tags should remain silent to avoid collision. The random sampling details are illustrated in Fig. 4.4, which shows the phase data sampled by the reader from multiple tags. A colored square indicates that a valid phase value is sampled from the tag, a blank square means the tag is not sampled in that time slot, and all the same colored squares are sampled from the same channel. Because the entire transmitting process is based on slotted ALOHA, the tags are sampled randomly. That is, phase can be collected from only one tag in each time slot, and the sampling interval for each tag is random.

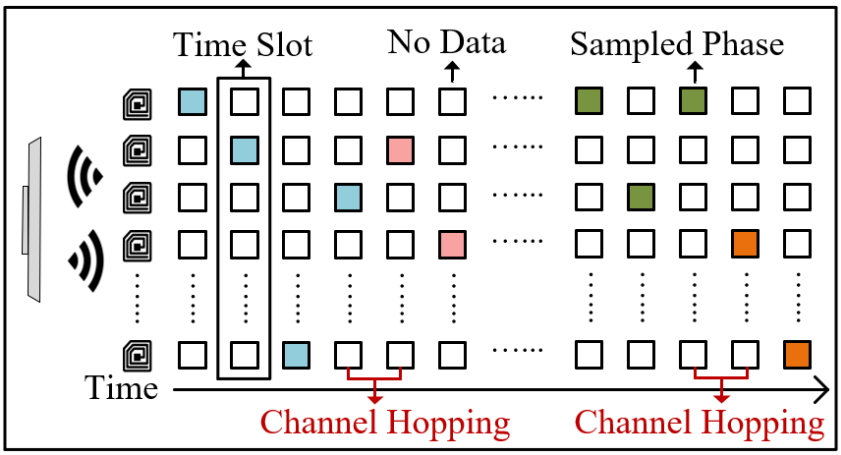


Figure 4.4: Slotted ALOHA based random sampling in RFID systems.

Because calculating phase difference between two tags requires to sample phase data from both tags in the same time slot, we need to estimate the missing phase data of all tags in every time slot, i.e., the blank squares in Fig. 4.4. Several compressed sensing approaches have been shown to achieve a good performance on recovering the missing RFID phase data [92, 93], where the number of deployed tags are less than 3. For example, in [92], a missing phase is estimated from neighboring sampled phase values with a Blackman window, where a pair of tags are used. However, these existing techniques are not suitable for our system, where more tags are deployed. When there are more tags, the sparsity of data becomes higher, thus resulting in a lower recovering accuracy.

Tensor completion is another powerful tool of compressed sensing, which has been adopted to recover missing data in RFID systems in our recent work [52]. We propose a tensor completion based method to recover the phase sequences for all tags in every time slot. Rather than directly recovering the phase sequence as in [52], we recover the sequence of *phase variation* for each tag, and then calculate phase by integrating the recovered phase variation. This is motivated by the fact that phase variation is not affected by frequency hopping, as proven in Section 4.4.1.

We first define the *ideal matrix* we aim to recover, given by

$$V_T \doteq \begin{bmatrix} v_{1t_1} & v_{1t_2} & v_{1t_3} & \dots & v_{1t_n} \\ v_{2t_1} & v_{2t_2} & v_{2t_3} & \dots & v_{2t_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{mt_1} & v_{mt_2} & v_{mt_3} & \dots & v_{mt_n} \end{bmatrix}, \tag{4.5}$$

where m is the tag index, t_n is the nth time slot, and v_{mt_n} is the real phase variation of tag m at time t_n. Due to slotted ALOHA, only one value in each column of V_T can be sampled. The goal is to estimate a matrix $\hat{V}_T$, such that $\hat{V}_T \approx V_T$, based on collected sparse samples. We first build a *sampled matrix* $\bar{V}_T$, which is of the same size as V_T, but all the missing data elements are set to 0. We then filter out thermal noise from the phase variation signal using a low-pass filter with a 15 Hz cutoff frequency, and map the signal to the sampled elements in $\bar{V}_T$. Moreover, all the distorted phase variations in $\bar{V}_T$ that satisfy (4.4), are also set to 0 (i.e., dropped) to avoid the influence of frequency hopping.

In the proposed system, 4 to 8 tags are attached to the seat belt and the frequency of time slots is about 220 Hz, which means the sparsity of $\bar{V}_T$ is higher than 75%, and the number of columns is much larger than the number of rows. Because of the high sparsity and the limited number of rows, the traditional singular value decomposition (SVD) based matrix completion method would not be effective for such $\bar{V}_T$, Therefore, we transform the data into a tensor form and apply tensor completion to estimate the missing data. Specifically, we reshape each row into a *generalized* Hankel matrix as the frontal slice of the tensor, which has the same format of Hankel matrix but is not square. The sparse matrix $\bar{V}_T$ is thus transformed into a tensor, given by

$$\bar{\mathcal{V}}_{(:,:,m)} \doteq \begin{bmatrix} v_{mt_1} & v_{mt_2} & \dots & v_{mt_{(n-r+1)}} \\ v_{mt_2} & v_{mt_3} & \dots & v_{mt_{(n-r+2)}} \\ \vdots & \vdots & \vdots & \vdots \\ v_{mt_r} & v_{mt_{(r+1)}} & \dots & v_{mt_n} \end{bmatrix}, \tag{4.6}$$

where r denotes the number of rows of the generalized Hankel matrix. Note that a large row number in Hankelization will lead to high complexity, while a small number of rows (i.e., less than 10) could considerably affect the recovering accuracy. Thus, we set $r = 20$ in our system (see Section 4.5.2).

Since thermal noise is filtered before tensor construction, the ideal tensor $\mathcal{V}$, which is constructed by the ideal phase variation matrix V_T, can be considered as low rank data. Thus, $\mathcal{V}$ can be estimated from the sampled sparse tensor $\bar{\mathcal{V}}$ by low-rank tensor completion, which is accomplished by solving the following optimization problem [168]:

$$\min_{\hat{\mathcal{V}}} \|\hat{\mathcal{V}}\|_*, \ \text{ s.t. } \Omega * \hat{\mathcal{V}} = \Omega * \bar{\mathcal{V}}, \tag{4.7}$$

where $\hat{\mathcal{V}}$ is an estimate of the ideal tensor $\mathcal{V}$, and Ω is a tensor of 0 and 1 elements, where $\Omega_{ijk} = 1$ when $\bar{\mathcal{V}}_{ijk}$ is sampled, and $\Omega_{ijk} = 0$ otherwise. $\|.\|_*$ denotes the trace norm of tensor [168].

We adopt the HaLRTC algorithm for tensor completion, which can solve the optimization problem (4.7) with the Augmented Lagrange Multiplier Method (ADMM) [146]. Compared with other tensor completion algorithms, HaLRTC usually achieves a higher accuracy at an acceptable complexity. A comparison of HaLRTC and a classic matrix completion method on phase variation recovery is shown in Fig. 4.5.

To make the comparison, we interrogate a single tag for 10 s, and repeat for 5 times as if there were 5 virtual tags. Since there is only one tag being interrogated each time, no data will be missing in V_T. We thus obtain the ideal phase variation data V_T as ground truth, and then remove some elements in V_T according to the slotted ALOHA protocol to obtain $\bar{V}_T$. The remaining elements used for recovery are marked with $\diamond$ in Fig. 4.5. Since there are 5 virtual tags in the emulated data, the sparsity is higher than 80%. Fig. 4.5 shows the first 500 samples recovered by tensor completion and matrix completion, respectively. It can be seen that the recovered signal by HaLRTC is very similar to the original signal, while the recovered signal by matrix completion is not good. Many values recovered by matrix completion are still very close to 0. This is because there are only 5 rows in $\bar{V}_T$, but the number of columns is

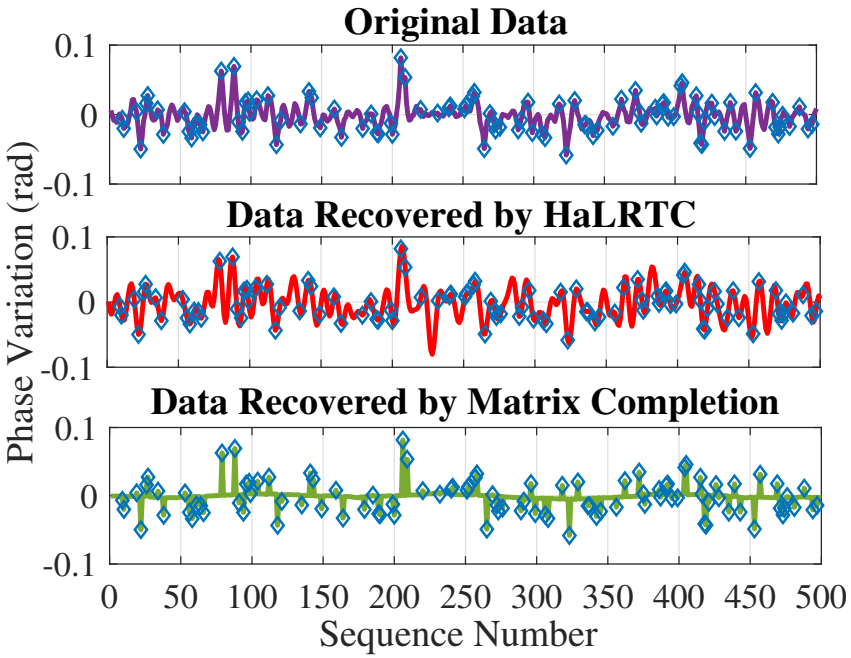


Figure 4.5: Recovered signals using HaLRTC and Matrix Completion.

now 500. SVD based matrix completion cannot obtain sufficient singular values for accurate estimation. In contrast, although tensor completion also requires singular values for estimation, the unfolding process can provide a sufficient matrix size for decomposing singulars values.

Once $\hat{V}_T$ is successfully recovered by HaLRTC, we can easily recover the phase sequence of each tag for all time slots by integrating the corresponding phase variations. Furthermore, the recovered phase data will not be affected by frequency hopping, because all the distortion related phase variations are deleted when building $\bar{V}_T$.

4.4.3 Dealing with Vehicle Vibration and Body Movements

Interference caused by vehicle vibration and body or environment movements is another big challenge for breath monitoring in driving environments, which should be mitigated before tensor decomposition. The driving movement could have different impacts for different tags, because the tags are deployed at different parts of the seat belt. Fortunately, most of the driving related body movements are not fast, and the resulting interference can be considered as a direct current (DC) component (with frequencies around 0) in the recovered phase signal. We thus apply a Hampel filter with a windows size of 3 s and a threshold of 0.001 to extract the DC component and then subtract it from the original signal.

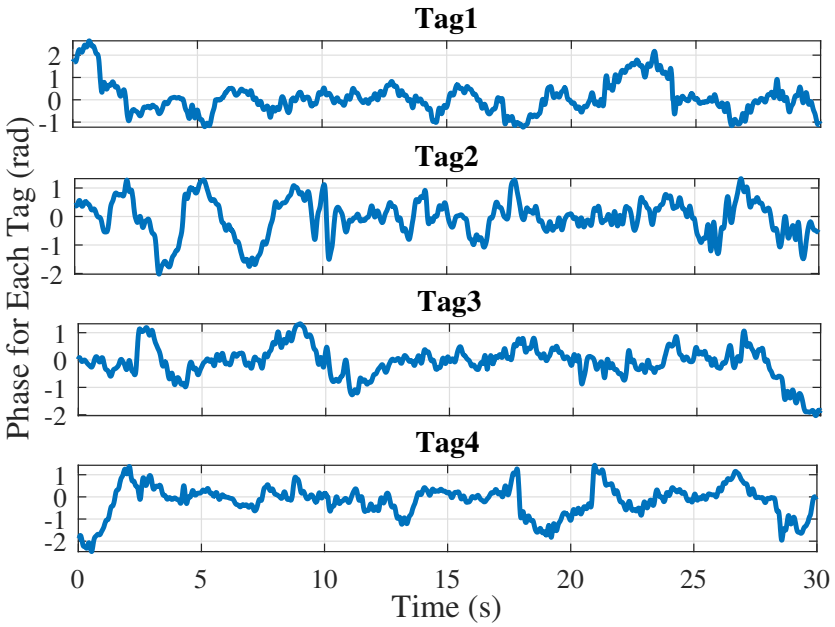


Figure 4.6: Recovered phase after DC removal.

The filtered signals from 4 tags are illustrated in Fig. 4.6. We find that the DC components are successfully removed, but the respiration signal is still hard to see. This is because the phase signals are also influenced by vibration of the moving vehicle. The noise generated by vibration is hard to be estimated, because both strength and frequency of the noise are related to road conditions and speed of the car.

However, since all the tags are attached to the seat belt, the noise generated by vibration has a similar effect on them. With filtered recovered phase variation signal, we can obtain the *phase difference* for each tag pair in each time slot. The resulting phase difference is plotted in Fig. 4.7, where most phase difference curves exhibit strong periodicity, meaning the influence of vibration has been effectively mitigated.

4.4.4 CPD based Respiration Signal Separation

Tensor Data Construction with Hankelization

Extracting a clean respiration signal from the calibrated phase difference data for different tag pairs, as shown in Fig. 4.7, is still challenging, because the periodicity strength of different tag pairs could vary with the movements of the driver and in the driving environment. It means no such tag pairs that could always reveal the highest periodicity, and the worst pairs could have no

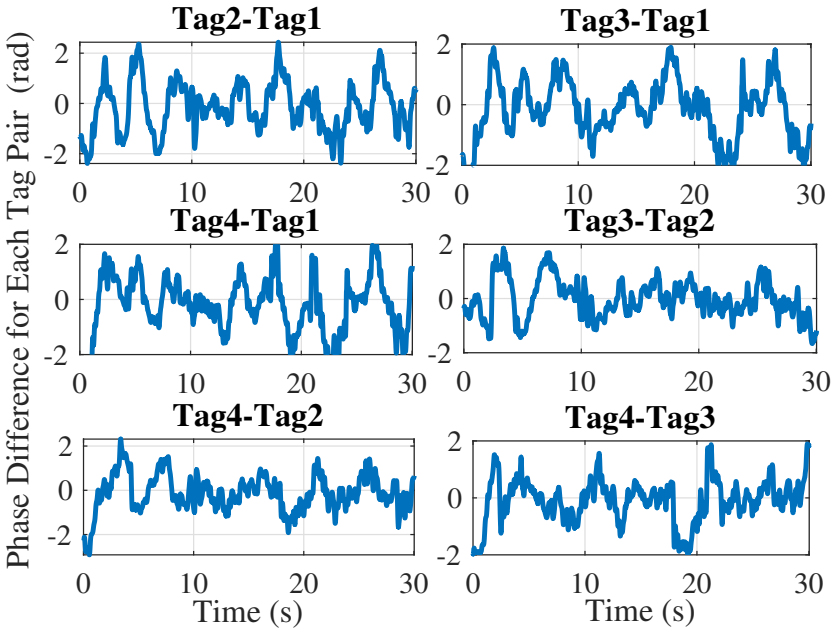


Figure 4.7: Phase difference for each pair of RFID tags.

periodicity in the estimated phase difference. Some traditional signal processing techniques, such as discrete wavelet transform (DWT), can extract the respiration signal from the signal with sufficient periodicity, but the performance could be poor when DWT is applied to noisy signals.

However, since the respiration signal captured by phase difference is the same for all tag pairs, each phase difference sequence can be considered as the same respiration signal corrupted by different noises. Rather than employing signal processing techniques for each individual tag pairs, we can analyze the entire group of estimated phase difference data. Tensor decomposition is widely used to separate a correlated signal from multiple datasets [19, 95]. For example, in TensorBeat [19], CPD is used to separate multiple breathing signals from WiFi phase difference data.

We can decompose the breathing signal from noises in driving environments using a *CPD tensor decomposition* method. The detailed process of respiration signal extraction is shown in Fig. 4.8.

As Fig. 4.8 shows, the respiration signal is extracted from the separated signals from CPD, so that the breathing rate could be estimated based on the interval of the detected signal peaks.

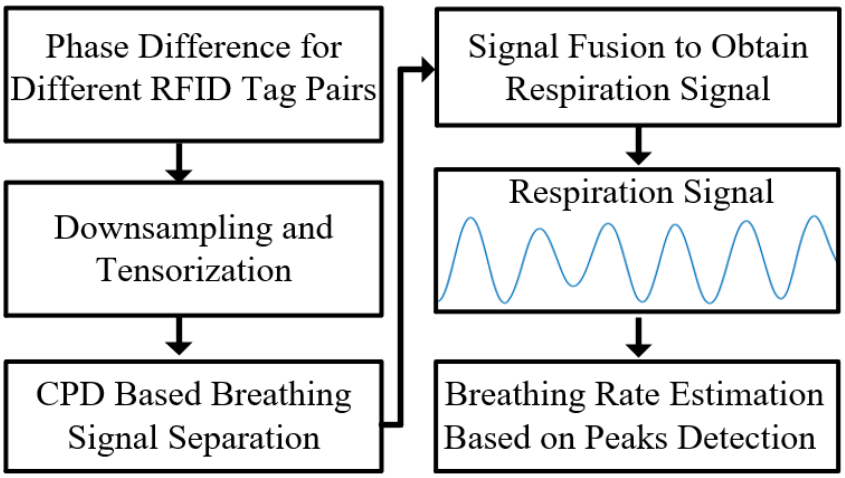


Figure 4.8: Flow-chart of the CPD based respiration extraction method.

Before decomposing the data with CPD, we transform the phase difference matrix into a tensor structure, by reshaping each phase difference sequence into a generalized Hankel matrix. Hankelization is essential before CPD decomposition is applied, because its special structure helps to separate a periodic signal (i.e., the breathing signal) from the data. We summarize the relationship between periodic signals and the proposed Hankel matrix in Theorem 4.1. The proof is provided in Appendix 4.7.

Theorem 4.1. *If the generalized Hankel matrix* $\boldsymbol{H}$ *with* r *rows is constructed from a sinusoidal wave with length* n*, it can be decomposed as:* $\boldsymbol{H} = \sum_{i=1}^{2} \alpha_i \cdot \beta_i^T$*, where both* $\alpha_i \in \mathbb{R}^{r,1}$ *and* $\beta_i \in \mathbb{R}^{(n-r+1),1}$ *represent sinusoidal signals with the same frequency as the original signal.*

Theorem 4.1 provides the theoretical underpinning for estimating respiration rate from the generalized Hankel matrix derived from the phase difference sequence. It also determines the number of sinusoidal components to be decomposed from the generalized Hankel matrix. In our system, we build the generalized Hankel matrix with 30 s of data and leverage 10 s of data to build each column of the matrix. This is because 10 s of data can guarantee that at least one full period of the breathing signal can be decomposed in α_i.

CPD Based Breathing Signal Separation

With the tensor constructed by Hankelization, denoted by Γ, the respiration signal can be separated from noise by applying CPD, which decomposes the tensor into a sum of the outer

products of three vectors as [96]

$$\Gamma \in \mathbb{R}^{I,J,K} \approx \sum_{m=1}^{M} \vec{a}_m \circ \vec{b}_m \circ \vec{c}_m, \tag{4.8}$$

where M is the tensor rank used for CPD, which also indicates the number of components in the decomposition result; $\vec{a}_m$, $\vec{b}_m$, $\vec{c}_m$ are the vectors at the mth position for the three dimensions, respectively. We have $\vec{a}_m \in \mathbb{R}^{I,1}$,$\vec{b}_m \in \mathbb{R}^{J,1}$, $\vec{c}_m \in \mathbb{R}^{K,1}$, and $(\vec{a}_m \circ \vec{b}_m \circ \vec{c}_m)(i,j,k) = \vec{a}_m(i)\vec{b}_m(j)\vec{c}_m(k)$, for all i,j,k. According to the definition of outer product, we can construct the matrix from the vectors for each dimension. For example, matrix $\boldsymbol{A}$ is defined as $[\vec{a}_1, \vec{a}_2, ..., \vec{a}_M]$, and matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ are defined similarly with $\vec{b}_m$ and $\vec{c}_m$, respectively.

The process of CPD is implemented by solving the following optimization problem.

$$\min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \|\Gamma - \sum_{m=1}^{M} \vec{a}_m \circ \vec{b}_m \circ \vec{c}_m\|_F^2. \tag{4.9}$$

Although the above problem is not convex, CPD leverages the Alternating Least Squares (ALS) algorithm to optimize one matrix while fixing the other two. With the ALS algorithm, the optimization problem can be reduced to a linear least squares problem, and the three matrices can be finally estimated.

In CPD, the number of components M should be prescribed, which is determined by the target signal and the uniqueness of the decomposition. Thus, we propose Theorem 4.2 to determine the range of M, which is proved in Appendix A.

Theorem 4.2. *The tensor rank used for CPD in the proposed system should satisfy* $2 \leqslant M \leqslant 4$ *and the CPD is unique.*

The noise in the tensor is usually considerably large in driving environments (i.e., much larger than the respiration signal). Some other noise components could also be decomposed. To separate noise and precisely extract the respiration signal, we set $M = 4$ for CPD. The decomposed matrix $\boldsymbol{B}$ is illustrated in Fig. 4.9. We can see that two sinusoidal signals with the same period are decomposed by CPD (i.e., the 3rd and 4th components), while the other components (i.e., the first and 2nd components) are for the residual noise after data preprocessing.

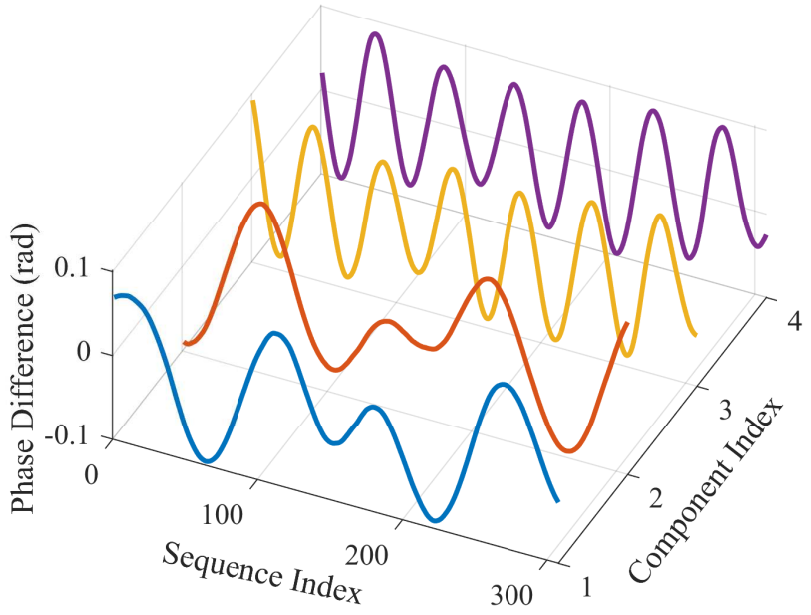


Figure 4.9: Decomposed signals by CPD.

4.4.5 Breathing Rate Estimation

With the respiration signal separated by CPD, we can estimate the breathing rate using a peaks detection algorithm. However, we still need to differentiate the periodic signal from all decomposed signals before we estimate the breathing rate. Thus, we leverage a frequency spectrum based method to figure out which decomposed signal is for the breathing signal. We first calculate the proportion of the power spectrum between 0.2 Hz and 0.5 Hz in the frequency domain, which is the range of normal human breathing. Then, we search for the two signals with the first two largest proportions and fuse them to obtain the final breathing signal. The fused signal is illustrated in Fig. 4.10. It can be seen that the respiration signal is precisely extracted and the noises from vehicle vibration and other movements in the driving movement are effectively removed. The intervals τ between each pair of adjacent peaks are calculated and the breathing rate F is determined by the average interval τ_{ave} as $F = 60/\tau_{ave}$.

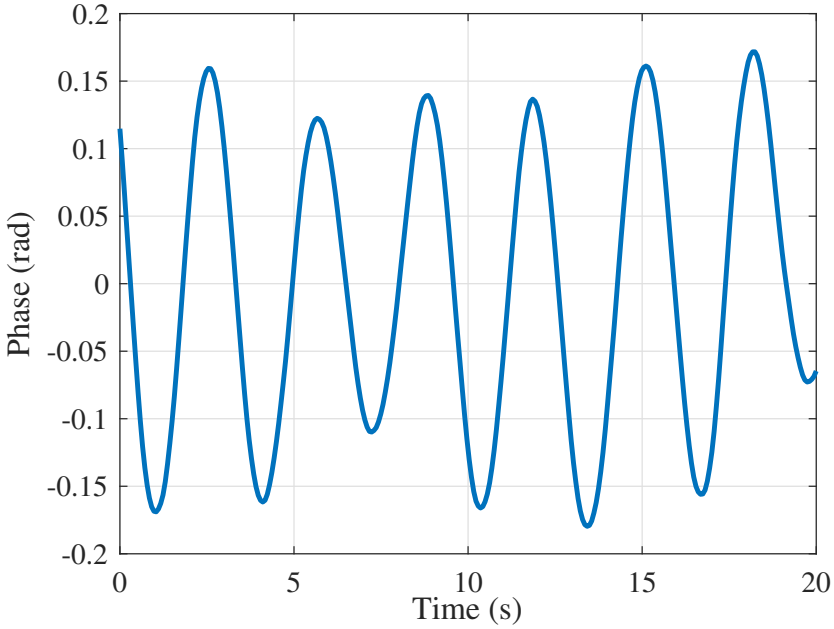


Figure 4.10: Signal fused by all breathing related components.

4.5 System Performance Evaluation

4.5.1 Experiment Configuration

To evaluate the performance of our respiration monitoring system in the driving environment, we implement a prototype system with an Impinj R420 reader equipped with a polarized antenna S9028PCR.[1] The setup up of the system is illustrated in Fig. 4.11. As the figure shows, multiple ALE-9470 RFID tags are attached to the seat belt of the driver. The size of the polarized antenna utilized in the system could be smaller, because many small-sized antennas have been developed, such as Keonn Advantenna-p11 and Thingmagic EL6E.

Although the cost of the RFID reader used in the prototype system is not very low, some other cheaper readers could be adopted for reduced cost. For example, since only one antenna is required in our system, one port reader like Impinj R120 can be used. Furthermore, medium range readers, such as Feig MRU102-PoE, will be another low-cost option, because the interrogate range for car environment monitoring is not demanding. Furthermore, our system is

[1] Although we used such commodity RFID devides in our proof-of-concept prototyping and experiments, the size of the system can be greatly reduced by using smartphones with an attached reader, e.g., made by Zebra.

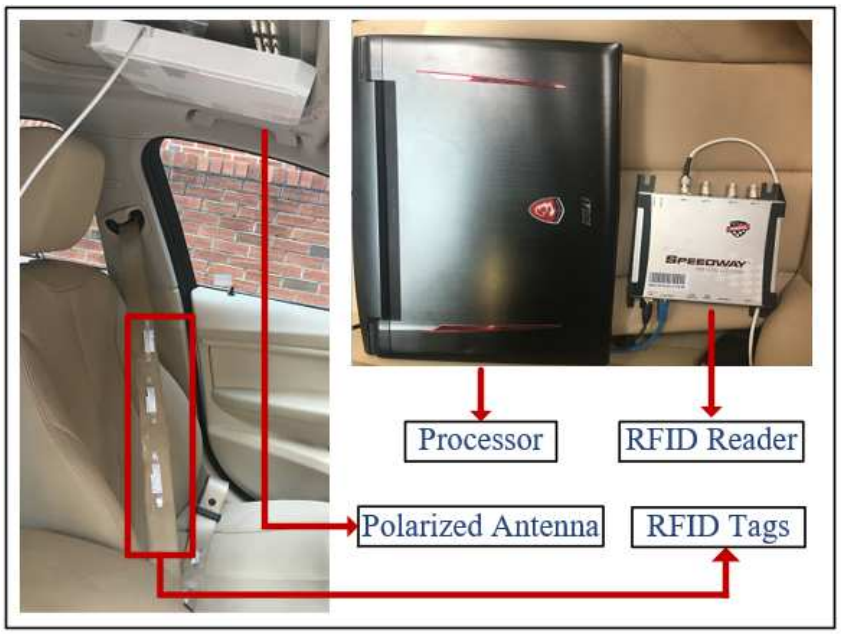


Figure 4.11: Illustration of the system setup in a car in our experiments.

currently composed of multiple commodity devices. The cost of the future system could be further reduced, if customized readers are used and mass produced.

The channel used by the reader hops every 0.2 s among 50 channels from 902 MHz to 928 MHz when interrogating RFID tags. The processor used for signal processing is an MSI laptop with a Navidia GTX 1080 GPU and Intel Core i7-6820HK CPU. The model of the test vehicle is BMW 328i. Five volunteers (1 female and 4 males) are tested with our prototype system. The breathing rates of the volunteers are also estimated by a NEULOG sensor, which is considered to be the ground truth in our experimental result analysis.

4.5.2 Results and Discussions

Overall Accuracy for Different Rates

Our system is tested by five volunteers with different breathing rates. The cumulative distribution function (CDF) of estimation errors is presented in Fig. 4.12. The figure shows that the median errors for the three ranges of breathing rates are 0.11 bpm, 0.12 bpm, and 0.14 bpm, respectively, which are all very close to each other. However, the maximum error for the $10 \sim 15$ bpm range is 0.33 bpm, which is much smaller than that for the $16 \sim 23$ bpm range and that for the $24 \sim 30$ bpm range. This implies that the accuracy of the system is higher when the

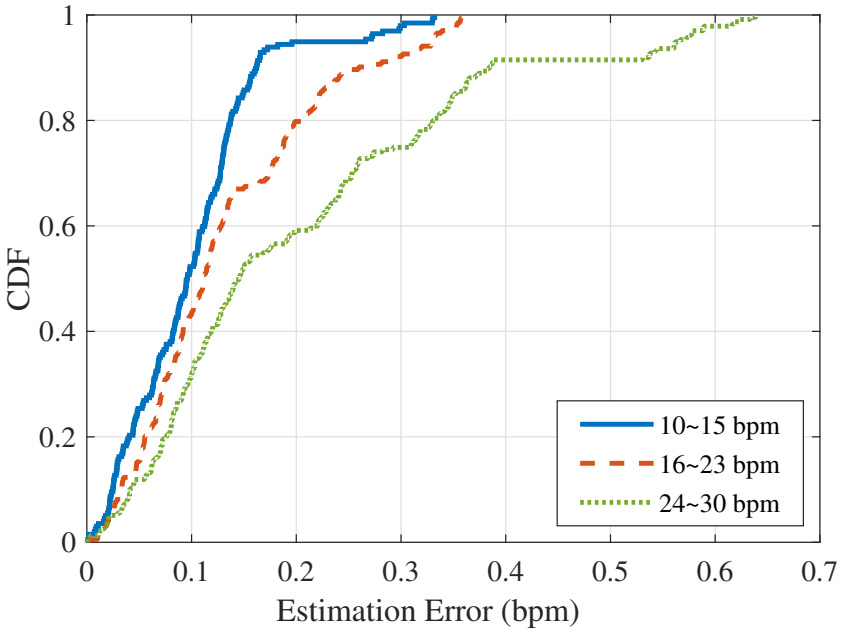


Figure 4.12: System performance for different breathing rates.

driver breaths slowly. This is because breathing rate is calculated by $F = 60/\tau_{ave}$, where τ_{ave} is estimated by peak detection. As the driver's breathing gets faster, τ_{ave} will become smaller, and the estimation error in τ_{ave} will be amplified in F.

We next compare with the traditional RFID based method presented in [8,58] in Fig. 4.13, when the driver is driving on a highway (i.e., interstate I-85). Other RFID based respiration monitoring techniques, such as Tagbreathe [27], can achieve high accuracy in a stationary testing environment. However, the Tagbreathe system is not well suited for operation with Ultra High Frequency (UHF) RFID devices in the US, which requires frequency hopping. Thus, we only provide the experimental result of the RFID breathing monitoring system in a stationary environment proposed in [8, 58] to illustrate the robustness of our system in noisy driving environments.

The CDFs of estimation errors achieved by the traditional and proposed methods are presented in Fig. 4.13. We find that the median error of the traditional method is 1.71 bpm and 36.11% data has an estimation error larger than 2 bpm. In contrast, the median error of our proposed method is 0.12 and the maximum error is 0.36 bpm. From the results, we can conclude that the performance of the traditional method is quite limited in driving environments, because of the considerable interference caused by the vehicle, driver, and passengers. However, due

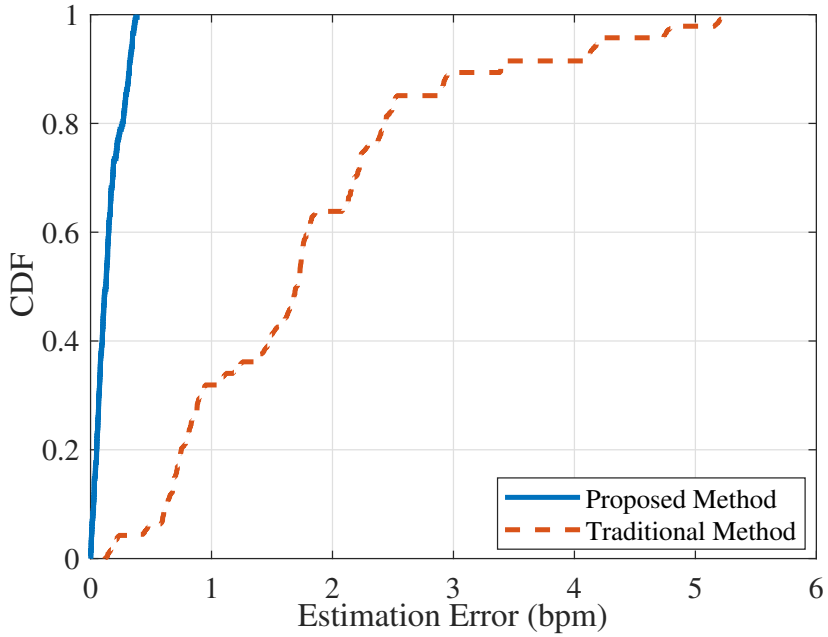


Figure 4.13: System performance compared with a traditional RFID based respiration sensing technique for stationary environments [8, 58].

to effective noise mitigation, our proposed method can achieve high accuracy in respiration monitoring in such noisy environments.

To compare the accuracy of breathing rate estimation for different types of systems, we summarize the mean estimation error as provided in related papers [66, 77] in Table 4.1 (note that we compare mean error rather than median error here, since those are provided in the two related papers). As shown in the table, the video camera based method and UWB radar based method have higher estimation errors, which are 0.79 bpm and 0.31 bpm, respectively. The mean estimation error of the proposed system is 0.11 bpm, which is much lower than the other two. This is because RFID tags can provide high accuracy as wearable sensors, and the effect of interference from other passengers is limited too. Moreover, with measured phase from multiple RFID tags, the CPD based approach can effectively mitigate the influence of vehicle vibration and other noises in driving movements.

Accuracy in Different Driving Scenarios

The system is evaluated in 3 different scenarios, including (i) driving on a highway, (ii) driving in a city street, and (iii) parked. The CDF of errors for the three scenarios are plotted in

Table 4.1: Comparison of Different Breathing Rate Monitoring Systems for the Driving Environment

Employed Device	*Mean Estimation Error*
Video Camera (Kinect)	0.79 bpm
UWB Radar	0.31 bpm
RFID Tags and Reader	0.11 bpm

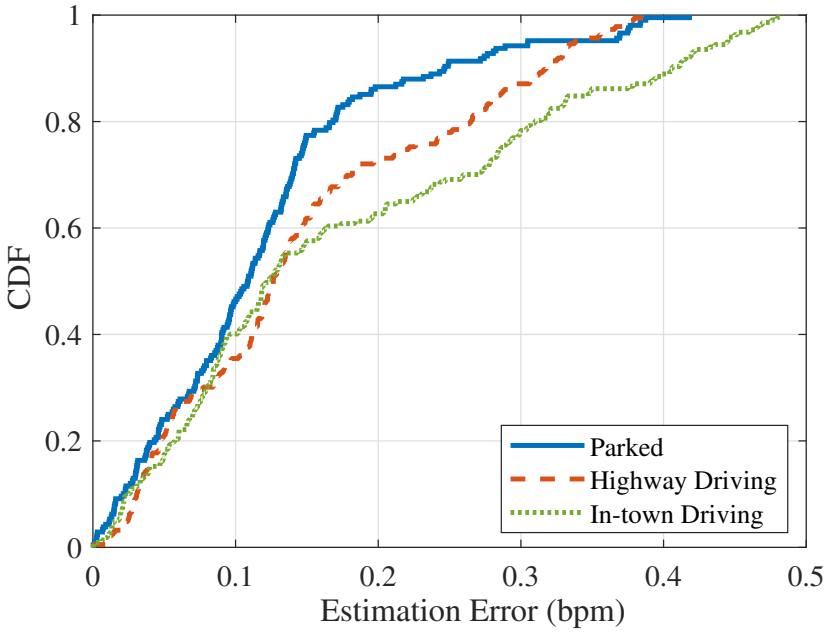


Figure 4.14: System performance for different driving scenarios.

Fig. 4.14. The median error is 0.11 bpm for the parked scenario and 0.12 bpm for the two driving scenarios. It proves that the influence of the driving environment is effectively mitigated by the proposed scheme. The figure also shows that the maximum error for in-town driving is 0.48 bpm, which is the highest among the three scenarios. This is because the driver needs to turn the wheel frequently and stops from time to time when driving in town, and the influence of the driving movements cannot be completely eliminated.

Impact of Passenger Movement

We also test the performance of our system when there are different numbers of passengers. In these experiments, all the passengers are asked to move naturally. As Fig. 4.15 shows, when no passenger is present, the median error is 0.09 bpm, which is the smallest among all the cases.

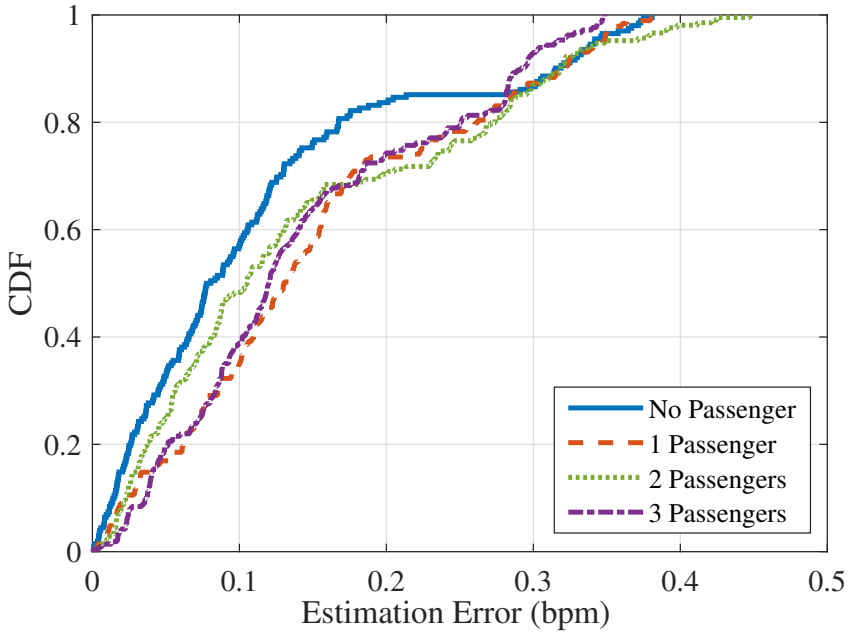


Figure 4.15: System performance with different numbers of passengers.

The error increases to 0.11 bpm when there is one passenger in the car. The movement of the passenger does affect the sampled phase data, although the impact is small. The median error is not obviously affected by increased number of passengers, and the maximum errors are almost the same for the different cases. This result shows that the performance of our system is not sensitive to the movement of passengers, because the tags are only attached to the seat belt of the driver as well as the near-field nature of RFID communications.

Impact of Antenna Position

We try different antenna deployment in our experiments to identify the most suitable way to place the antenna. The first scenario is to attach the antenna on the ceiling above the driver seat, as shown in Fig. 6.12. The second deployment is to attach the antenna to the front control panel of the vehicle. The third scenario is to bond the antenna on the side of the back of the co-pilot seat. The CDF results are plotted in Fig. 4.16. It can be seen that the deployment of the antenna has a considerable impact on the system performance. The median error is 0.73 bpm when the antenna is placed on the side of the diver, because the phase changes caused by the movement of human chest is harder to detect. The accuracy is also not high when the antenna is attached to the front panel, with a median error of 0.28 bpm. This is because the range of the

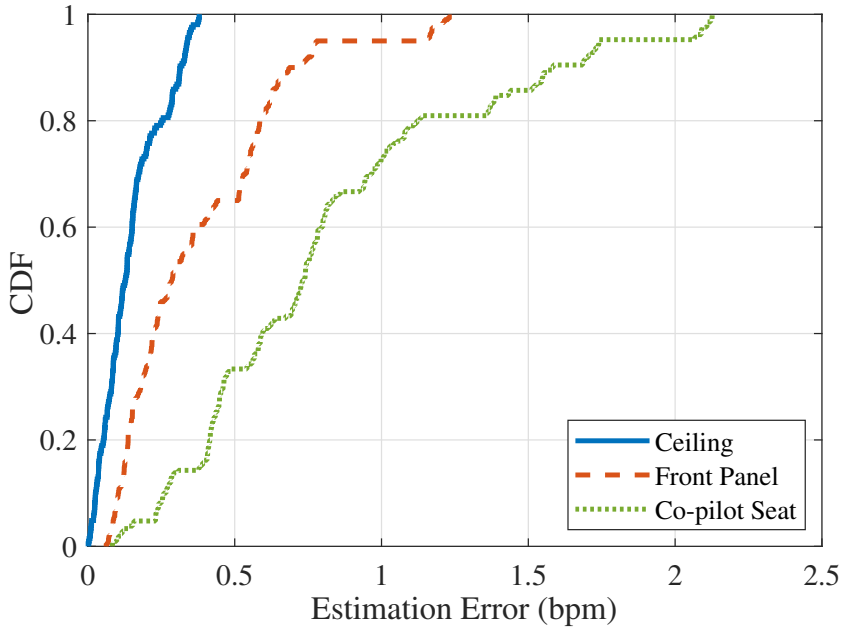


Figure 4.16: System performance under different deployment locations of the polarized antenna.

polarized antenna is limited, and some of the tags cannot be scanned by the antenna. Thus, we conclude that the best deployment position of the antenna is the ceiling, which can guarantee full-tag interrogation and high sensitivity of detecting the respiration signal.

Impact of Number of Tags and Coupling Effect

We next evaluate the accuracy of our system with different numbers of tags. Since we need to calculate phase difference between tag pairs and build the tensor data, the minimum number of tags is 3 in our system. Fig. 4.17 shows the accuracy under different numbers of tags. The mean error is relatively high when only 3 tags are deployed, which is 0.45 bpm. This is because the 3 tags can only generate 3 phase difference sequences for CPD. The respiration signal is hard to be decomposed from the small tensor constructed by 3 tags. We also observe that when 5 or more tag are deployed, the error is reduced to 0.11 bpm. However, the error becomes 0.21 when 9 tags are used, because the error in recovering the missing samples becomes large with 9 tags. Moreover, more time will be consumed in data recovering and tensor decomposition when more tags are used. We use 5 tags on the seat belt in our system.

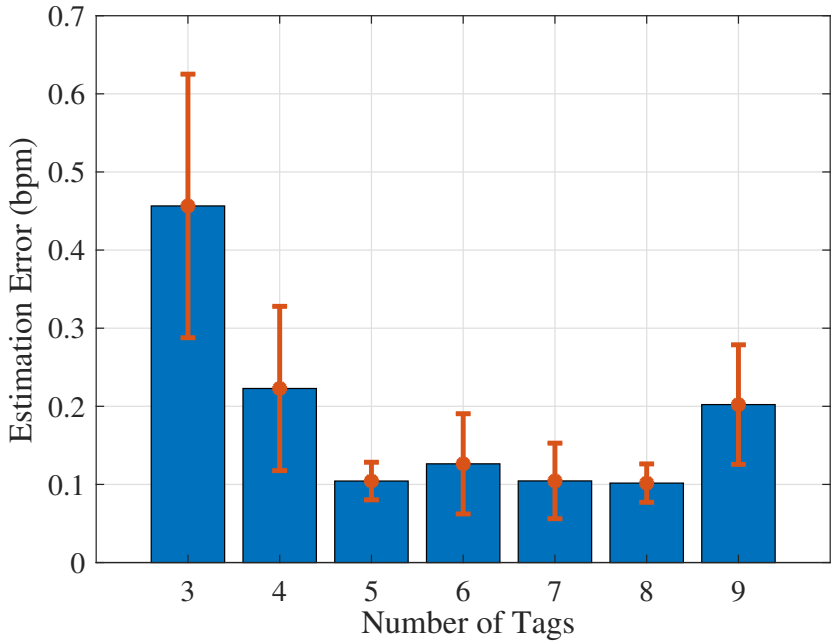


Figure 4.17: Estimation error for different numbers of deployed tags.

Fig. 4.17 also shows that the accuracy of the system is not seriously affected by the coupling effect, because the estimation error remains small (around 0.11 bpm) even when 8 tags are used. With 8 tags attached to the seat belt, the density of RFID tags is quite high, which generates large mutual coupling among these tags. Fortunately, the coupling effect only affects the phase value received by the reader [23], while the breathing rate estimation is dependent on the periodicity of the signal. Thus the system is robust to the mutual coupling effect. The figure also shows that, the error increases a little (to 0.2 bpm) when 9 tags are deployed. This is because the considerable larger mutual coupling effect may introduce low backscattering power from the tags in this case. Since the reader uses a power threshold for receiving tag response, tags with low backscattering power will hardly be interrogated by the reader. To make sure that all tags can provide enough information to the reader, we conclude that 8 tags are the maximum number of tags for the prototype system.

Complexity Reduction

Since the system requires solving two optimization problems for data recovery and tensor decomposition, the complexity of the algorithm could be a problem that needs to be investigated. We aim to reduce the complexity of HaLRTC and CPD by reducing the size of the tensors and

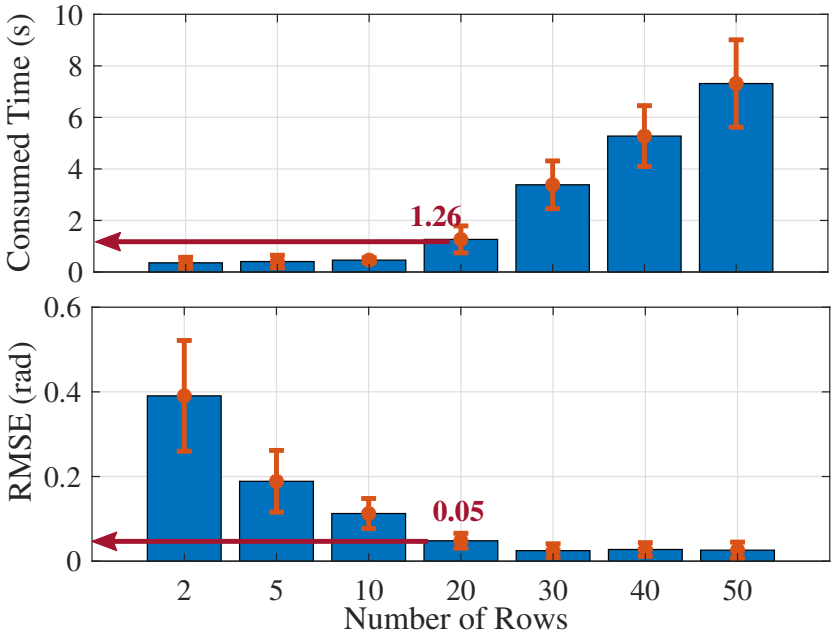


Figure 4.18: Impact of Hankelization size on HaLRTC complexity and recovering accuracy.

by downsampling. For phase recovery, we can reduce the size of the tensors by adjusting the number of rows in Hankelization. We utilize the 20 s data used in Fig. 4.5 to test the influence of the number of rows on complexity and accuracy. The results are shown in Fig. 4.18. We can see that as the number of rows is increased, the consumed time increases exponentially, while the Root Mean Square Error (RMSE) decreases exponentially. To trade off between accuracy and complexity, we choose 20 as the row number for building generalized Hankel matrix.

Downsampling is not implemented in the phase recovering process, because the sparsity of the data is very high and downsampling could considerably affect the recovery performance. However, downsampling is an effective way to reduce the complexity of CPD. Following Theorem 4.1, the rows of the generalized Hankel matrix in CPD should be large enough, or the periodicity of the signal can hardly be revealed in α_i. Thus, we fix the number of rows and test the performance of CPD with different downsampling indices. As Fig. 4.19 shows, when the index is larger than 12 the consumed time by CPD is shorter than 0.82 s, and the estimation error is smaller than 0.2 bpm when the index is smaller than 14. When the downsampling index gets higher, the accuracy decreases sharply, because the remaining data is not enough for CPD to precisely separate the breathing signal. To achieve high accuracy with an acceptable complexity, we downsample the data by 12 before constructing the tensor input for CPD.

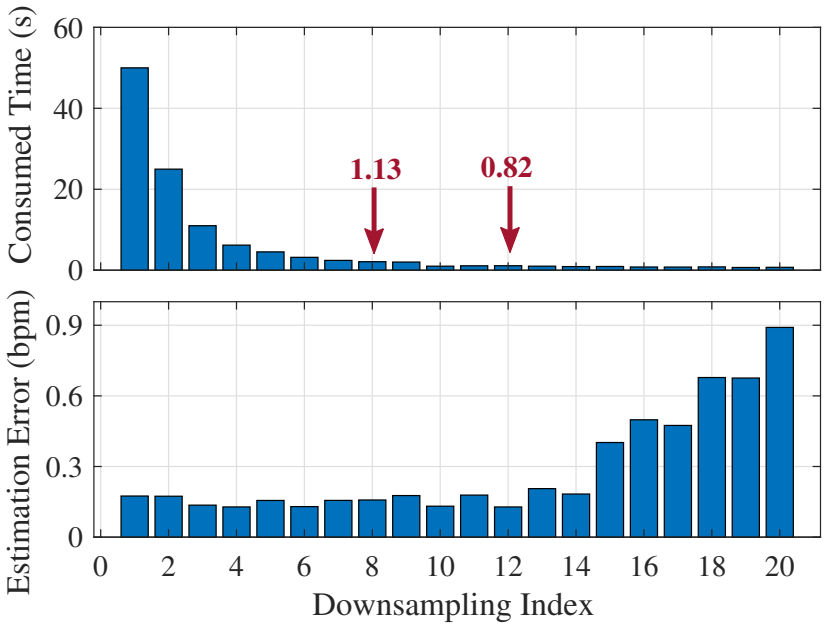


Figure 4.19: Impact of downsampling on CPD complexity and system performance.

In conclusion the total time consumed for data processing in our system is about 2 s. Since drowsy driving is a relatively long, slowly developing process (i.e., people do not fall asleep suddenly), such a latency should be sufficient to warn the driver in advance (e.g., the safe distance to the vehicle ahead of one's car is 2 or 3 s). In addition, the processing speed can be improved by specific embedded hardware. The complexity of the prototype system is acceptable.

4.6 Conclusions

In this paper, we proposed an RFID based respiration rate monitoring system for the driving environment. The proposed system included several novel components, including data collection, breathing data preprocessing, CP tensor decomposition, and respiration signal estimation, to combat the strong noise caused by frequency hopping, random sampling, vehicle vibration, and other movements in the environment. The proposed system was implemented with commodity RFID tags and reader, and was evaluated under real driving scenarios. Our experiments showed that the tensor completion and CPD approaches were effective for respiration monitoring in driving environments.

4.7 Proof and discussion of the theorems in CPD processing

4.7.1 Proof of Theorem 4.1

Proof. To analyze the features of the generalized Hankel matrix, we first assume the original signal is a noise-free, discrete sinusoidal signal sampled at a constant period t, which is given by $y(n) = a\sin(wtn + \varphi_0)$, where a, w, and φ_0 are the amplitude, frequency, and initial phase offset of the sinusoidal signal, respectively.

For convenience, we define

$$\vec{S}_p^q \doteq [\sin(wtp + \varphi_0), \sin(wt(p+1) + \varphi_0), ..., \sin(wtq + \varphi_0)],$$

and rewrite the generalized Hankel matrix as

$$\boldsymbol{H} = a \cdot \left[\vec{S}_1^{(n-r+1)},\ \vec{S}_2^{(n-r+2)},\ \cdots,\ \vec{S}_r^n\right]^T. \tag{4.10}$$

We also define

$$\vec{C}_p^q \doteq [\cos(wtp+\varphi_0), \cos(wt(p+1)+\varphi_0), ..., \cos(wtq+\varphi_0)].$$

Since $\sin(wt(n+1) + \varphi_0) = \sin(wtn + \varphi_0)\cos(wt) + \cos(wtn + \varphi_0)\sin(wt)$, we can convert the time-shift of Hankelization to a summation format. The generalized Hankel matrix $\boldsymbol{H}$ can be rewritten as

$$\boldsymbol{H} = a\cdot\begin{bmatrix} \vec{S}_1^{(n-r+1)}\cos(0) + \vec{C}_1^{(n-r+1)}\sin(0) \\ \vec{S}_1^{(n-r+1)}\cos(wt) + \vec{C}_1^{(n-r+1)}\sin(wt) \\ \vec{S}_1^{(n-r+1)}\cos(w2t) + \vec{C}_1^{(n-r+1)}\sin(w2t) \\ \cdots \\ \vec{S}_1^{(n-r+1)}\cos(w(r-1)t) + \vec{C}_1^{(n-r+1)}\sin(w(r-1)t) \end{bmatrix}.$$

Note that both vectors $[\cos(0), \cos(wt), ..., \cos(w(r-1)t)]$ and $[\sin(0), \sin(wt), ..., \sin(w(r-1)t)]$ are discrete sinusoidal signals with the same period w. Thus, we can rewrite $\boldsymbol{H}$ as $\boldsymbol{H} =$

$\boldsymbol{\psi}_1^T \cdot \boldsymbol{\psi}_2$, where $\boldsymbol{\psi}_1 = [\alpha_1, \alpha_2]^T$ given by

$$\boldsymbol{\psi}_1 = 1 \cdot \begin{bmatrix} \cos(0), \cos(wt), ..., \cos(w(r-1)t) \\ \sin(0), \sin(wt), ..., \sin(w(r-1)t) \end{bmatrix}, \tag{4.11}$$

and $\boldsymbol{\psi}_2 = [\beta_1, \beta_2]^T$ given by

$$\boldsymbol{\psi}_2 = a \cdot \left[S_1^{(n-r+1)},\ C_1^{(n-r+1)} \right]^T, \tag{4.12}$$

where each row of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ is a sinusoidal signal with period w. Therefore, both α_i and β_i are sinusoidal signals with the same period w. □

4.7.2 Discussion of Theorem 4.2

Since only the driver's breathing signal is captured by the phase difference sequence, the modified Hankel matrices can be regarded to be generated by a single sinusoidal signal (i.e., the breathing signal) plus noises from the driving environment. Following Theorem 4.1, the two related components should be decomposed, and thus we have $M \geqslant 2$.

Next we consider the upper bound of M to satisfy the uniqueness requirement of CPD. It is given that the output of CPD is unique only when $R_A + R_B + R_C \geq 2M + 2$ [96], where R_A, R_B, and R_C are the number of independent rows in matrix A, B, and C. In the constructed tensor, R_C is determined by the number of independent signals, while R_A and R_B are determined by the sampled signal used for hankelization. In our system, each tag can be considered as an independent sensor. So R_C equals to the number of deployed tags, which means $R_C \geqslant 4$ (since 4 to 8 tags are used). In addition, we have $R_A = R_B \geqslant 3$ because the breathing signal itself takes 2 rows and the noise in sampled data occupies at least one independent row. Therefore, to satisfy the uniqueness condition of CPD, we should have $2M + 2 \leqslant \min\{R_A\} + \min\{R_B\} + \min\{R_C\} = 3 + 3 + 4 = 10$, i.e., $M \leqslant 4$. Thus, we conclude that the tensor rank used for CPD in the proposed system should satisfy $2 \leqslant M \leqslant 4$ and the CPD is unique.

Chapter 5

SparseTag: RFID Tag Localization with a Sparse Tag Array

5.1 Introduction

With the rapid growth of the Internet of Things (IOT), the Radio Frequency Identification (RFID) technology has been regarded as an effective and low-cost solution for many emerging IoT applications. In addition to the wide adoption in traditional identification applications in various fields, such as retailing, sports, library, manufacturing, and supply chain management, positioning of RFID tags has attracted increasing interest from researchers in recent years. Rather than reading the stored Electronic Product Code (EPC) from RFID tags, the low level data of the RFID channel, such as received signal strength indication (RSSI) and phase, can be collected from the received tag responses and leveraged for tag localization.

RSSI-based technique has been proposed to localize RFID tags [31], but the accuracy of such systems is usually limited by the low resolution and randomness of the RSSI data. Active RFID tags have been adopted in prior works, which usually has a much higher cost than the passive tags. For passive tag based localization, phase angle has been widely utilized because of its high resolution and stability. However, due to the wide beam of polarized reader antenna and the multipath effect, high-accuracy positioning of passive tags is still a big challenge. To achieve narrow beams of the reader antenna for high-accuracy localization, multiple antennas can be utilized [33, 35], but at a higher cost. Systems with a single moving antenna or moving RFID tags are then proposed for reduced cost [21, 22, 34], which generate additional virtual antennas instead of using real ones. These techniques can achieve high localization accuracy,

but the moving the antenna or tag incurs time delay and requires careful calibration of the system.

Recently, RFID tag array has been leveraged to improve the accuracy and the robustness of RFID-based sensing systems [98]. For example, Tagyro uses a hologram-based method to transform phase offset to orientation of the tag array [23] for tracking the 3D orientation of passive objects. RF-Wear is developed for orientation estimation with a uniform linear array (ULA) for body-frame tracking [99]. The accuracy of direction of arrival (DOA) based localization techniques can be improved by utilizing more antennas. Thus leveraging a tag array with more RFID tags is an effective way to achieve high positioning accuracy. Compared with the systems with multiple polarized antennas, the cost of building a passive RFID tag array is negligible.

However, the technical challenges still exist for tag array-based localization, such as how to mitigate the multipath effect from the propagation environments and the phase distortion caused by mutual coupling between RFID tags. To deal with the influence of multipath effect, some existing techniques leverage a mobile antenna to localize a tag array in different positions [100, 101]. Although the mobile antenna can reduce the cost, the specialized mobile shelf and motor incur additional cost. For DOA based localization, the multipath effect could be effectively mitigated by utilizing an RFID tag array with sufficient number of tags.

However, when the traditional ULA tag array is used, the tag density could be high when many tags are placed on a small surface of the object, such as a book or a small package. In such scenarios, the accuracy will be influenced by the strong mutual coupling effect, which introduces additional frequency offset as well as amplitude offset of the resonance peak [102, 103]. It has been proved by several existing systems that mutual coupling generates considerable interference to the collected phase angle of RFID tags, which degrades the localization performance [105, 106]. Furthermore, the backscattered signal from RFID tags may not be sufficiently strong to be detected by the antenna, because the strength of the signal is also affected by mutual coupling. Thus, a special tag array with a lower density than ULA is needed, to be resilient to mutual coupling and deliver accurate DOA estimation.

In this paper, we propose a novel sparse RFID tag array for tag localization [57]. We first an analysis of the mutual coupling effect and prove that the phase difference from pairs of tags used in our system is independent to the coupled voltage and mutual impedance. Next, we present the SparseTag system, i.e., a **Sparse** RFID **Tag** array system for high-precision backscatter indoor localization, which comprise a sparse tag array and an RFID reader with two antennas. We analyze our sparse array processing for DOA estimation, which is quite different from the traditional MUSIC algorithm based methods using a ULA [107]. The key idea is to obtain a new signal vector with a different co-array, which is a longer array whose antenna locations are not evenly spaced. In addition, we design a new sparse RFID tag array, which has a symmetric structure and is effective for mitigating the mutual coupling effect. We derive its difference co-array and prove its several important properties, such as its hole-free feature, degrees of freedom (DOF), and weight function. We analytically show why the proposed sparse tag array can outperform ULA on DOA estimation. Then, we develop a DOA estimation scheme using the difference co-array of the proposed sparse tag array with a spatial smoothing method. Finally, we provide a localization method based on the two estimated DOAs, while a robust channel selection method is proposed for mitigating the multipath effect. We implement SparseTag with off-the-shelf RFID tags and reader, and evaluate its performance in two environments, including a computer laboratory and an anechoic chamber, where superior DOA estimation and location performance over the ULA-based benchmark scheme are demonstrated.

Compared with the brief introduction in the previous work, we have greatly enhanced the description and presentation in the current journal version, and extensive new experimental results were added to demonstrate the superiority of the proposed tag array. In section 5.2, we have newly conducted an experiment to validate the influence of different types of tags on mutual coupling. Besides, a new theorem 5.2 is added to show the drawbacks of the normal ULA tag array and analyze the influence of mutual coupling on the back-propagating power. Furthermore, multiple figures are newly added to illustrate the extensive addition experimental results.

The main contributions made in this paper are summarized as follows.

- We justify the feasibility and advantages of utilizing a sparse tag array for DOA based indoor localization through analysis and experiments. To the best of our knowledge, this is the first work to leverage sparse tag arrays for backscatter indoor localization, which does not require to move either the tags or the antenna(s).

- We design the SparseTag system, which includes sparse array processing, difference co-array design, DOA estimation using a spatial smoothing based method, and a localization method. We propose a new sparse tag array design and analytically prove its superior performance over the traditional ULA design. In addition, a robust channel selection method based on the sparse tag array is proposed for mitigating the indoor multipath effect.

- We implement SparseTag with off-the-shelf RFID tags and reader, and evaluate its performance in two indoor environments, including a computer laboratory and an anechoic chamber, with extensive experiments. The experimental results verify the effectiveness and strengths of the proposed SparseTag system.

The remainder of this paper is organized as follows. We analyze the mutual coupling effect on RFID phase difference and the success rate of sampling in Section 5.2. The proposed SparseTag system is presented in Section 5.3 and its performance is evaluated in Section 5.4. Section 5.5 discusses related work and Section 5.6 concludes this paper.

5.2 Analysis of Mutual Coupling

5.2.1 Phase Angle and Phase Difference

The FCC requires frequency hopping to avoid interference for readers. The readers use the spectrum between 902.5 MHz and 927.5 MHz, which is divided into 50 channels. The reader uses the Low-Level Reader Protocol (LLRP) to interrogate the tags, which can provide the RF phase angle, Doppler frequency, and Peak RSSI of the RFID channel [38]. In particular, the

phase angle, denoted by ϕ, can be written as

$$\phi = \text{mod}\left(\frac{2\pi \cdot 2l}{\lambda} + \theta_t + \theta_r + \theta_{tag}, 2\pi\right), \tag{5.1}$$

where l is the distance between the tag and the reader antenna, λ is the wavelength of the signal, and θ_t, θ_r, and θ_{tag} are the offsets introduced by the reader's transmitting circuit, the reader's receiving circuit, and the RFID tag's backscattering circuit, respectively. The challenge for RFID-based sensing techniques is how to translate the measured phase ϕ to distance l, under strong interference from the phase offsets and frequency hopping.

To mitigate the impact of phase offsets, we propose to adopt an RFID tag array. Rather than using the phase angle from each individual tag, the difference between a pair of neighboring tags is used. Following (5.1), the phase difference between Tags 1 and 2 is given by

$$\Delta\phi_{1,2} = \text{mod}\left(\frac{2\pi \cdot 2(l_1 - l_2)}{\lambda} + \theta_{tag_1} - \theta_{tag_2}, 2\pi\right), \tag{5.2}$$

where l_1 and l_2 are the distances from Tags 1 and 2 to the reader antenna, respectively; and θ_{tag_1} and θ_{tag_2} are the phase offsets due to Tags 1 and 2's circuit, respectively. It can be seen that the phase offsets introduced by the reader's circuits, i.e., θ_t and θ_r, are canceled in (5.2). In addition, the incident wave from the reader antenna is similar to a plane wave if the tag-antenna distance is sufficiently long. In this case, the phase difference can be translated to the DOA, if the pair of tags are placed closer than $\lambda/4$ [106].

5.2.2 The Mutual Coupling Effect

When a tag array is deployed, the mutual coupling effect becomes a limiting factor of the sensing performance. The inductive coupling of neighboring RFID antennas causes transfer of energy between closely placed tags, which usually affect the measured phase angles and the received signal strength at the reader. In the remainder of this section, we will provide an analysis of the effects of mutual coupling on phase difference and sampling effectiveness.

Impact on Phase Angle and Phase Difference

The Gen 2 protocol is adopted for the interrogation process to avoid collision of simultaneous responses to a query from multiple tags [108]. With this protocol, among the tags that respond to the reader's query with their RN16 (a 16-bit random number), only one tag, to which the reader echoes with its RN16, will be activated to send its EPC to the reader in every round of interrogation.

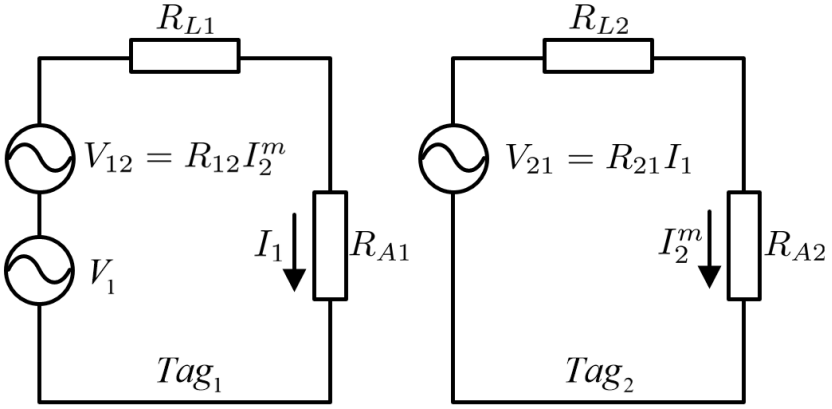
In Fig. 5.1, we present the corresponding circuit models of two tags under mutual coupling [109]. The upper plot is for the case when Tag 1 is activated, and the lower plot is for the case when Tag 2 is activated. In the circuit models, V_1 and V_2 represent the source voltages, and I_1 and I_2 are the source currents, when Tag 1 or Tag 2 is activated, respectively; R_{L1} and R_{L2} are the impedance of the microchip, and R_{A1} and R_{A2} are the impedance of the antenna input, of the two tags, respectively.

Theorem 5.1. *Consider two RFID tags under strong mutual coupling. If the tags have identical chip impedance and antenna input impedance, i.e., $R_{L1} = R_{L2}$ and $R_{A1} = R_{A2}$, the ratio of their equivalent source currents equals to the ratio of their equivalent source voltages. That is, we have $I_1/I_2 = V_1/V_2$.*

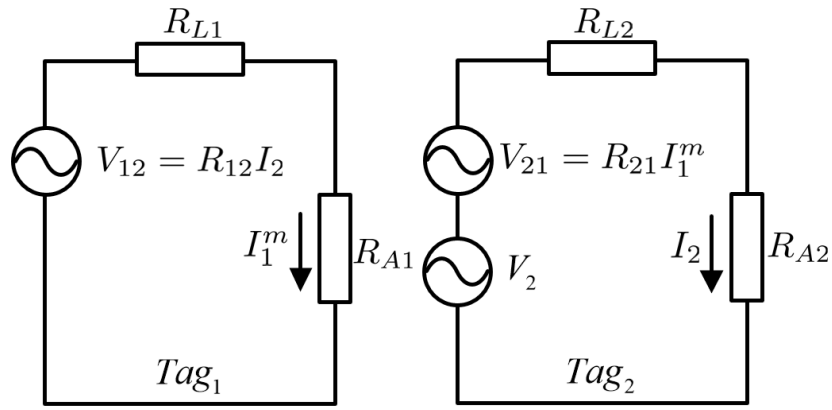
Proof. Consider the case when Tag 1 is activated by the reader. Due to induced coupling of the two tags' antennas, Tag 1's current I_1 will trigger a coupled voltage $V_{21} = R_{21} \cdot I_1$ in Tag 2. Here R_{21} is the mutual impedance of Tag 2 with respect to Tag 1. The coupled voltage V_{21} will then induce a current I_2^m in Tag 2, which next produces a coupled voltage $V_{12} = R_{12} \cdot I_2^m$ back at Tag 1, Here R_{12} is the mutual impedance in Tag 1 with respect to Tag 2. Assuming the two tags are of the same type, it follows that

$$I_1 \cdot (R_{L1} + R_{A1}) = I_1 \cdot R_0 = V_1 + R_{12} \cdot I_2^m \tag{5.3}$$

$$I_2^m \cdot (R_{L2} + R_{A2}) = I_2^m \cdot R_0 = R_{21} \cdot I_1, \tag{5.4}$$



(a) When Tag 1 is activated.



(b) When Tag 2 is activated.

Figure 5.1: The equivalent circuit model of two tags under mutual coupling.

where $R_0 = R_{L1} + R_{A1} = R_{L2} + R_{A2}$ is a constant. Assume the two tags have identical mutual impedance [109], i.e., $R_{12} = R_{21} = R_m$. Then current I_1 can be written as

$$I_1 = \frac{V_1}{R_0 - R_m^2/R_0}. \tag{5.5}$$

We can drive the same relationship for the case when Tag 2 is activated, as

$$I_1^m \cdot (R_{L1} + R_{A1}) = I_1^m \cdot R_0 = R_{12} \cdot I_2 \tag{5.6}$$

$$I_2 \cdot (R_{L2} + R_{A2}) = I_2 \cdot R_0 = V_2 + R_{21} \cdot I_1^m, \tag{5.7}$$

where I_1^m is the induced current in Tag 1 by the coupling voltage V_{12}. We can solve for the current I_2 in Tag 2 as

$$I_2 = \frac{V_2}{R_0 - R_m^2/R_0}, \tag{5.8}$$

Then we conclude from (5.5) and (5.8) that $I_1/I_2 = V_1/V_2$. □

Rewrite the complex current and voltages as $I_i = |I_i| \angle I_i$ and $V_i = |V_i| \angle V_i$, $i = 1, 2$, where $|\cdot|$ is the amplitude and $\angle$ is the phase angle. According to Theorem 5.1, we have

$$\angle I_1 - \angle I_2 = \angle V_1 - \angle V_2. \tag{5.9}$$

The measured phase angle by the reader is determined by the distance and the phase of the current that generates the tag response signal [72]. But the measured phase difference will be independent to the coupling voltage and mutual impedance. That is, *mutual coupling has a negligible impact on the phase difference*, although the phase angle itself is highly susceptible to the coupling effect, as shown in (5.5) and (5.8). Theorem 5.1 and the following analysis justify the feasibility of leveraging tag arrays for DOA estimation in the presence of mutual coupling effect.

We designed three experiments to validate the above analysis. The *first* experiment is to measure the phase angle from a tag, while placing another tag next to it at various distances. To assess the interference induced by mutual coupling, we also measure the ground truth phase angle when the second tag is absent. The measured phase errors (i.e., the difference between with or without the second tag) are presented in the upper plot of Fig. 5.2 for various distances between the two tags. It can be seen from the plot that the phase errors are all quite big until the second tag is placed at a large distance, e.g., 16 cm, from the target tag. Therefore, in order to avoid the large phase interference induced by mutual coupling, the tags should be placed at least 16 cm away from each other.

The *second* experiment is to measure the phase difference by placing the two tags in parallel on the same plane at various distances. The experiment is conducted as illustrated in

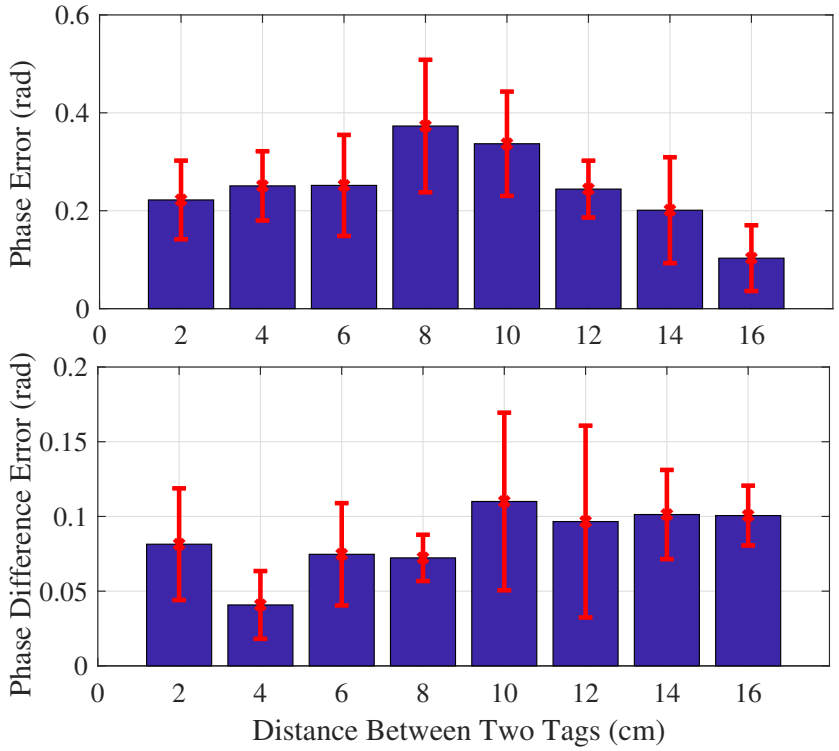


Figure 5.2: Impacts of mutual coupling on measured phase angle (the upper plot) and phase difference (the lower plot).

Fig. 5.3. In the first part of the experiment, we place Tag 1 at each of the locations, which are separated 2 cm apart. The phase angles from Tag 1 is measured at each of these locations. Then the ground truth phase difference is calculated by subtracting the phase angle at the left-most position from that measured at any other locations. This approach allows us to obtain the phase differences at various tag-tag distances without the mutual coupling effect. In the second part of the experiment, two tags are deployed: Tag 1 is fixed at the left-most location, while Tag 2 is put at each of the other locations. The phase differences under mutual coupling is then measured at different tag-tag distances and are plotted in the lower plot of Fig. 5.2. It can be seen that the phase difference errors are all smaller than 0.1 radian except when the distance is 10 cm. The phase difference errors are mainly due to the multipath effect and random noise; the impact of mutual coupling on phase difference is much weaker than the case of phase angles. This experiment validates Theorem 5.1.

It is worth noting that the above two experiments are different from the Tagyro scheme [23], where the change of phase difference is measured when a tag rotates around another fixed tag.

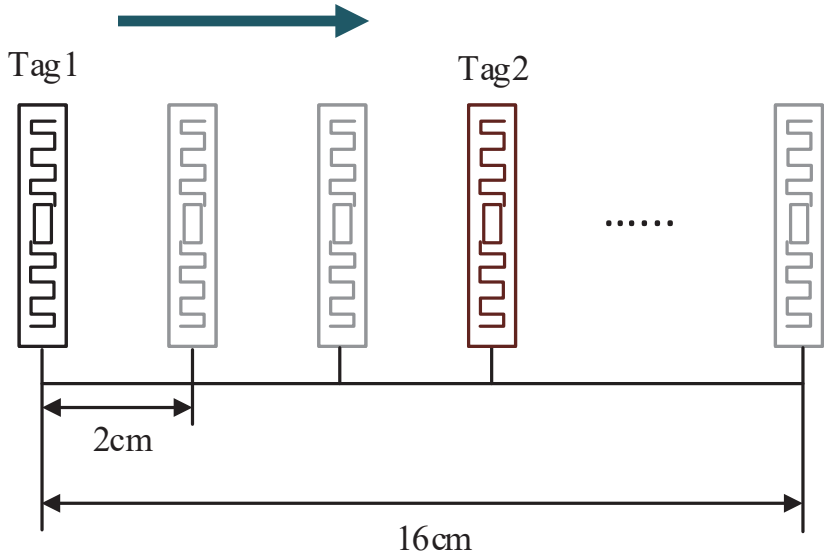


Figure 5.3: The setup of the second experiment for assessing the impact of mutual coupling on measured phase difference.

The relative orientation of the pair of tags has a big impact on the mutual impedance. Thus the mutual coupling effect of Tagyro varies with the different rotation angles of the tags. Due to this reason, Tagyro requires careful calibration with a hologram-based approach in order to translate phase offset into the tag array's orientation.

Due to the imperfections in tag production, the impedance of different tag could still be different even if the tags are of the same type and produced by the same manufacturer. We design the *third* experiment to assess the impact of different tag types and different tags of the same type on phase difference measurement. Specifically, we repeat the same experiment with three different types of tags and three different groups of tags of each type. The average phase difference error is summarized in Table 5.1. It can be observed from the table that, although tags used in each group are of different types, the phase difference error of each group are similar. This experiment validates that the effect of imperfect tag production is negligible on phase difference measurements.

Impact on the Success Rate of Sampling

Consider, for example, a ULA tag array. The reader keeps on interrogating the tags in the array for a certain period of time. Let n_i be the number of phase angle samples read from tag i in

Table 5.1: Impact of Different Types of Tags on Phase Difference Error

Type of RFID tag	Group A	Group B	Group C
ALN-9740	0.08 rad	0.06 rad	0.08 rad
SMARTRAC DogBone	0.11 rad	0.08 rad	0.09 rad
SMARTRAC ShortDipole	0.08 rad	0.07 rad	0.07 rad

the array. The *success rate of sampling* of tag i is defined as the ratio of n_i over the maximum number of samples collected from any of the tags in the ULA, i.e.,

$$\xi_i = \frac{n_i}{\max_i\{n_i\}}. \tag{5.10}$$

To measure the mutual coupling effect on tags' success rate of sampling, let P_r denote the received power at the reader, given by [110]

$$P_r = \left(\frac{\lambda}{4\pi l}\right)^4 P_r G_r^2 G_t^2 \frac{4R_A^2}{(R_L+R_A)^2+(X_L+X_A)^2}, \tag{5.11}$$

where P_r is the reader's transmit power, G_t and G_r are the antenna gains of the tag and the reader, respectively, R_A and X_A are tag antenna's radiation resistance and reactance, respectively, and R_L and X_L are tag chip's radiation resistance and reactance, respectively. To get a valid sample, the received power P_r should exceed the detection threshold P_{th}, i.e., $P_r \geq P_{th}$ [110]. Otherwise, the tag cannot be detected by the reader.

Theorem 5.2. *Assume the tag chip's impedance R_L is constant. If the tag antenna and chip's reactance satisfy $X_A = -X_L$, the received power at the reader P_r will be an increasing function of R_A when the tag-reader distance l is fixed.*

Proof. If the tag antenna's and chip's reactances satisfy $X_A = -X_L$ (assuming perfect tag production), we have from (5.11) that

$$P_r = \left(\frac{\lambda}{4\pi l}\right)^4 P_r G_r^2 G_t^2 \frac{4}{(R_L/R_A+1)^2}.$$

It can be easily verified that the received power P_r is an increasing function of R_A, when all other parameters are fixed. □

When the two tags are placed closer, the tag antenna's radiation resistance R_A will become smaller due to the mutual coupling effect [110]. According to Theorem 5.2, the reader's received power will be lower if the two tags are placed closer to each other. On the other hand, with mutual coupling, $(X_A + X_L)^2$ will not be zero anymore [110], which also reduces the received power as given in (5.11). In the tag array, mutual coupling could reduce the received powers of some tags, leading to a low success rate of sampling for such affected tags.

Fig. 5.4 presents the success rates of sampling of a ULA comprising five tags, placed at 2.1 m away from the reader antenna. We find that the success rates of sampling of all the tags are over 90% when the distance between tags is 6 cm, and the success rates of sampling of all tags are higher than 70% at a 4 cm tag intervals. However, when the tags are placed at 2 cm apart, the success rate of sampling of Tag 2 becomes lower than 10%. That is, this tag cannot be effectively detected on most channels, which generally happens for a tag placed at the center of the ULA, because the mutual coupling effect caused by the tags on both sides are strong. To ensure that all tags in the array be effectively sampled by the reader, the density of the tag array should not be too high. This observation motivates us to design a sparse array for localization.

5.3 The SparseTag System

5.3.1 Overview

Fig. 5.5 provides an overview of the proposed SparseTag system, where an RFID tag array and a reader with two antennas are utilized. We assume the position of the tag array is unknown (i.e., to be detected), while the locations of the two reader antennas are known as a prior. The main idea of the SparseTag design is to utilize a sparse tag array to detect the DOAs at the center of the array from both antennas. Then the position of the center of the array will be solved from the known locations of the two antennas and the estimated DOA values.

In typical applications, the tag array is attached to, or even woven into, a small object (e.g., a book, a tablet, or a shirt). The key challenge in the design of SparseTag is how to accurately

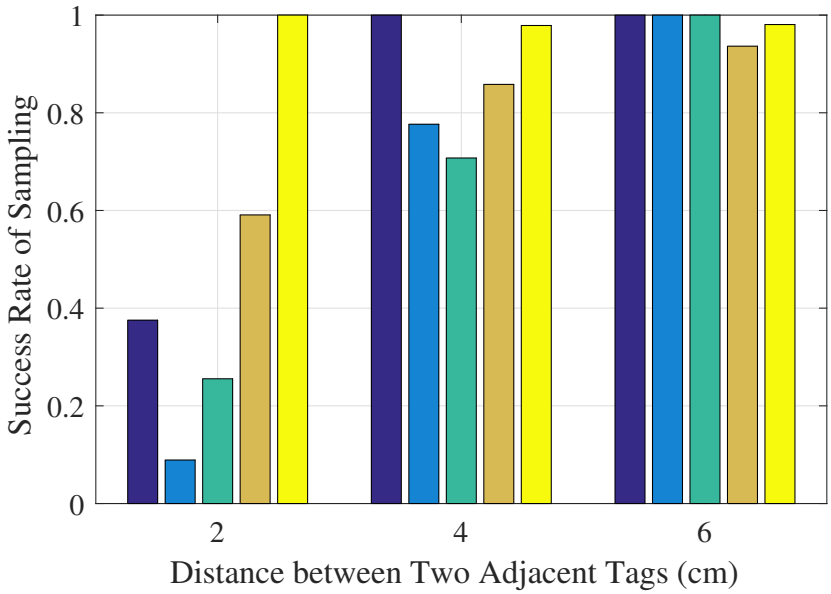


Figure 5.4: Impact of mutual coupling on the success rate of sampling, when five tags are placed at 2 cm, 4 cm, and 6 cm intervals.

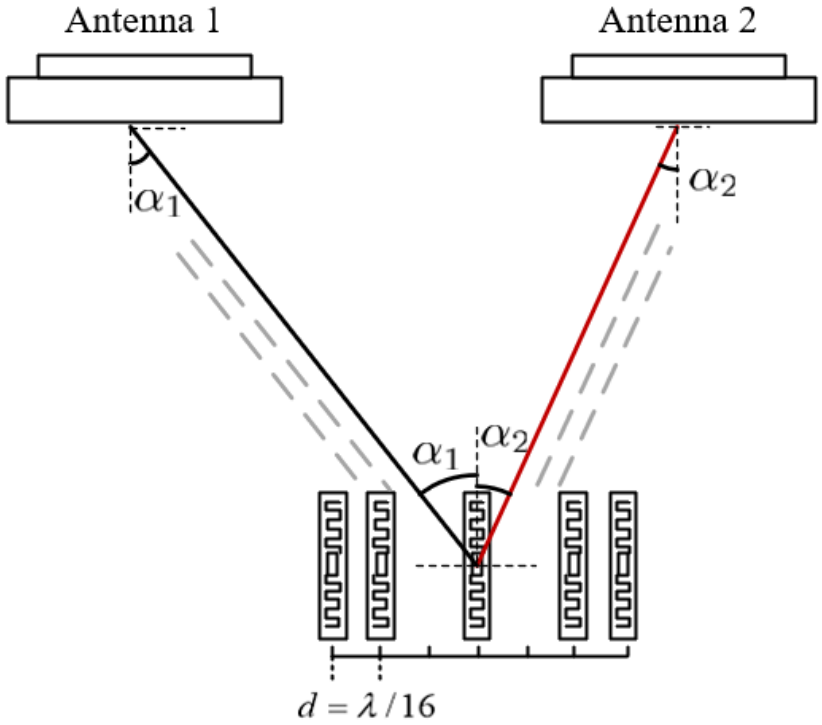


Figure 5.5: An overview of the proposed SparseTag system, comprising of a sparse tag array and a reader with two antennas. The antenna locations are known and the center of the tag array is to be localized.

estimate the DOAs by leveraging the sparse tag array. With a traditional ULA, the sensor element spacing should be smaller than half of the wavelength, while the MUSIC algorithm

can be applied to estimate the DOA [107]. As discussed in Section 5.2, a ULA may not be suitable for positioning a small object. This is because, for RFID systems operating in the 900 MHz band, half of a wavelength is already 16 cm. Furthermore, with an N-element ULA, the MUSIC algorithm only estimates up to $(N-1)$ DOAs. Usually, spatial smoothing is adopted to decorrelate uncorrelated sources, which takes half of the elements and consequently, the maximum number of estimated DOAs will be halved [116]. In this paper, a novel sparse RFID tag array structure is proposed to achieve high success rate of sampling for the tags in the array, while the minimum spacing of the tags can be as small as $\lambda/16$. Since usually a tag, e.g., an ALN-9740 tag, is about 1 cm wide, the minimum spacing of the tags should be no smaller than 1 cm such that the tags will not overlap with each other. Consequently, the minimum spacing of the proposed array structure is set to $\lambda/16$, which is roughly 2 cm for the 900 MHz band. Moreover, we derive the difference co-array of the sparse tag array, which can provide a higher DOA resolution for more accurate localization performance.

The proposed SparseTag system mainly comprises four modules, i.e., (i) Sparse Array Processing, (ii) Co-array Design, (iii) DOA Estimation, and (iv) Location estimation. We describe the design of each of the modules in the following.

5.3.2 Sparse Array Design

In order to adopt tag arrays to localize small objects, the number of tags, as well as the tag spacing, cannot be too big. To address these issues, we propose to adopt a sparse array that comprises of N tags with a nonuniform linear placement, which is quite different from the traditional ULA plus MUSIC approach [107].

Let the steering vector for direction α be denoted by $\vec{a}(\alpha)$, with elements $\exp\{j\frac{2\pi}{\lambda}d_i \sin\alpha\}$, where d_i is the location of tag i and λ is the wavelength of the carrier frequency. Assume there are D multipath components from the propagation environment, each having direction α_i and

power σ_i^2, $i = 1, 2, ..., D$. The received signal at time t can be written as

$$\begin{aligned} \vec{g}[t] &= \sum_{i=1}^{D} \vec{a}(\alpha_i) s_i[t] + \vec{n}[t] \\ &= \mathbf{A}\vec{s}[t] + \vec{n}[t], \end{aligned} \tag{5.12}$$

where $\mathbf{A} = [\vec{a}(\alpha_1), \vec{a}(\alpha_2), ..., \vec{a}(\alpha_D)]$ denotes the array manifold matrix, $\vec{s}[t] = [s_1[t], s_2[t], ..., s_D[t]]^T$ denotes the source signal vector, and $\vec{n}[t]$ is the additive white noise vector. Assuming the multipath components are temporally uncorrelated, the source autocorrelation matrix will then assume a diagonal structure. Considering the second-order information of the received signal $\vec{g}(t)$, its covariance matrix, denoted by $\mathbf{R}_{gg}$, can be derived as

$$\begin{aligned} \mathbf{R}_{gg} &= \mathbb{E}[\vec{g}(t)\vec{g}(t)^H] \\ &= \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma_n^2\mathbf{I} \\ &= \sum_{i=1}^{D} \sigma_i^2 \vec{a}(\alpha_i)\vec{a}(\alpha_i)^H + \sigma_n^2\mathbf{I}. \end{aligned} \tag{5.13}$$

We next vectorize the $\mathbf{R}_{gg}$ in (5.13) to derive the measurement vector, which is given by

$$\begin{aligned} \vec{z} &= \text{vec}(\mathbf{R}_{gg}) \\ &= \text{vec}\left[\sum_{i=1}^{D} \sigma_i^2 \vec{a}(\alpha_i)\vec{a}(\alpha_i)^H\right] + \sigma_n^2\vec{1}_n \\ &= (\mathbf{A}^* \odot \mathbf{A})\vec{p} + \sigma_n^2\vec{1}_n, \end{aligned} \tag{5.14}$$

where $\vec{p} = [\sigma_1^2, \sigma_2^2, ..., \sigma_D^2]^T$, $\vec{1}_n = [\vec{e}_1^T, \vec{e}_2^T, ..., \vec{e}_N^T]^T$, and $\vec{e}_i$ is a column vector whose ith element is "1" and all other elements are "0." Here the measurement vector is regarded as the signal received at an array with a manifold of $(\mathbf{A}^* \odot \mathbf{A})$ [111], where $\odot$ represents the Khatri-Rao (KR) product. The matrix $(\mathbf{A}^* \odot \mathbf{A})$ can be regarded as the manifold of a longer array, whose antenna positions are determined by the different values in the set $\{\vec{x}_i - \vec{x}_j\}$, $1 \leq i$ and $j \leq N$, where $\vec{x}_i$ is the location vector of Tag i. This new array is termed the *difference co-array* [111].

In SparseTag, DOA is estimated with the difference co-array, which can effectively exploit the second-order statistics of received signal for an increased DOF.

5.3.3 Difference Co-array Design

In the following, we first present several basic definitions related to the difference co-array. We then present the design of the difference co-array for the sparse tag array used in SparseTag.

Definitions

Definition 1. *(Difference Co-Array). Consider a sparse, N-element tag array. Let $\vec{x}_i$ be the location vector of Tag i. The difference co-array of the sparse array is defined as [111]*

$$\mathcal{D} = \{\vec{x}_i - \vec{x}_j, 1 \leq i, j \leq N\}. \tag{5.15}$$

The difference co-array can be regarded as a new array, where the tags are placed at the locations given in the set $\mathcal{D}$. In addition, the values of the cross correlation elements in the covariance matrix of the received signal by the sparse tag array are determined by the number of elements in the difference co-array, which is helpful to improve the number of estimated DOAs.

Definition 2. *(Restricted Array). A sparse, N-element tag array is a restricted array if its difference co-array is hole-free [112].*

If the difference co-array is hole-free, it is also a ULA. Therefore, the traditional subspace based MUSIC algorithm can be employed to estimate DOA using a hole-free difference co-array. For instance, the tags are placed at the positions given by the set S, which is given by

$$\mathcal{S} = \{m \cdot d,\ m = 1, 2, 4\}. \tag{5.16}$$

where d is the minimum spacing between tags. The corresponding difference co-array can be derived as

$$\mathcal{D} = \left\{ -\vec{3}, -\vec{2}, -\vec{1}, \vec{0}, \vec{1}, \vec{2}, \vec{3} \right\}. \tag{5.17}$$

Although the position $3d$ is missing in this sparse array (see (5.16)), there is no missing vector in the difference co-array set $\mathcal{D}$ (i.e., it includes all the vectors from $-\vec{3}$ to $\vec{3}$, see (5.17)). Consequently, this array is still useful for DOA estimation using the MUSIC algorithm.

Definition 3. *(Degree of Freedom (DOF): The DOF of a sparse array is the cardinality of its difference co-array $\mathcal{D}$ [111].*

The DOF of a sparse array can be derived by the cardinality of its difference co-array $\mathcal{D}$, which indicates the maximum number of DOAs that can be estimated.

Definition 4. *(Weight Function). For a sparse, N-element tag array, its weight function $w(\vec{d})$ is defined as the number of tag pairs that can achieve the difference co-array element $\vec{d}$. The weight function is given by [111]*

$$w(\vec{d}) = \left| \left\{ (\vec{x}_i, \vec{x}_j) | \vec{x}_i - \vec{x}_j = \vec{d} \right\} \right|, \ \vec{d} \in \mathcal{D}. \tag{5.18}$$

The weight function indicates the how serious the mutual coupling effect is, which is helpful for our proposed sparse tag array.

Difference Co-array

Assume N is an odd number. Then the tag placement in the N-element sparse tag array are given in the set $\mathcal{S}$, which is

$$\begin{aligned} \mathcal{S} = \ & \{ m \cdot d, \ m = 1, ..., (N+1)/2 - 1, \\ & (N+1)/2 + 1, (N+1)/2 + 3, ..., N+2 \}, \end{aligned} \tag{5.19}$$

where the tag minimum spacing is set to $d = \lambda/16$; such small spacing allows us to use a small-sized tag array to localize small objects. The proposed sparse tag array has a symmetric structure; its left and right halves have the same tag spacing arrangement. In addition, the gaps from the two tags placed at positions $(\frac{N+1}{2} - 1)d$ and $(\frac{N+1}{2} + 3)d$ to the tag placed at the center of the array (i.e., at position $(\frac{N+1}{2} + 1)d$) is both $2d$. Thus the proposed array is a sparse array.

The sparse tag array used in SparseTag has the following three key advantages. First, its symmetric structure helps to suppress the mutual coupling effect, which usually limits the performance of traditional tag arrays. The reduced mutual coupling leads to less interference in measured phase difference and thus, higher accuracy in the estimation of DOAs. Second, the sparse structure also helps to mitigate the degradation of the success rate of sampling of the tags in the array, specifically, the tag at the center of the array. Third, it allows us to use tag arrays with a smaller physical dimension. Such smaller-sized tag arrays help to improve the DOA resolution and are easier to deploy, such as attached to small objects or woven into clothing.

The difference co-array corresponding to the proposed sparse array is given by the following placement set S_d.

$$\mathcal{S}_d = \{m \cdot d,\ m = -(N+1), -N, ..., N, (N+1)\}. \tag{5.20}$$

Theorem 5.3. *The proposed sparse array is a restricted array. That is, it is a hole-free difference co-array.*

Proof. The difference co-array of the proposed sparse array is a ULA. It is easy to verify that it is a hole-free difference co-array, with the given placement set $\mathcal{S}_d$. Thus we conclude that it is a restricted array. □

Corollary 5.3.1. *The proposed sparse array's co-array is the same as that of an* $(N+2)$ *ULA.*

Proof. The proposed sparse array has the same left-most and right-most tag positions at d and $(N+2)d$, respectively, as that of an $(N+2)$ ULA. In addition, both arrays are restricted arrays. Therefore, we conclude that the proposed sparse array has the same co-array as that of the $(N+2)$-element ULA. □

Theorem 5.4. *The DOF of the sparse N-element array is $2N + 3$.*

Proof. The proposed sparse array is given by the position set $\mathcal{S}$, and the cardinality of its difference co-array $\mathcal{S}_d$ is $2N + 3$. Therefore we conclude that the DOF of the sparse array is $2N + 3$ according to Definition 3. □

Theorem 5.5. *The weight function of the proposed N-element sparse array is $w(\vec{d} = \vec{0}) = N$ and $w(\vec{d} = \vec{1}) = N - 3$.*

Proof. When $\vec{x}_i = \vec{x}_j$, we have the case $\vec{d} = \vec{0}$. This case occurs N times for an N-element array, i.e., when $i = j = 1, 2, ..., N$. Therefore we have $w(\vec{d} = \vec{0}) = N$ according to Definition 4.

Furthermore, consider two different subarrays given by sets $\mathcal{S}_l = \{m \cdot d, m = 1, 2, ..., \frac{N+1}{2} - 1\}$ and $\mathcal{S}_r = \{m \cdot d, m = \frac{N+1}{2} + 3, \frac{N+1}{2} + 4, ..., N + 2\}$, respectively. The case $\vec{d} = \vec{1}$ takes place for $(\frac{N+1}{2} - 2)$ times in each subarray. Furthermore, for the subarray given by set $\mathcal{S}_c = \{m \cdot d, m = \frac{N+1}{2} - 1, \frac{N+1}{2} + 1, \frac{N+1}{2} + 3\}$, the case $\vec{d} = \vec{1}$ does not arise at all. It follows that $w(\vec{d} = \vec{1}) = (\frac{N+1}{2} - 2) * 2 = N - 3$. □

We make the following observations from the above theorems and corollary.

- The proposed N-element sparse array has the same DOF as an $(N + 2)$-element ULA. Using the proposed sparse array, we can achieve a higher maximum number of estimated DOAs than using a ULA with the same amount of tags and the MUSIC algorithm.

- Using the proposed sparse array can achieve a higher sampling rate of the tags. This is because its weight function, i.e., $w(\vec{d} = \vec{1}) = N - 3$, is smaller than that of an $(N + 2)$-element ULA. In Fig. 5.6, we compare a 7-tag ULA with a 5-tag proposed sparse array. In each figure, the upper plot shows the placement of the tags, while the lower plot presents the corresponding weight function $w(n)$. It can be seen that the 5-tag sparse array has the same DOF, i.e., 13, as the 7-tag ULA, since they share the same difference co-array. Furthermore, it shows that $w(\vec{1}) = 2$ for the 5-tag sparse array and $w(\vec{1}) = 6$ for the 7-tag ULA.

- There are other types of sparse arrays, e.g., the co-prime array [113], nested array [111], and super nested array [114], which can also achieve a larger DOF than the proposed sparse array. However, such arrays may not be suitable for the deployment of tag arrays. This is because such arrays all require a relatively larger physical space, which may not be available for small objects. Moreover, such arrays' structures are not symmetric. Therefore, they may incur more serious mutual coupling among the tags, leading to large interference in the RFID phase and phase difference samples.

5.3.4 Estimation of DOA

The difference co-array of the proposed sparse array is then leveraged for DOA estimation. A spatial smoothing based method is employed, which is different from the existing approach that utilizes spatial smoothing to mitigate correlated sources [111]. The SparseTag approach constructs an observation matrix for the difference co-array, which does not require using high-order cumulative signals.

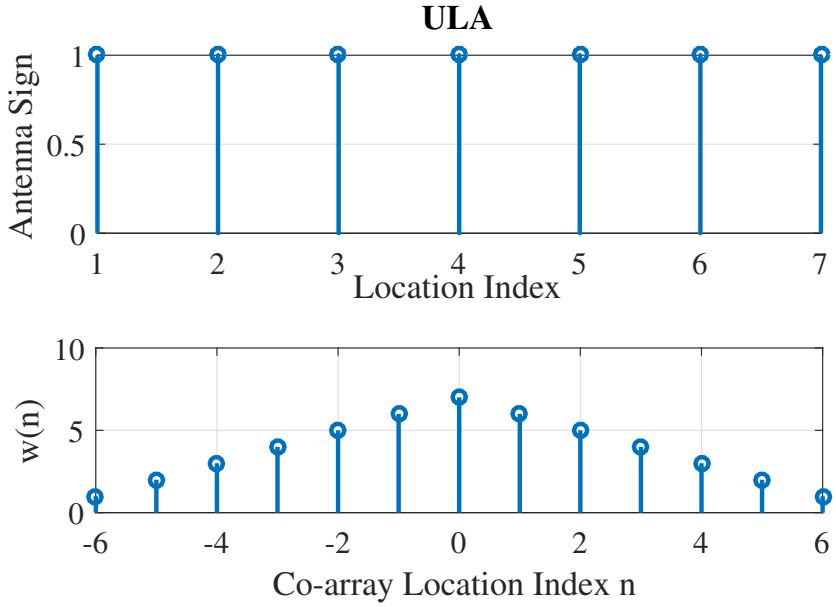
Specifically, we first derive the array manifold $(\mathbf{A}^* \odot \mathbf{A})$ following (5.14), which has a dimension $N^2 \times D$. According to Theorem 5.4, we next construct a matrix $\mathbf{B}$ with dimension $(2N+3) \times D$ by removing the repeated rows from the array manifold. Next, we sort this constructed matrix to ensure that row i corresponds to the tag position $(-N-1+i)d$ in the proposed difference co-array. Then we obtain a new vector $\vec{y}$, which is written as

$$\vec{y} = \mathbf{B}\vec{p} + \sigma_n^2 \vec{e}, \tag{5.21}$$

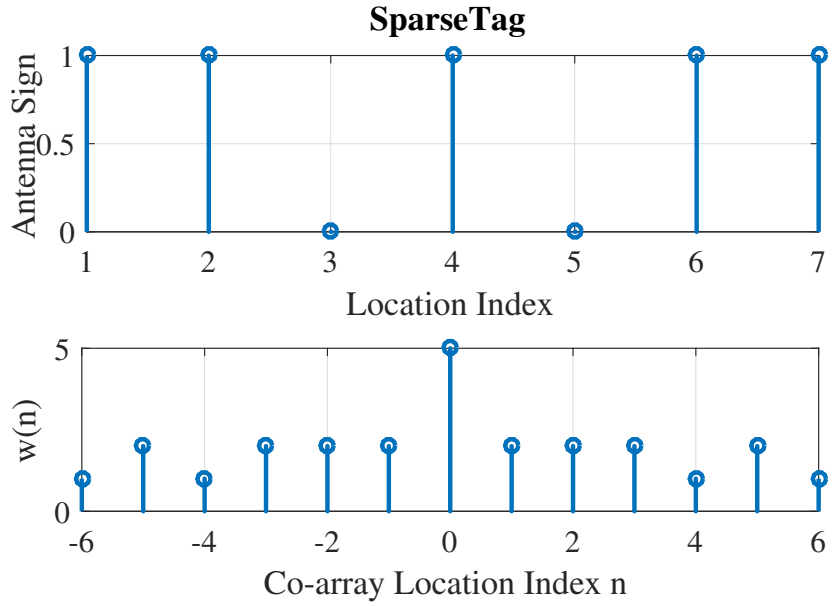
where $\vec{e} \in \Re^{(2N+3)\times 1}$ is a vector whose $(N+1)$th element is "1" and all other elements are "0."

Following the placement set (5.20), we divide the co-array into $(N+1)$ overlapping subarrays, each having $(N+1)$ elements, while the ith subarray is given by the following placement set:

$$\mathcal{S}(i) = \{(-i+1+m)\cdot d,\ m = 0, 1, ..., N\}. \tag{5.22}$$



(a) A 7-tag ULA and its weight function $w(n)$. Upper: antenna sign is 1 means a tag is placed at the corresponding location; Lower: the weight function.



(b) A 5-tag sparse array and its weight function $w(n)$. Upper: antenna sign is 1 means a tag is placed at the corresponding location; Lower: the weight function.

Figure 5.6: A 7-tag ULA versus a 5-tag sparse array.

Let $\vec{y}(i)$ be a new vector for subarray i that comprises the same elements of $\vec{y}$ ranging from the $(N+1-i+1)$th element to the $(2N+1-i+1)$th element:

$$\vec{y}(i) = \mathbf{B}(i)\vec{p} + \sigma_n^2 \vec{e}(i), \tag{5.23}$$

where $\mathbf{B}(i)$ is a matrix with dimension $(N+1) \times D$, comprising the same rows of $\mathbf{B}$ ranging from the $(N+1-i+1)$th row to the $(2N+1-i+1)$th row; and $\vec{e}(i)$ is a vector whose ith element is "1" and all other elements are "0." Consequently, the spatially smoothed matrix $\mathbf{R}_s$ is obtained as

$$\mathbf{R}_s = \frac{1}{N+1} \sum_{i=1}^{N+1} \vec{y}(i)\vec{y}(i)^H. \tag{5.24}$$

We then utilize $\mathbf{R}_s$ to estimate DOA. With our approach, N DOAs can be estimated, which is considerably larger than what can be obtained with the MUSIC plus ULA approach (i.e., $(N-1)/2$). SparseTag incorporates a directional antenna for increased range. The line-of-sight (LOS) component is dominant and a strong incident wave. The proposed sparse array can achieve a higher angle resolution than the existing approach.

5.3.5 Location Estimation with DOAs

The proposed SparseTag system comprises a tag array and a reader with two directional antennas, each of which operates on 50 channels in the 900 MHz band and samples phase angle of the received tag response. Some channel information may not be reliable due to the multipath propagation.

Fig. 5.7 presents the phase angles collected from a five-tag sparse array from the 50 channels. It can be seen the phase difference of two adjacent tags on most channels are similar. This is due to the small distance between the pair of tags (e.g., 2 cm or 4 cm); and the 0.5 MHz change of channel frequency caused by channel hopping can hardly cause a sufficient change in phase. We also find from Fig. 5.7 that the phase difference collected from some channels are highly different from others. Such difference is caused by the multipath effect on different channels. Some channels are more susceptible to the multipath effect; so the phase angles collected from such channels should be filtered out before DOA estimation.

To address this issue, SparseTag adopts a channel selection procedure. Denote $\phi_{(i,f_m)}(t)$ as the phase angle sampled from Tag i on channel f_m at time t. The phase difference between

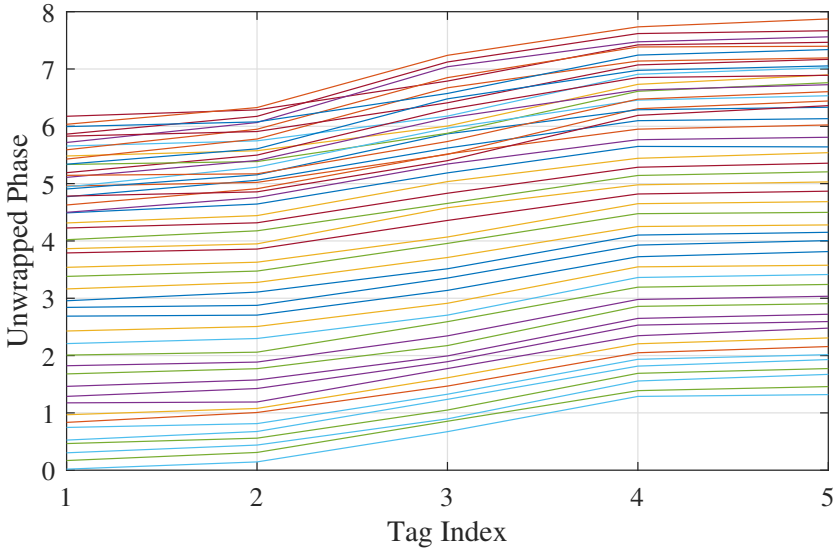


Figure 5.7: Phases angles sampled from a 5-tag sparse array over 50 channels (each line corresponds to a different channel).

Tag i and Tag $i+1$ at time t, denoted by $\Delta_{(i,f_m)}(t)$, is given by

$$\Delta_{(i,f_m)}(t) = \phi_{(i+1,f_m)}(t) - \phi_{(i,f_m)}(t),\ i = 1, 2, ..., N-1. \tag{5.25}$$

We select the medium value of all the phase differences from all the channels for robustness, since in many cases only a few channels are impaired. After selecting the right channel, we recalculate the phase angles of all the tags in the array. Using the Tag 1 phase angle as a reference and assuming $\phi_{(1,f_m)}(t) = 0$ at time t, the phase value of Tag i is

$$\phi_{(i,f_m)}(t) = \phi_{(i-1,f_m)}(t) + \Delta_{(i-1,f_m)}(t),\ i = 2, 3, ..., N. \tag{5.26}$$

The received signal is next reconstructed as

$$\hat{g}(t) = [e^{(j(2\pi - \phi_{(1,f_m)}(t)))}, ..., e^{(j(2\pi - \phi_{(N,f_m)}(t)))}]. \tag{5.27}$$

Note that we have the terms $(2\pi - \phi_{(i,f_m)}(t))$ in (5.27) due to the reader operation of the phase angle. Two DOAs are estimated using multiple snapshots of received signal (each comprising samples from all the 50 channels and each tag), one for each reader antenna. In Fig. 5.8, we plot

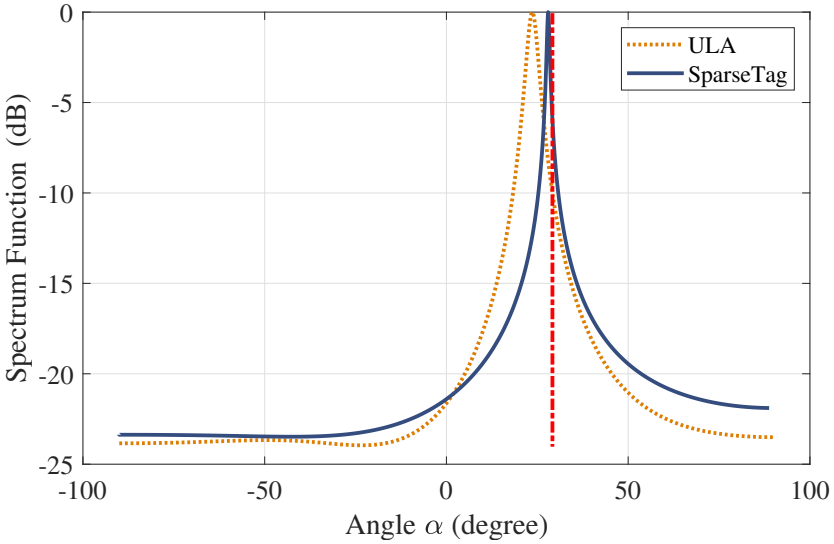


Figure 5.8: DOA estimation results obtained using SparseTag and ULA. The red vertical dashed line marks the ground truth of 28°.

the power spectrum density obtained by SparseTag and ULA arrays from the same experimental setting. The ground truth of DOA is marked by the red vertical line in the figure, which indicates 28°. Fig. 5.8 shows that the peak of the SparseTag curve is considerably sharper and closer to the ground truth than the peak of the ULA curve obtained using the MUSIC algorithm. DOA estimation with SparseTag is more accurate than with ULA, because SparseTag achieves larger DOFs than ULA.

The center of the tag array can be derived from the two estimated DOAs and the known coordinates of the two antennas. Consider a coordinate system where the direction of the tag array is the x-axis and the y-axis be perpendicular to the tag array. Assume (x_i, y_i) is the known coordinates of antenna i, $i = 1, 2$, and let (x_c, y_c) denote the coordinates of the center of the tag array. The two DOAs and the coordinates satisfy the following conditions (see Fig. 5.5).

$$\cot(\alpha_1) = \frac{y_c - y_1}{x_c - x_1} \tag{5.28}$$

$$\cot(\alpha_2) = \frac{y_c - y_2}{x_c - x_2}. \tag{5.29}$$

The cotangent function in (5.28) and (5.29) is given by $\cot(\alpha) = \cos(\alpha)/\sin(\alpha)$. We then solve (5.28) and (5.29) for the coordinates of the center position of the RFID tag array (x_c, y_c),

which is given by

$$x_c = \frac{x_1 \cot(\alpha_1) - x_2 \cot(\alpha_2) + y_2 - y_1}{\cot(\alpha_1) - \cot(\alpha_2)} \tag{5.30}$$

$$y_c = \frac{(x1 - x2) \cot(\alpha_1) \cot(\alpha_2) + y_2 \cot(\alpha_1) - y_1 \cot(\alpha_2)}{\cot(\alpha_1) - \cot(\alpha_2)}. \tag{5.31}$$

5.4 Experimental Validation

5.4.1 System Implementation and Experiment Setup

We develop an implementation of SparseTag using a commodity Impinj R420 RFID reader equipped with two circular polarized antennas and different types of RFID tags. When scanning the tags, the reader hops among 50 channels in the range of 902.5 MHz to 927.5 MHz, as required by the FCC. A Lenovo Thinkpad S3 laptop is used to control the reader and process the collected data. The RFID reader samples channel related data from received tag responses such as time stamp, phase angle, RSSI, and Doppler shift using a Low-level Reader Protocol (LLRP) [38]. Furthermore, we build the RFID tag arrays using three different types of passive tags, including ALN-9740, SMARTRAC DogBone, and SMARTRAC ShortDipole.

Extensive experiments are conducted in two different environments, including a 7.5×5.6 m^2 computer laboratory and an 8×2.4 m^2 anechoic chamber, which are illustrated in Fig. 5.9. The computer lab is a more cluttered environment with computers and furniture, which cause the multipath propagation of RFID signals. We also try to introduce more severe multipath effect by placing chairs in the LOS path between the tag array and antennas. In the anechoic chamber setup, most multipath effects are eliminated due to the special absorbing material mounted on the wall, ceiling, and floor. In the experiments, we mark various positions on the floor, which are considered as ground-truth. The tagged object (e.g., a book) is held by an easel to be in the same horizontal plane as the two reader antennas. As discussed, the target of localization is the center of the tag array. The same experiments are conducted using the ULA tag array with identical hardware and environment setup to assess the strengths of the proposed sparse tag array.

(a) The computer lab setup. (b) The anechoic chamber setup.

Figure 5.9: The setup of two experimental scenarios for SparseTag performance evaluation.

5.4.2 Evaluation in Different Localization Scenarios

We conduct experiments in both the lab and anechoic chamber environments. In Fig. 5.10 and Fig. 5.11, we plot the cumulative distribution function (CDF) of the DOA and location errors obtained by SparseTag with a 5-tag sparse array. The median error of DOA estimation in the anechoic chamber scenario is $1.125°$, while the median error in the computer lab scenario is $1.872°$. Fig. 5.10 also shows that the maximum error in an anechoic chamber is only $4.024°$. The angle estimation accuracy is higher when the system is tested in the anechoic chamber than the computer lab experiments, because the multipath effect is eliminated in the anechoic chamber setup. Fig. 5.11 also shows that the location error in the anechoic chamber scenario is smaller than that in the computer lab scenario. The median location error in the anechoic chamber environment is 3.419 cm, and the median location error in the computer lab environment is 5.012 cm.

To validate the performance of SparseTag in an environment with stronger multipath effect, we place some chairs as obstacles in the LOS path between the tag array and the reader antennas. The mean localization errors of all the three scenarios are plotted in Fig. 5.12, where both the SparseTag errors and the ULA errors are provided. It is shown in the figure that, the mean estimation error of the 5-tag SparseTag array in the rich multipath environment is 5.637 cm, while the mean error in the typical Computer Lab environment is 5.158 cm. For the 5-tag

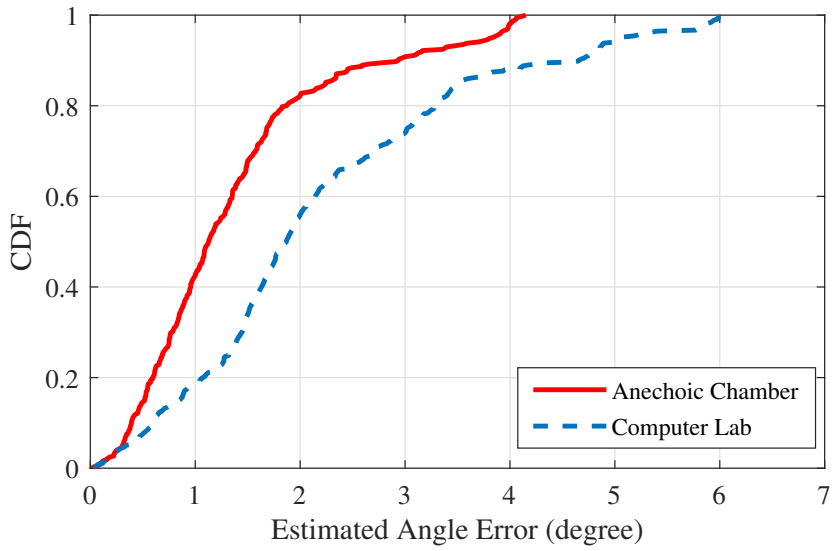


Figure 5.10: CDFs of DOA errors achieved by SparseTag with a 5-tag sparse array in the computer lab and anechoic chamber scenarios.

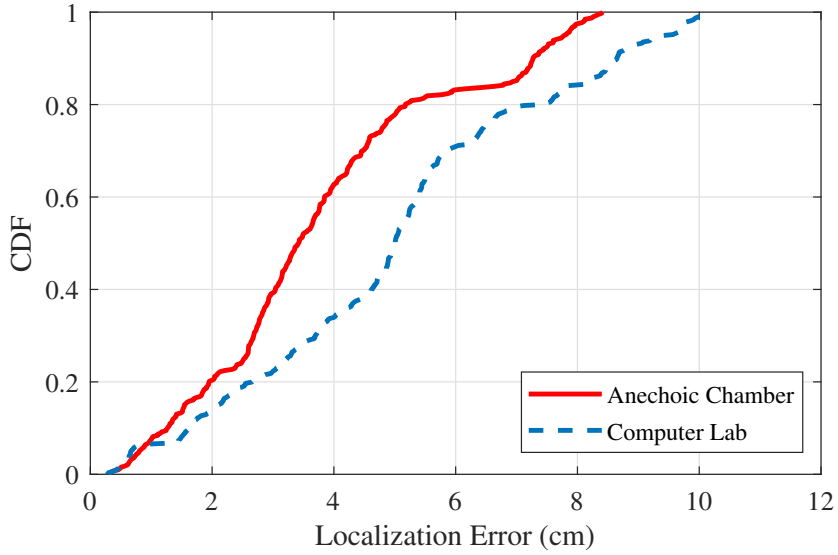


Figure 5.11: CDFs of localization errors achieved by the 5-tag SparseTag in the computer laboratory and anechoic chamber scenarios.

ULA array, the errors in the same environments increase to 8.967 cm and 7.450 cm, respectively. These results validate that the proposed spare array is more robust to the multipath effect than the ULA array with the same number of tags.

5.4.3 Comparison with Baseline Scheme

To validate the strengths of the proposed sparse tag array, we conduct more experiments to compare SparseTag with the traditional ULA plus MUSIC localization system [99, 107]. Fig. 5.13

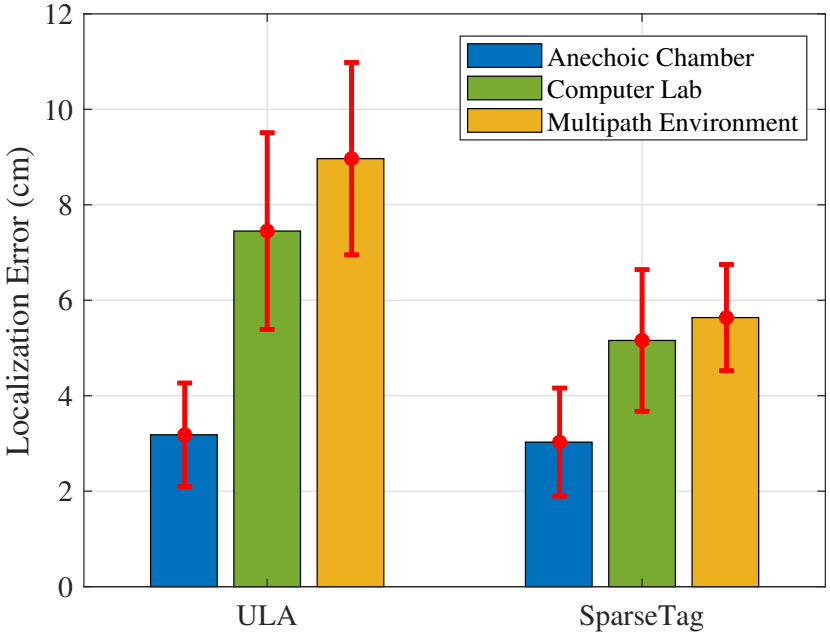


Figure 5.12: Mean localization errors achieved by the 5-tag ULA array and SparseTag array in three different scenarios.

presents a comparison of SparseTag with ULA in the computer lab environment. The CDFs of DOA errors obtained with a 5-tag ULA and a 5-tag SparseTag systems are plotted. We find that the maximum estimated DOA error of ULA is $9.198°$, while the maximum DOA error of SparseTag is $6.161°$. The median errors for ULA and SparseTag are $2.909°$ and $1.831°$, respectively. In addition, 90% of SparseTag estimated DOA errors are below $5°$. We conclude that SparseTag is more accurate for DOA estimation than ULA, because SparseTag achieves a higher angle resolution than ULA. With the spatially smoothed matrix $\mathbf{R}_s$ in (5.24), the number of estimated DOAs is more than that of ULA with the same number of tags, which means the sparse tag array performs better in rich multipath environments than the ULA array.

Fig. 5.14 presents the CDFs of localization errors obtained with the 5-tag SparseTag and 5-tag ULA. The same localization estimation method is used with the two different tag arrays in the same environment. We find that the median error of SparseTag is 4.985 cm, while the median error of ULA is 7.611 cm. Usually an UHF passive tag is about 10 cm long. For instance, the ALN-9740 tag used in our experiments is 98.2 mm $\times$ 12.3 mm. The SparseTag's median error is about half of the tag length; therefore it is sufficiently accurate for many practical applications. Fig. 5.14 also shows that the maximum error of SparseTag is 10.114 cm,

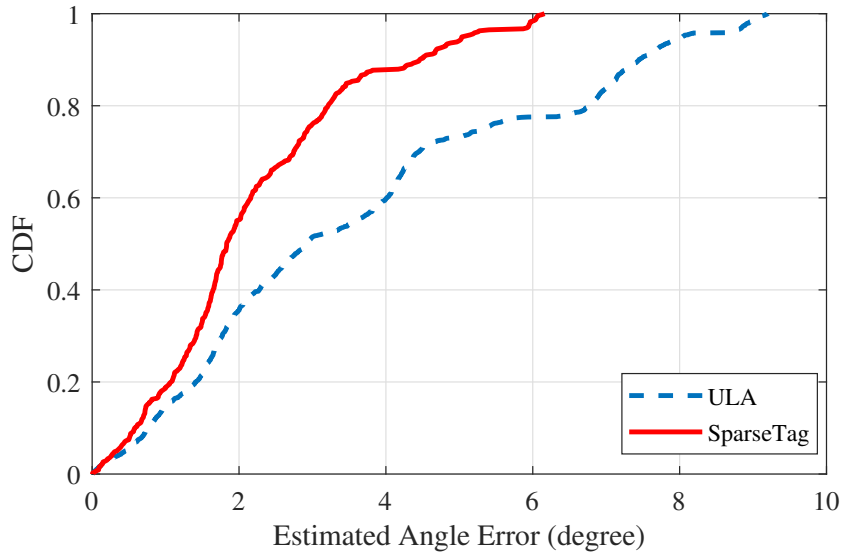


Figure 5.13: CDFs of DOA errors achieved by a 5-tag SparseTag and a 5-tag ULA in the computer lab experiment.

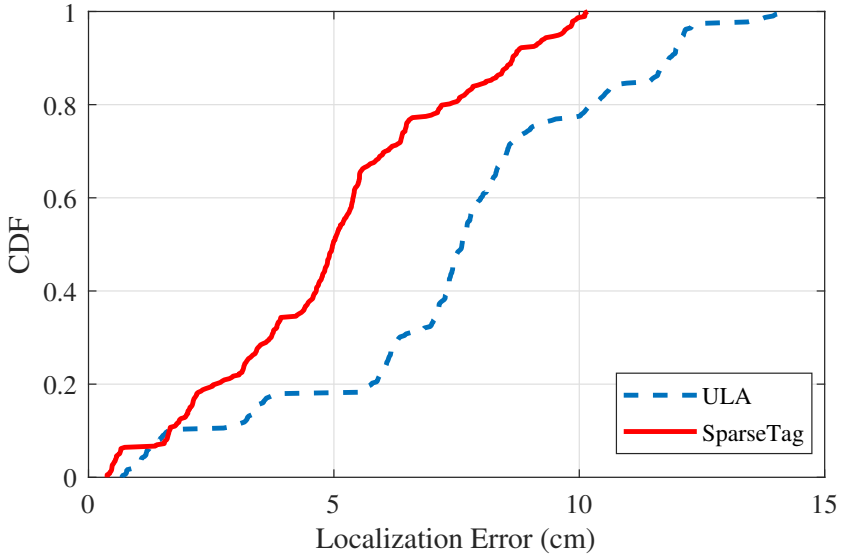


Figure 5.14: CDFs of localization errors achieved by a 5-tag SparseTag and a 5-tag ULA in the computer lab experiment.

which is much smaller than the ULA's maximum error. Thus it is validated that SparseTag can achieve a higher accuracy of localization than ULA.

We also compare the proposed system with an existing RFID tag array based localization technique using a mobile antenna [100] in the same rich multipath environment where the results in Fig. 5.12 are obtained. Rather than leveraging two antennas, a single mobile antenna is utilized for positioning of 5-tag ULA array and the sparse array. The mean estimation errors are presented in Fig. 5.15. As the figure shows, the ULA's localization error decreases from

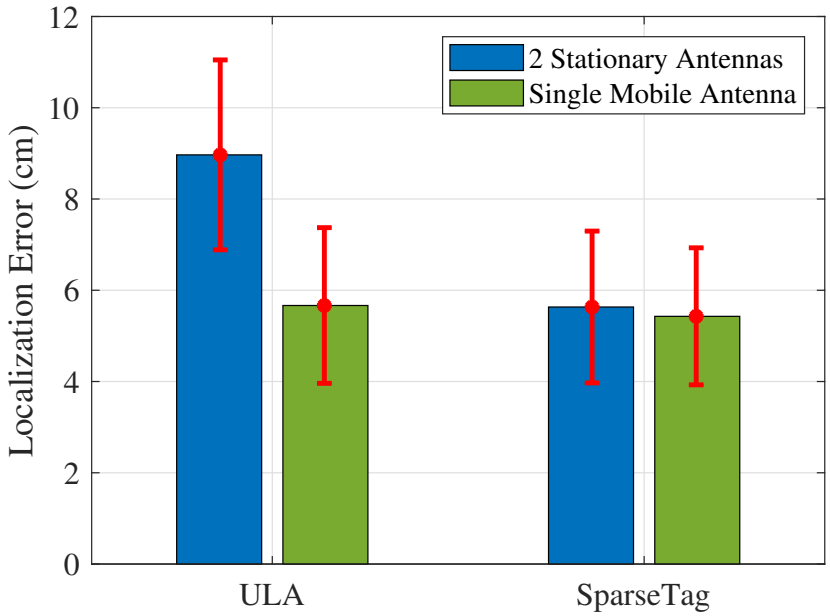


Figure 5.15: Mean localization errors obtained by the 5-tag ULA array and the SparseTag array with two stationary antennas and with a single mobile antenna.

8.967 cm to 5.666 cm, when the mobile antenna is leveraged in the system. This is because, the mobile antenna can be considered as multiple virtual antennas, which help to mitigate the multipath effect. However, for SparseTag, the improvement in accuracy brought about by the use of the mobile antenna is not obvious. This experiment result indicates that the multipath effect has already been effectively mitigated by using the proposed 5-tag sparse array, so use of the mobile antenna does not further improve the localization accuracy. We thus conclude that the proposed SparseTag system achieves high accuracy without using moving antennas, making it easier to deploy and more adaptable.

5.4.4 Impact of System Design Factors

To further assess the proposed system, we conduct more experiments on the influence of several design factors. Since directional antennas are used in our experiments, we also evaluate how the relative angle of the directional antenna affects the estimation results. Fig. 5.16 shows the estimation errors for different antenna angles, including -30°, -15°, 0°, 15°, and 30°, where 0° means the antenna directly faces the tag array. From Fig. 5.16, we can see that the estimation errors at different angles are all around 2°, and the error does not increase as the angle of the

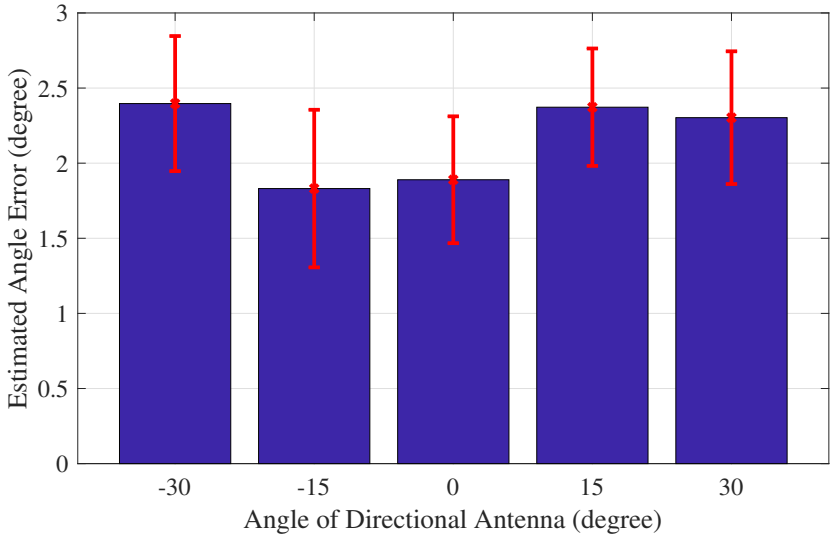


Figure 5.16: Impact of the angle of the directional antenna.

antenna is changed. This experiment shows that the estimated DOA is not seriously affected by the relative angle of the directional antenna.

We also examine the effect of the number of snapshots on DOA estimation. Recall that each snapshot comprises samples from all the 50 channels and from each tag. Fig. 5.17 shows that the DOA error obtained by one to 10 snapshots. It can be seen that when there are less than three snapshots, the DOA error is about 3°, while the error remains at about 2° with five or more snapshots. This is because only one or two snapshots cannot effectively remove the white noise in the RFID signals. If the number of snapshots is larger than nine, the effect of noise can be mostly removed using a spatial smoothing based method. As a result, we choose 10 snapshots in our SparseTag system.

Fig. 5.18 illustrates the impact of the difference between the heights of the reader antennas and the tag array. The height difference is represented by the angle between the horizontal plan and the line connecting the antenna and the center tag, as illustrated in Fig. 5.19. When the sparse tag array is not on the same horizon plane as the antennas, the DOA estimation error will increase quickly. This is because SparseTag mainly focuses on 2D localization; when the antennas are at a different height from the tag array, an additional phase offset will be introduced. The phase offset is related to the height difference, and so the DOA error will increase as the height difference becomes larger.

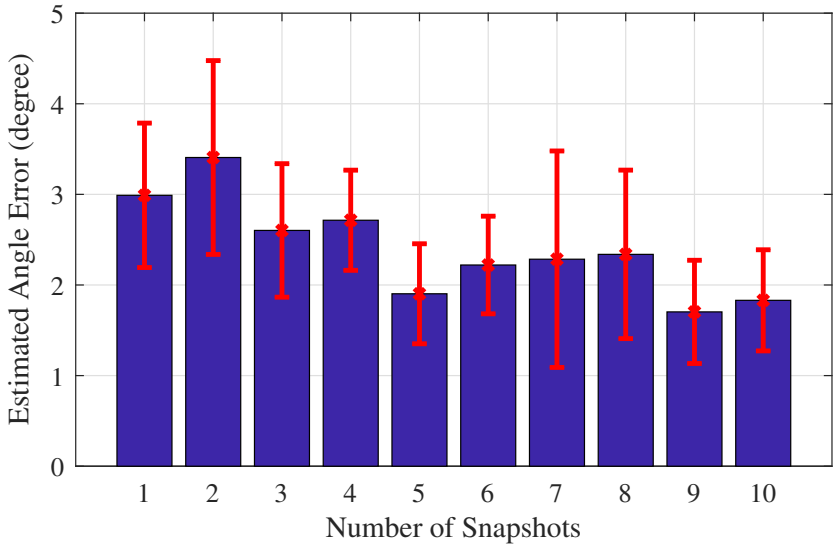


Figure 5.17: Impact of the number of snapshots.

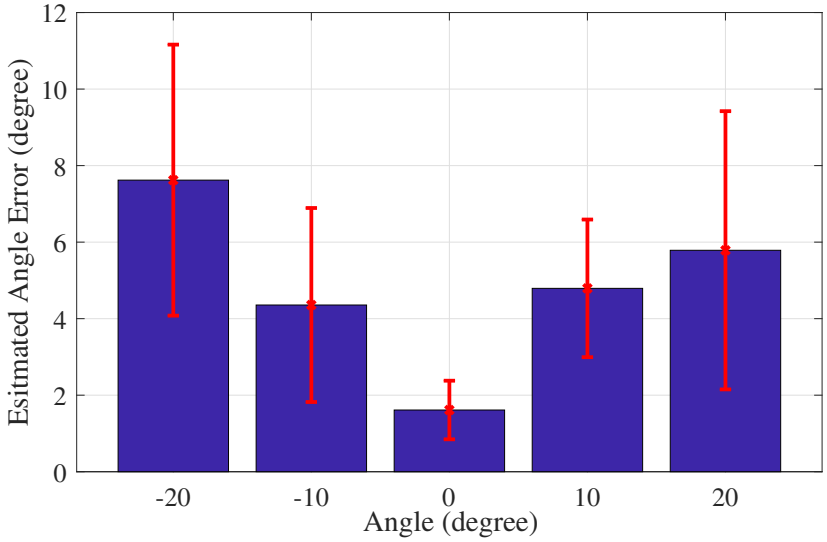


Figure 5.18: Impact of the height difference between the tag array and the antennas, which is represented by the angle as shown in Fig. 5.19.

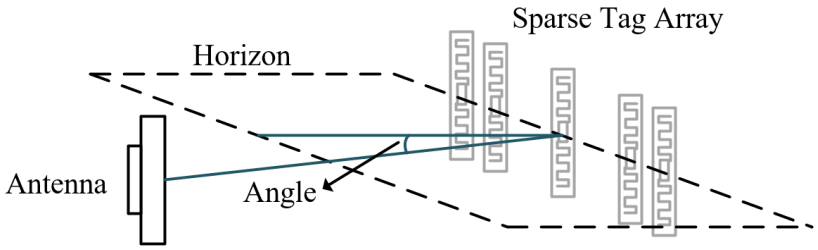


Figure 5.19: The height difference between the tag array and the antennas is represented by the angle between the horizontal plan and the line connecting the antenna and the center tag.

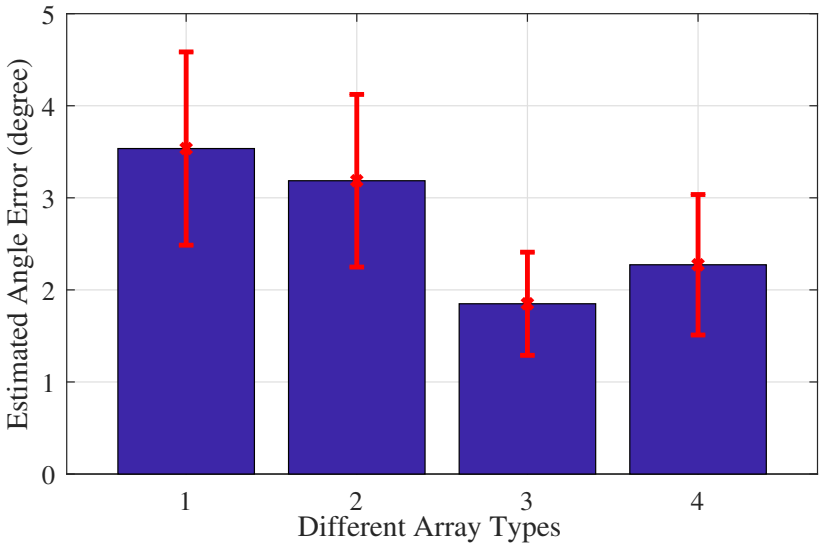


Figure 5.20: Impact of different array types on DOA estimation error. The first and second arrays are ULA with 3 and 5 tags, respectively, while the third and fourth arrays are sparse tag arrays, with 5 tags at positions (0, d, $3d$, $5d$, $6d$), and 7 tags at positions (0, d, $2d$, $4d$, $6d$, $7d$, $8d$), respectively.

Fig. 5.20 presents the estimated DOA errors obtained using different types of arrays. In our experiments, we evaluate the DOA estimation accuracy of 4 types of tag arrays. The first and the second arrays are ULA with 3 and 5 tags, respectively, while the third and the fourth arrays are sparse tag arrays, which consist of 5 tags at positions (0, d, $3d$, $5d$, $6d$), and 7 tags at positions (0, d, $2d$, $4d$, $6d$, $7d$, $8d$), respectively. From Fig. 5.20, we can see that the angle errors of ULA are both higher than $3°$, while both sparse tag arrays achieve lower errors about $2°$. This is because the sparse tag array achieves a higher angle resolution than ULA with the same number of tags. Fig. 5.20 also shows that the errors of the 5-tag sparse array is close but lower than that of the 7-tag sparse array. The 5-tag sparse array is sufficient to estimate DOA accurately.

5.4.5 Evaluation of the Near-field Effect

According to the FCC regulation on transmit power, the distance between the reader antenna and the tag array is usually not large. The tag array is placed within 3 m from the polarized antennas in all our experiments. Otherwise, the tag array can hardly be detected by the reader due to extremely weak RSS. The MUSIC algorithm is adopted for DOA estimation, which

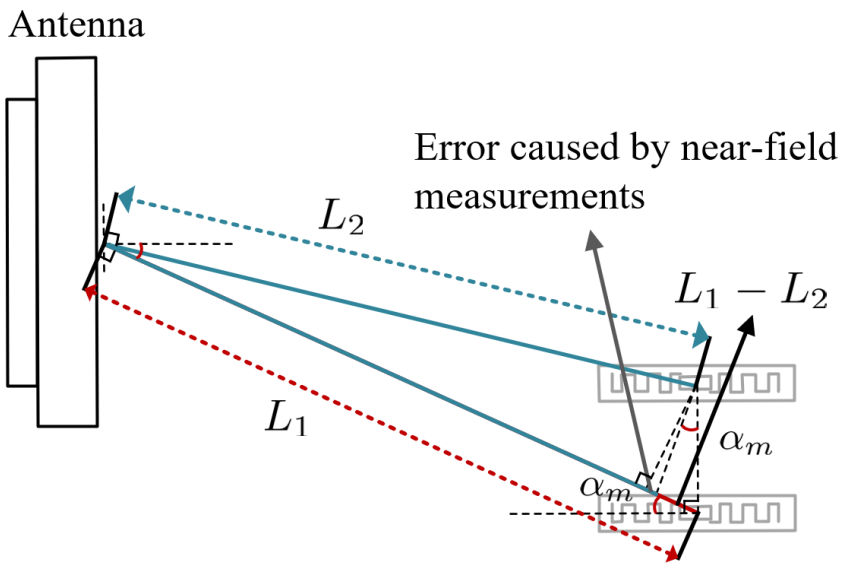


Figure 5.21: Illustrate the error introduced by the near-field measurements.

assumes the incident wave to the array is a plane wave. Such an assumption may not be rigorous in near-field communications scenarios, and will cause extra estimation errors. Fig. 5.21 shows an antenna and a simple 2-tag array, where the tags are placed d apart from each other. The tag-to-reader distances are L_1 and L_2 for Tag 1 and Tag 2, respectively. Here α_m is the DOA to be estimated. If we assume the incident wave to Tag 1 and the wave to Tag 2 are along two parallel lines (i.e., the plane wave assumption holds true), $(L_1 - L_2)$ can be consider as an edge of the right angled triangle as shown in the figure. Thus α_m can be easily computed as:

$$\alpha_m = \arcsin\left(\frac{d}{L_1 - L_2}\right), \tag{5.32}$$

where $(L_1 - L_2)$ can be estimated from the phase difference of the two tags. However, when the two tags are close to the antenna, the two incident waves will not be parallel and the relationship (5.32) will not hold true, which leads to additional DOA errors.

To evaluate the influence of such near-field effect, we conduct two experiments to find out the effective range of the SparseTag system. We first test the influence of the distance between the tag array and the antennas. We estimate DOAs under different tag-to-reader distances, ranging from 0.5 m to 2.5 m, and the results are presented in Fig. 5.22. We find that when the distance is 0.5 m or lower, the DOA error will be higher than $3.6°$. When the distance is 1 m

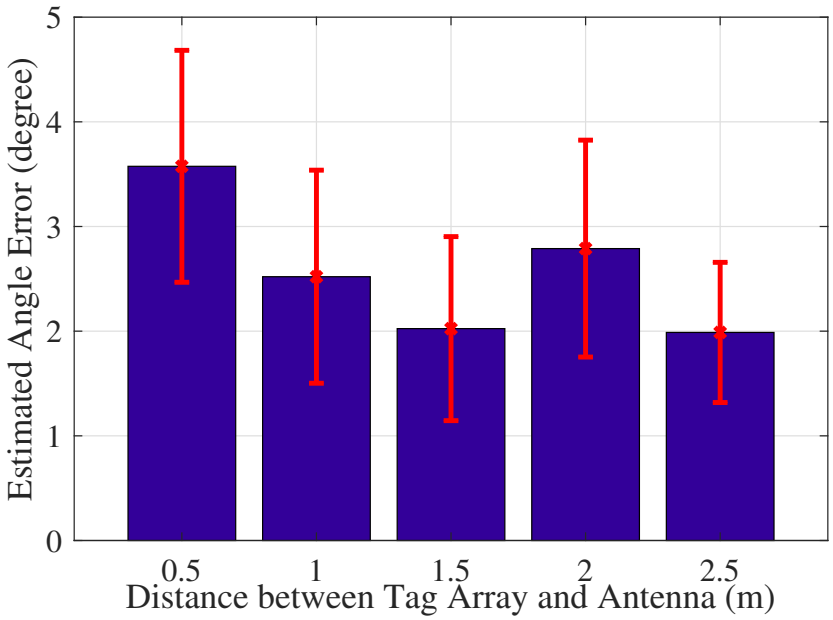


Figure 5.22: Impact of the distance between the tag array and the antennas.

or larger, the DOA error will be lower than $3°$. These results show that the influence of the tag-to-antenna distance is not strong since in typical applications the tag array will placed more than 1 m away from the antenna.

The second experiment examines the influence of different measuring angle on DOA error. We place the tag array at a fixed distance (e.g., 2 m) from the reader, and estimate the DOA from different relative antenna-tag array positions, where the ground truth DOA ranges from $-75°$ to $75°$. The estimation errors are presented in Fig. 5.23. We find that the DOA error is lower than $3.5°$ when the tag array is place between $-60°$ and $60°$. However, the error becomes considerably large when the DOA is over $75°$. This is because when the angle is too large, the polarized antenna can hardly collect phase values from all the channels and from each tag in the array. From these two experiments, we conclude that the effective range of our proposed system is from 0.5 m to 2.5 m and the effective range of estimated DOA should be between -$75°$ and $75°$.

5.5 Related Work

With the rapid development of Internet of Things, indoor localization attracts increasing attention in recent years. As an RFID-based indoor localization system, our work is closely related

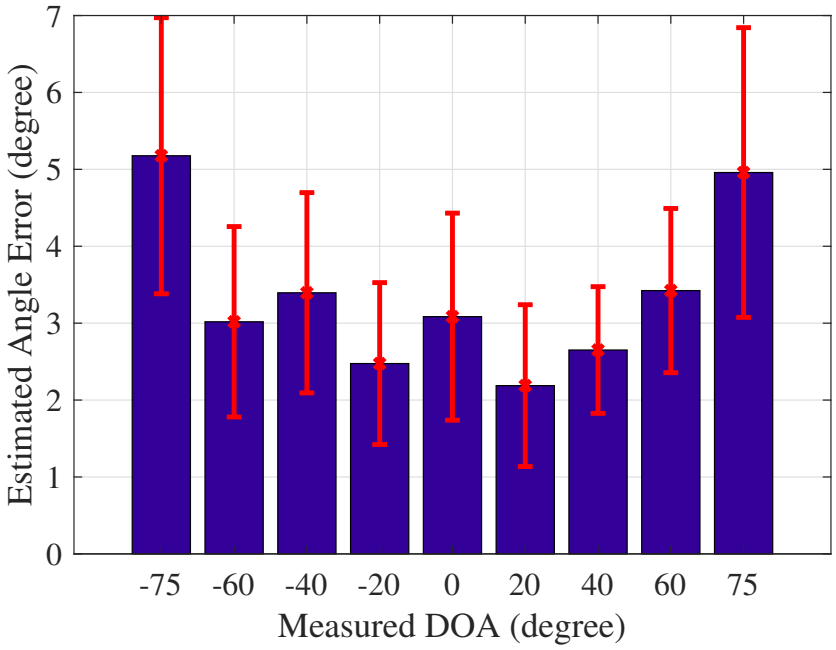


Figure 5.23: Impact of different ground truth DOA values.

Table 5.2: Features in Different RFID Tag Localization Techniques

Localization Technique	Hardware Modification	Tag Array	Antenna Array	Dynamic Tags or Antennas
LANDMARC	No	No	Yes	No
RF-IDraw	No	No	Yes	No
Tagoram	No	No	No	Yes
RFfind	Yes	No	No	No
SparseTag	No	Yes	No	No

to the RF based localization techniques in prior work. In this section we mainly focus on WiFi based techniques and RFID based techniques.

WiFi signals are widely utilized for indoor localization because of its low-cost, wide coverage, and ubiquitous deployment. Among various techniques, Angle of Arrival (AoA) is a typical method to estimate the location of the transmitter [115], but the accurate AoA is hard to estimate because of the multipath effect on the WiFi signal. To mitigate the multipath effect, antenna array-based systems are proposed to estimate the angle of multiple incoming paths of WiFi signal and distinguish the Line-of-sight (LoS) component [116, 117]. In addition, rather than directly calculate the AoA of the LOS path, some prior works leverage machine learning to estimate the position of the transmitter by learning the location features from collected channel state data. For example, Radar is a WiFi fingerprinting scheme using RSS [118]. Channel

State Information (CSI) is regarded as fine-grained representation of the WiFi channel and can achieve more accurate localization performance [7]. However, a well-trained neural network is usually sensitive to changes in the environment, the network parameters need to be updated once the testing environment is changed. Compared with these antenna array based systems, our sparse tag array can achieve high resolution of angle estimation as well as having a low cost.

The RFID technology has been regarded as an effective and low-cost solution for many emerging IoT applications [25, 42, 54, 58, 157]. Although RFID-based systems are limited by the short communication range, the multipath effect on RFID systems is usually much smaller than that on WiFi systems. Thus, various RFID based localization schemes have been proposed to achieve higher accuracy and convenient deployment than WiFi-based systems.

Existing works on RFID tag localization can be classified into received signal strength Indicator (RSSI)-based and phase-based methods. These works mainly focus on locating a single tag, i.e., one tag is located at a time. For RSSI-based methods, a large number of reference tags are deployed at known locations. By comparing the RSSI data with reference tags, the position of the target tag can be determined [31]. In fact, RSSI values are raw channel information and are not stable, due to the factors such as multipath propagation, tag's orientation, RFID reader's transmit power, etc. RSSI based methods usually do not achieve high accuracy in indoor localization. On the other hand, phase based methods have been developed for estimating distance and direction of arrival (DOA) [33]. However, the measured phase is periodic, which leads to phase ambiguity and makes it less useful. Moreover, considerable measured phase errors are introduced by the reader antennas and the tag itself.

To address these issues, the synthetic aperture radar (SAR) technique is proposed for DOA estimation by moving the reader antenna around [21]. The second solution is the hologram technique, which computes the probability of each known position as the tag source within an area of interest and then chooses the most likely position as the tag location [22, 34]. Another solution is the hyperbolic-based method for distance estimation, which locates a static tag [35]. However, this solution does not achieve high localization accuracy due to limited number of reader antennas. In addition, the RFind system achieves higher localization accuracy using a

large virtual bandwidth to estimate time-of-flight, but it requires a special hardware [91]. The features of different RFID tag localization techniques are further summarized in Table5.2.

5.6 Conclusions

In this paper, we investigated the problem of localizing an RFID tag array. The proposed system was termed SparseTag, i.e., a sparse RFID tag array system for high accuracy backscatter indoor localization. The SparseTag system comprised four key components: (i) sparse array processing, (ii) difference co-array design, (iii) DOA estimation using a spatial smoothing method, and (iv) a DOA-based localization method. We implemented the SparseTag system using off-the-shelf RFID tags and reader, and assessed its performance with extensive experiments in two settings. The experimental results validated the effectiveness and high location accuracy of the proposed SparseTag system.

Chapter 6

RFID-Pose: Vision-aided 3D Human Pose Estimation with RFID

6.1 Introduction

In recent years, human pose tracking becomes an important topic in computer vision, evolving from 2D [173] to 3D poses [174]. The accuracy of human pose tracking technique is continuously improved by more advanced hardware and machine learning (i.e., deep learning) techniques. Camera-based techniques have been shown effective for human pose tracking. However, such vision-based techniques also raise security and privacy concerns. It is usually annoying if one is being watched by a video camera all day. It is reported that millions of wireless security cameras deployed around the world are at risk of being hacked [175]. The video data used for pose tracking could be intercepted and illegally used by hackers. The privacy issue draws increasing concerns in the age of Internet of Things (IoT), where eHealth based on IoT is an important part. Many techniques have been proposed to improve the privacy and reliability of the IoT [124–126].

With rapid development of machine learning, deep learning has been highly promising for improving the safety and reliability of personal software and the IoT, which usually relies on sufficient and high-quality data [127–129]. If the human pose data is obtained without using a camera, people will no longer worry about their privacy being threatened. To address this issue, several radio frequency (RF) sensing based schemes have been proposed for human pose estimation, such as WiFi [131,178], Frequency-Modulated Continuous Wave (FMCW) radar [176], and mmWave radar [177]. Unlike camera-based techniques, such RF sensing based schemes

estimate the human joints from a confidence map constructed by RF signals, so the user's privacy will be preserved. For example, channel state information (CSI) is utilized in WiFi based systems [131], and the human pose can be estimated with a deep neural network such as a convolutional neural network (CNN). However, due to the multipath effect, WiFi signals are highly sensitive to interference (e.g., movements) in the surrounding environment. Although FMCW radar is more robust to the environment interference than WiFi based systems, the cost of the system is higher than commodity WiFi, which hinders its wide deployment.

To this end, radio frequency identification (RFID) provides a promising solution for human pose estimation. Compared with the above contact-free RF sensing systems, RFID tags can be used as wearable sensors because of their small size. The interference caused by the multipath effect is much smaller in the RFID system. Furthermore, the cost of RFID systems is lower than the advanced radar based systems such as the FMCW radar. However, because of the low data rate in RFID systems, generating a joint confidence map for all joints, as in other RF based systems, is highly challenging. Consequently, the existing RFID based pose tracking systems are focused on monitoring the movements of one particular limb using the phase data sampled from multiple tags [134, 135]. When multiple joints are moving simultaneously, the performance could be affected by the disturbance of other RFID tags (e.g., the mutual coupling effect) or the inter-tag collisions. Thus, tracking the entire body with RFID tags is still a challenging and open problem.

In this paper, we address the challenges in human pose estimation using RFID tags with a novel vision-aided, deep learning solution. We propose the RFID-Pose system for tracking the movements of multiple human limbs in realtime. In the proposed system, RFID tags are attached to the target human joints. The movement of the tags are captured by the phase variations in the responses from each tag. We propose a vision-aided solution to help the proposed deep learning model to learn the features of tag phase variations, rather than localizing these tags with traditional tag localization techniques [57]. The collected RFID phase data is firstly preprocessed to improve the quality of the raw sampled data, in particular, to mitigate the phase distortion and estimate the large amount of missing samples. Then, we leverage a deep kinematic neural network to learn the features of RFID phase data, where a Kinect 2.0 is used to

obtain the ground truth (i.e., labeled data for training). With the assistance of vision data, the deep learning model transforms the phase variation into the spatial rotation angle of each human joint. Since the spatial rotation angle estimation does not require generating a confidence map, the low data rate limitation of RFID systems is no longer an issue. In realtime estimation, human pose is reconstructed by estimated rotation angles from RFID data and the initial human skeleton. The vision data will not be needed anymore in this stage, and so the user's privacy can be well protected.

The main contributions of this paper are summarized as follows.

- To the best of our knowledge, this is the first work for 3D human pose estimation using commodity RFID reader and tags, which can effectively monitor multiple human joints simultaneously in realtime.

- We propose a novel data preprocessing approach to mitigate the severe RFID phase distortion and compensate the large amount of missing data in sampled raw RFID data. The tensor completion technique is utilized for data imputation, so that phase data for all RFID tags can be estimated. The greatly improved data quality leads to more effective learning for human pose estimation.

- We propose a vision-aided solution for training the proposed deep kinematic neural network, to transform sensed RFID phase variations to the spatial rotation of each limb. The proposed approach effectively addresses the challenges of the low data rate in RFID systems, because rotation angle estimation requires much less data than generating a joint confidence map.

- We develop a prototype system with commodity RFID devices and Kinect 2.0, to evaluate the system performance. Our experimental study validates that the proposed RFID-Pose system can effectively track the human pose with different types of motions in realtime.

In the following, we review related work in Section 6.2 and present the RFID-Pose system overview in Section 6.3. The challenges and solutions to RFID data preprocessing are presented in Section 6.4. The challenges and solutions to RFID based pose estimation are analyzed and

introduced in Section 6.5. We present our prototype system evaluation in Section 6.6 and conclude this paper in Section 6.7.

6.2 Related Work

This work is closely related to prior works on RFID based sensing [10] and human pose estimation [186]. We mainly focus on these two classes of systems in the following.

Recently, passive RFID tags have attracted great interest because of their easy deployment and low-cost features [179]. The Low Level Reader Protocol used by the Reader can provide useful low-level information such as received signal strength indicator (RSSI), phase, Doppler frequency shift, timestamp, etc. [38]. As a result, many RFID-based sensing techniques have been developed for many applications, such as indoor localization [21, 22, 34, 57, 91], vital sign monitoring [8, 27, 58, 88, 89, 138, 139], user authentication [68], material identification [69], object orientation estimation [23], vibration sensing [70], anomaly detection [71], temperature sensing [52], and drone localization and navigation [53, 54, 90]. Particularly, the RF-wear system [135] and RF-Kinect system [134] utilize RFID tags attached to the human joints to estimate the movement of a particular limb, such as front arms, front legs, and thighs [134, 135]. We adopt the same approach in RFID-Pose. However, these systems may not be suitable for realtime human pose estimation, especially when multiple moving joints need to be tracked simultaneously. These RFID based sensing systems inspire us to develop an RFID based pose estimation system.

Prior works on human pose estimation are mainly based on computer vision techniques [186, 187]. For human pose estimation using video data, deep learning based method has been shown effective for 2D human pose with conventional RGB cameras [173, 188], and 3D human pose with RGB-Depth cameras [189] and VICON systems [190]. These camera-based techniques can achieve high accuracy, but all require sufficient lighting condition and may raise privacy concerns.

These limitations motivate the development of RF based pose estimation techniques, because detecting RF signals do not require any lighting [191]. Moreover, since no video is used in the RF systems, the privacy issues are effectively addressed. However, collecting labeled

pose data from RF signals is very challenging. Therefore, several RF based techniques leverage vision data as labeled pose data to train the deep learning network. This approach is also taken in the proposed RFID-Pose system. For example, RFPose is the first work to use RF signals with an FMCW radar for 2D human pose estimation, where a teacher-student deep learning model is utilized [176]. RFPose3D is the later version for 3D human pose estimation with FMCW radar [191]. Moreover, mmwave Radar is also utilized for human pose estimation with deep learning [177]. Recently, WiFi CSI has been exploited to create 2D skeletions [178] and 3D human poses [131] using cross-modal deep learning techniques. However, Radar and WiFi based human pose estimation are easily influenced by the environment noise and interference, and the FMCW radar technique is limited by the relatively higher cost (e.g., implemented with Universal Software Radio Peripherals (USRP)).

The proposed RFID-Pose system, to the best of our knowledge, is the first to apply RFID based sensing for 3D human pose estimation. The proposed system consists of a novel and effective solutions for cross-modal 3D human pose estimation using RFID and computer vision, which is much more robust compared with WiFi and Radar based methods.

6.3 RFID-Pose System Overview

In this paper, we propose an RFID based sensing system, termed RFID-Pose, to estimate and track 3D human pose in realtime. The RFID-Pose system can sense the 3D positions of all the RFID tags attached to the human body by exploiting the phase data collected at the reader antennas. The training process of the system is supervised by the labeled vision data collected by a Kinect2.0 device, but only RFID data will be required for online human skeleton estimation. Human pose can be effectively constructed by mapping the positions of the attached RFID tags into 3D coordinates. The overview of the RFID-Pose system architecture is presented in Fig. 6.1, which is mainly composed of four components, including (i) RFID phase data collection, (ii) Kinect skeleton data collection, (iii) RFID data preprocessing, and (iv) Skeleton reconstruction using a deep kinematic neural network.

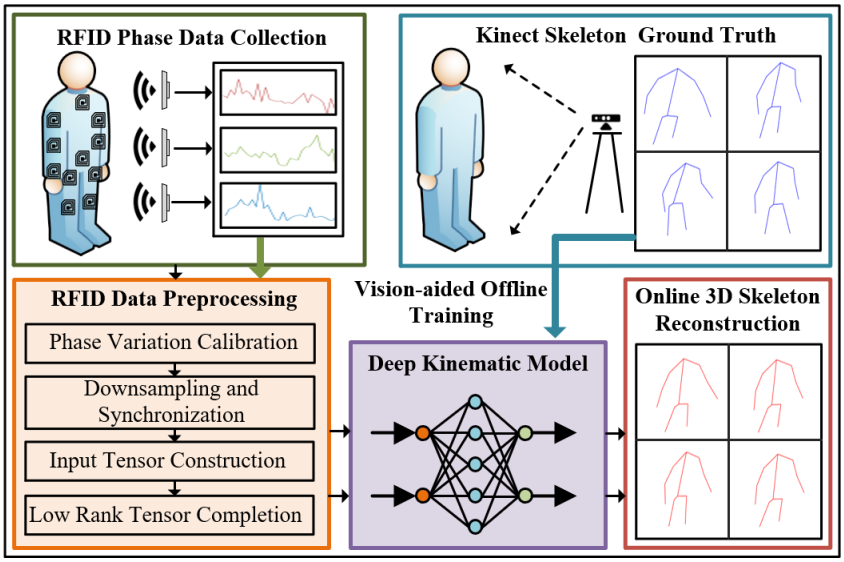


Figure 6.1: Overview of the RFID-Pose system architecture.

6.3.1 RFID Phase and Kinect Pose Data Collection

In the proposed system, training data is sampled by both the RFID antennas and the Kinect 2.0 device simultaneously. The collected RFID data will be used as the input to the deep kinematic neural network, and the Kinect 3D pose data will be used as labeled data for the supervised training. To collect RFID data, we attach passive RFID tags on the 12 joints of the human body. Three reader antennas are used to collect the phase and timestamp data from all the attached RFID tags. Kinect 2.0 is a depth camera widely used for capturing 3D poses in interactive video games. The 3D position of each human joint is estimated by both the RGB camera and the infrared sensors, and all measured joint positions are stored as 3D coordinates.

6.3.2 RFID Data Preprocessing

Since the sampled RFID raw phase data suffers from considerable distortion caused by channel hopping and phase wrapping, the RFID phase calibration must be applied to cleanse the data before using it to train the deep neural network. We first calibrate the phase variation to mitigate the influence of channel hopping and phase wrapping. Next, we downsample the calibrated RFID data and synchronize it with the 3D pose time sequence obtained by Kinect. However, because of the slotted ALOHA-like transmission in the RFID system, tags are not evenly interrogated by the antennas. In order to synchronize the RFID data with the collected

pose data from Kinect, we should obtain the phase for all tags corresponding to each Kinect data frame. To this end, we propose to employ low rank tensor completion to estimate the missing phase values from the tags. Finally, the calibrated phase data is used as input to train the deep neural network for human skeleton reconstruction.

6.3.3 Human Skeleton Reconstruction with a Deep Kinematic Neural Network

In RFID-Pose, we incorporate the deep kinematic neural network to learn the features of the RFID phase data. Unlike monitoring one particular limb movement as in traditional RFID based skeleton tracking systems [134, 135], the deep kinematic neural network is designed to simultaneously estimate the spatial rotation of all human joints relative to their parent joints. Once the initial human skeleton (i.e., the length of the limbs of target) is given, the network could effectively learn the features of calibrated RFID tensor data, and reconstruct the positions of human joints with estimated rotation angles. In RFID-Pose, the Kinect pose data is only used as benchmark for evaluating the accuracy of 3D pose reconstruction in the online testing process.

6.4 Challenges and Solutions: RFID Phase Distortion Mitigation and Data Imputation

The proposed RFID-Pose system reconstructs 3D human pose from RFID phase data with a deep kinematic neural network. However, the raw RFID phase data cannot be directly used for training and testing. The raw phase dataset from one of the tags sampled by a reader antenna in 500 time slots is plotted as diamond in Fig. 6.2. The figure shows that the collected RFID phase data is severely interfered during transmission by channel hopping and phase wrapping. Furthermore, there are many samples with a 0 value, which means the tag is not successfully sampled in the time slot. This is due to the Slotted ALOHA transmission in RFID systems; only one tag is allowed to respond to the reader's query in each time slot. Such sparse, low quality RFID data makes the RFID based 3D pose tracking highly challenging unless an appropriate data preprocessing is conducted.

Therefore, we propose the following RFID data preprocessing for the sampled RFID phase data, as illustrated in Fig. 6.3. In the preprocessing procedure, we first calibrate the overall

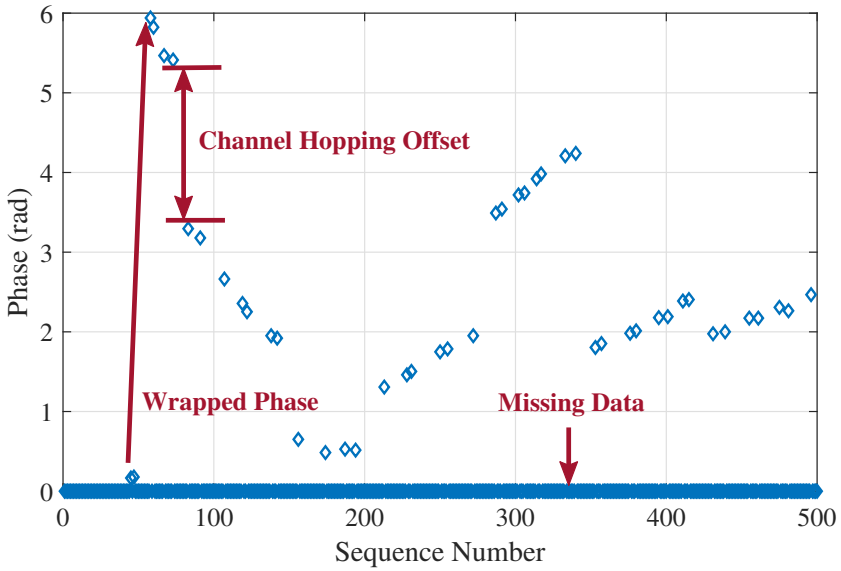


Figure 6.2: Raw phase sampled from one of the RFID tags by a single Reader antenna.

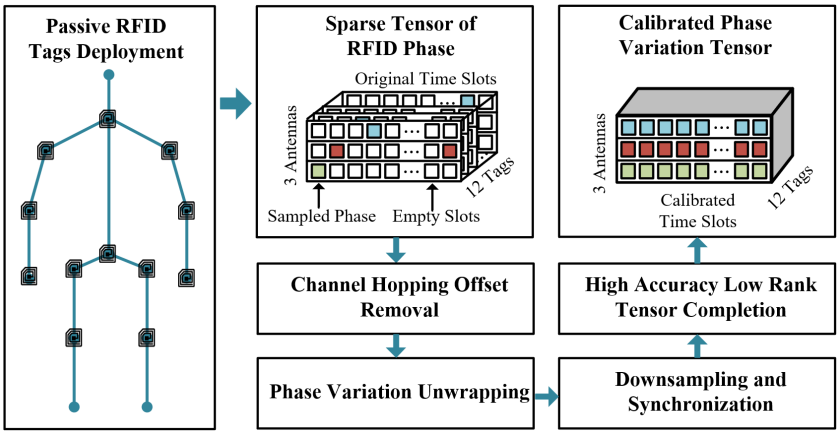


Figure 6.3: Flow chart of RFID data preprocessing.

phase interference in the raw data and then synchronize the RFID phase data with the collected Kinect data (used as labels for training) Next, the RFID data is used to construct a 3rd-order tensor, where the element at location (x, y, z) is the data collected from antenna x in time slot y from RFID tag z. We leverage High Accuracy Low Rank Tensor Completion (HaLRTC) to recover the missing samples and form the input data tensor, which is fed into the deep kinematic neural network for training and inference. More details are provided in the following.

6.4.1 Combating Collected Phase Interference

Frequency Hopping Offset Mitigation

In the proposed system, we leverage an RFID reader to extract the phase data from received RFID tag responses using the Low Level Reader Protocol, which is indicative of the tag-to-antenna distance [38]. The phase value is obtained when the RFID reader receives the Electronic Product Code (EPC) from the interrogated tag. The sampled phase value can be written as:

$$\Phi = \text{mod}\left(\frac{4\pi S f}{c} + \Phi_{tag} + \Phi_a,\ 2\pi\right), \tag{6.1}$$

where S denotes the distance between the interrogated tag and the reader antenna; and Φ_{tag} and Φ_a represent the phase offset caused by the circuits in the RFID tag and the reader antenna, respectively; f is the center frequency of the channel; and c is the speed of light. The equation shows that the phase value is indicative of the variation of the tag-to-antenna distance S, but it is also affected by the phase offset caused by the tag Φ_{tag} and the antenna Φ_a.

According to the FCC regulations, the Ultra-High Frequency (UHF) RFID system should hop among 50 channels during operation to avoid collisions among multiple RFID readers. In (6.1), the sum phase offset $\Phi_\alpha = \Phi_{tag,\alpha} + \Phi_{a,\alpha}$ is determined by both the hardware and the current frequency f_α used for the interrogation. So a considerable phase offset will be generated each time when the system hops to a new channel. As shown in Fig. 6.2, the severe phase offset is caused by channel hopping, which leads to considerable interference in the collected phase data. To mitigate the interference, we first rewrite the sampled phase in (6.1) from each channel α as:

$$\Phi = \text{mod}\left(\frac{4\pi S f_\alpha}{c} + \Phi_\alpha,\ 2\pi\right),\ \alpha = 1, 2, ..., 50, \tag{6.2}$$

where α is the RFID channel index ranging from 1 to 50. The equation shows that the channel hopping offset is a constant value for each particular channel, which can be canceled by subtracting two phase samples on the same channel. Thus, rather than using the RFID phase data,

we calculate the RFID phase variation on the same channel to mitigate the interference caused by the channel hopping offset.

The phase variation is calculated by subtracting a sampled phase data from the previous one on the same channel α, as:

$$\phi = \text{mod}\left(\frac{4\pi(S_n - S_{n-1})f_\alpha}{c},\ 2\pi\right), \qquad \alpha = 1, 2, ..., 50,\ n = 2, 3, ..., \tag{6.3}$$

where S_n represents the tag-to-antenna distance for the nth sampled data on the current channel. It can be seen that the phase variation in (6.3) is not affected by the phase offset anymore. Since $(S_n - S_{n-1})$ is the change of distance relative to the previous sample, phase variation is also suitable for tracking the movement of RFID tags. Therefore, to mitigate the interference caused by the frequency hopping offset, the input RFID data to the deep kinematic network is composed of the phase variation calculated for each RFID channel.

Phase Data Unwrapping

After calculating the phase variation for each channel, the phase distortion caused by channel hopping will be effectively mitigated. However, as shown in Fig. 6.2, since the sampled phase is wrapped in $[0, 2\pi]$ rad, the wrapped phase data also leads to severe interference in calculated phase variation. For example, if the phase changes from 0.1 rad to -0.1 rad, calculated phase variation will be $2\pi - 0.2$ rad, but the real phase variation is only -0.2 rad. To avoid the influence of phase wrapping, we apply a simple algorithm to unwrap the phase variation.

Considering that the frequency range of the reader antenna is 902MHz–928MHz with a wavelength about 33cm, we assume that all the tag position variations between two adjacent samples is smaller than 16.5cm (half of the weave length), which is reasonable given the 110Hz sampling rate. Thus, we calibrate the calculated phase variation when its absolute value is larger than π as follows:

$$\phi' = \phi - 2\pi\frac{\phi}{|\phi|},\ \text{if } |\phi| > \pi. \tag{6.4}$$

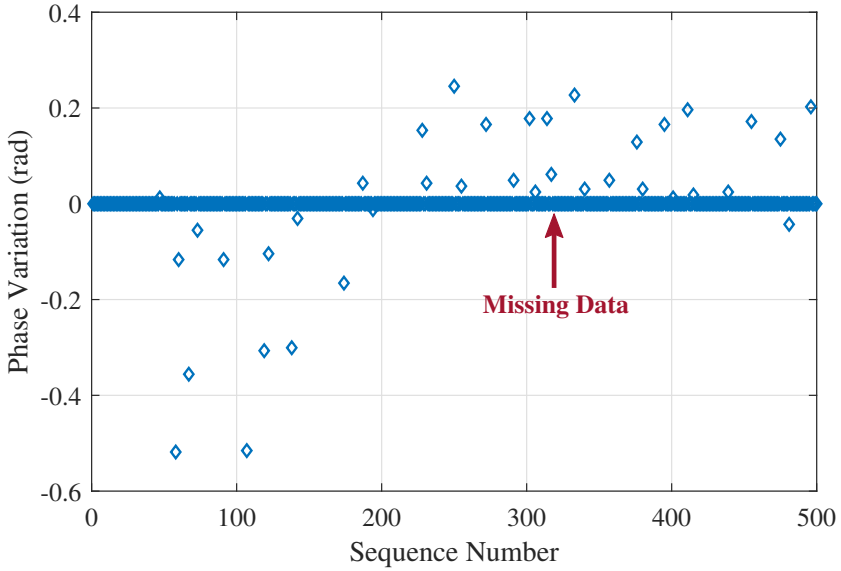


Figure 6.4: Calibrated phase variation data from one of the RFID tags (the raw data is plotted in Fig. 6.2).

In (6.4), $\phi/|\phi|$ returns the sign of ϕ. Then depending on whether the phase variation is positive or negative, a -2π or a 2π offset is added to ϕ. The calibrated phase variation, for the raw phase data shown in Fig. 6.2, is presented as diamonds in Fig. 6.4. We can see that, the channel hopping offset is eliminated in the calibrated data, as well as the phase distortion caused by phase wrapping. Notice that there are still missing data samples, which should be addressed. Otherwise, the input data still contains too many empty units (i.e., it is still highly sparse).

6.4.2 RFID Data Imputation

Following FCC regulations, the communications between the RFID reader and tags are based on Slotted ALOHA. It means the back propagation data of all the tags are received randomly, and only one tag can respond to the reader in each time slot (i.e., only one phase sample can be collected from one of the tags at a time). In RFID-Pose, we employ a commodity RFID reader with three antennas to scan the 12 tags attached to the human joints. The sampling rate for each tag is thus very low. From the calibrated phase variation data in Fig. 6.4, we can see that this antenna only collects 38 samples for that tag in 500 time slots, while ideally we expect 500 samples. This means more than 90% of the data are missing for this tag. Learning features

from such sparse datasets is highly challenging, and we should estimate the missing samples for more effective learning.

Downsampling and Synchronization

With N_p antennas and N_q tags, we can create a $N_p \times N_q$ phase variation matrix for all the tags and antennas and extend it into an order-3 tensor structure for various time slots. The data tensor for N_p antennas, N_q tags, and N_t time slots is constructed as:

$$\boldsymbol{\psi}(:,:,q) = \begin{bmatrix} \phi_{q1}^1 & \phi_{q2}^1 & \cdots & \phi_{qN_t}^1 \\ \phi_{q1}^2 & \phi_{q2}^2 & \cdots & \phi_{qN_t}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{q1}^{N_p} & \phi_{q2}^{N_p} & \cdots & \phi_{qN_t}^{N_p} \end{bmatrix}, q = 1, 2, ..., N_q.$$

In the data tensor, ϕ_{qt}^p represents the calibrated phase variation data from tag q sampled by antenna p in time slot t. Note that only one phase variation can be sampled in each $\boldsymbol{\psi}(:,t,:)$. So only up to N_t samples are non-empty in this $N_p \times N_t \times N_q$ tensor, i.e., it is highly sparse. The RFID-Pose system utilizes 12 tags and 3 antennas. Thus the sparsity of the data tensor is as high as 97.22%, which leads to poor learning performance. However, such highly sparse tensors are very hard to be accurately completed with traditional compressed sensing techniques.

Fortunately, since the frame rate of the Kinect data is 30 fps, we can compress the RFID data in multiple adjacent time slots to match the corresponding, single Kinect data frame. Furthermore, since the requirement on the frame rate is not very high for human pose tracking (which mostly involve slow body movements), we can further downsample the Kinect data so that more slices in the sparse tensor can be grouped into one. If we compress tensor $\boldsymbol{\psi}$ into $\boldsymbol{\Psi}$ with ratio ξ, the new tensor after synchronization could be denoted as:

$$\boldsymbol{\Psi}(:,:,q) = \begin{bmatrix} \bar{\phi}_{q1}^1 & \bar{\phi}_{q2}^1 & \cdots & \bar{\phi}_{qN_T}^1 \\ \bar{\phi}_{q1}^2 & \bar{\phi}_{q2}^2 & \cdots & \bar{\phi}_{qN_T}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \bar{\phi}_{q1}^{N_p} & \bar{\phi}_{q2}^{N_p} & \cdots & \bar{\phi}_{qN_T}^{N_p} \end{bmatrix}, q = 1, 2, ..., N_q,$$

where N_T is the number of synchronized time slots for RFID data, which is the same as the number of downsampled Kinect data units. As the equation shows, for each unit $\mathbf{\Psi}(n_p, n_t, q)$ in the tensor, the first coordinate n_p represents the index of the sampling antenna, the second coordinate n_t indicates the index of the time slot, and the third coordinate q is the index of the attached RFID tag. The tensor structure is also illustrated in the right-hand-side of Fig. 7.2. In addition, $\bar{\phi}^p_{q,T}$ is the mean phase variation from tag q sampled by antenna p in synchronized time slot T, which is calculated for the ξ adjacent values in $\boldsymbol{\psi}$ as:

$$\bar{\phi}^p_{qT} = \frac{1}{\xi} \sum_{t=T}^{T+\xi-1} \phi^p_{qt}. \tag{6.5}$$

After the downsampling process, the sampling period is also multiplied by ξ. Since phase variation represents the velocity of the overall phase changes, the mean value calculation still keeps the phase variation velocity unchanged. With downsampling and synchronization, the sparsity of the RFID data will be greatly reduced, as illustrated in Fig. 6.5, which is obtained with $\xi = 50$ from the calibrated phase variation data shown in Fig. 6.4. As the figure shows, there are now 38 valid data units in 70 time slots. Compared to the original data in Fig. 6.4, the sparsity is effectively reduced. However, there are still intervals of time with no effective sampled data, which will be addressed next.

High Accuracy Low Rank Tensor Completion (HaLRTC)

The commodity RFID reader used in RFID-Pose has three antennas. To accurately learn the RFID phase variation features, all tags should be sampled by all antennas in each time slot in the ideal case. However, the phase variations collected from different antennas could be treated as different samples from the same signal source (i.e., tag movement). Since the number of signal sources equals to the number attached RFID tags, the sparse tensor $\mathbf{\Psi}$ can be considered as a low-rank tensor, which can be recovered by low-rank tensor completion. This task is accomplished by solving the following optimization problem [168]:

$$\min_{\hat{\mathbf{\Psi}}} \|\hat{\mathbf{\Psi}} \text{ s.t.:} \mathbf{\Omega} * \hat{\mathbf{\Psi}} = \mathbf{\Omega} * \mathbf{\Psi}, \tag{6.6}$$

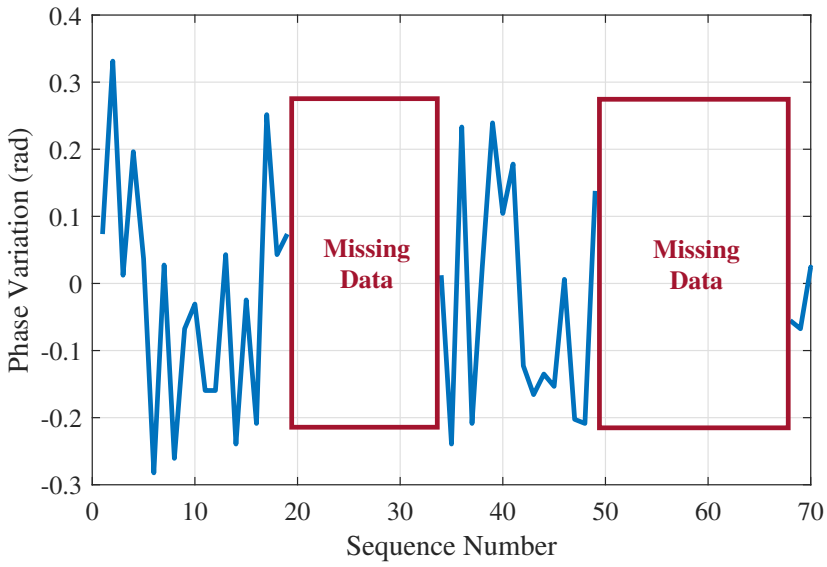


Figure 6.5: Downsampled and synchronized RFID phase variation from one RFID tag with $\xi = 50$.

where $\hat{\mathbf{\Psi}}$ is an estimation of the ideal tensor data $\mathbf{\Psi}_{ideal}$, which is composed of all the ideal phase variation data; and $\mathbf{\Omega}$ is a tensor of 0 and 1 elements, where $\mathbf{\Omega}_{IJK} = 1$ when $\mathbf{\Psi}_{IJK}$ is sampled, and $\mathbf{\Omega}_{IJK} = 0$ otherwise. In (6.6), $\|\cdot\|_*$ denotes the trace norm of tensors.

During the optimization procedure, the trace norm of the 3rd-order tensor $\mathbf{\Psi}$ is calculated with the combination of its unfolded matrix in different modes. The optimization problem is represented as [168]:

$$\begin{aligned} \min_{\hat{\mathbf{\Psi}}, \boldsymbol{M}_i} \quad & \sum_{i=1}^{3} h_i ||\boldsymbol{M}_i(i)||_* \\ \text{s.t.:} \quad & \mathbf{\Omega} * \hat{\mathbf{\Psi}} = \mathbf{\Omega} * \mathbf{\Psi}, \\ & \hat{\mathbf{\Psi}} = \boldsymbol{M}_i, \ i = 1, 2, 3, \end{aligned} \tag{6.7}$$

where h_i's are constants satisfying $\sum_{i=1}^{3} h_i = 1$, $\boldsymbol{M}_i$ is a tensor with the same size as $\hat{\mathbf{\Psi}}$, and $\boldsymbol{M}_i(i)$ is the matrix unfolded from tensor $\boldsymbol{M}_i$ in mode i. The equation shows that the trace norm of a tensor is a convex combination of norms for all matrices unfolded along each mode. In HaLRTC, the optimization problem (6.7) is solved with the Augmented Lagrange Multiplier

Method (ADMM) [146] with the augmented Lagrangian function defined as:

$$\boldsymbol{L}_\rho(\hat{\boldsymbol{\Psi}}, \boldsymbol{M}_i, \boldsymbol{Y}_i) = \sum_{i=1}^{3} h_i ||\boldsymbol{M}_i(i)||_* + \left\langle \hat{\boldsymbol{\Psi}} - \boldsymbol{M}_i, \boldsymbol{Y}_i \right\rangle + \frac{\rho}{2} ||\boldsymbol{M}_i - \hat{\boldsymbol{\Psi}}||_F^2,$$

where $\langle \cdot, \cdot \rangle$ represents the inner product of two tensors and $|| \cdot ||_F$ is the Frobenius norm of the tensor; $\boldsymbol{Y}_i$ is a zero tensor with the same size as $\hat{\boldsymbol{\Psi}}$, and $\rho > 0$ is the penalty factor in the algorithm. In our system we set $\rho = 1e^{-4}$. Rather than iterate recursively to optimize the target tensor $\hat{\boldsymbol{\Psi}}$. ADMM literately updates multiple variables, i.e., $\boldsymbol{M}_i$, $\hat{\boldsymbol{\Psi}}$, and $\boldsymbol{Y}_i$ as follows.

$$\begin{aligned}
&(i) \quad \boldsymbol{M}_i' = (\boldsymbol{M}_i) : \boldsymbol{L}_\rho(\hat{\boldsymbol{\Psi}}, \boldsymbol{M}_i, \boldsymbol{Y}_i) \\
&(ii) \quad \hat{\boldsymbol{\Psi}}' = (\hat{\boldsymbol{\Psi}}) : \boldsymbol{L}_\rho(\hat{\boldsymbol{\Psi}}, \boldsymbol{M}_i', \boldsymbol{Y}_i) \\
&(iii) \quad \boldsymbol{Y}_i' = \boldsymbol{Y}_i - \rho(\boldsymbol{M}_i' - \hat{\boldsymbol{\Psi}}').
\end{aligned}$$

These functions converge when the update between two adjacent iteration is sufficiently small. Thus, the update threshold is set to determine whether $\hat{\boldsymbol{\Psi}}$ is successfully estimated or not. To balance the data imputation performance and the convergence rate of the algorithm, we set the convergence threshold to $1e^{-6}$ to make sure the data is effectively recovered with an acceptable convergence rate. Compared with other low-rank tensor completion algorithms, HaLRTC can solve the optimization problem (4.7) more accurately with a lower complexity. The entire tensor completion process in our system only takes less than 0.1 second to execute because the downsampling reduces the input tensor size. As illustrated in Fig. 6.6, all the missing data can be effectively estimated by HaLRTC. So the reconstructed tensor $\hat{\boldsymbol{\Psi}}$ can be used by the deep learning model for 3D human pose estimation.

To evaluate the performance of the HaLRTC algorithm, we compare compare it with a conventional interpolation method, i.e., the bilinear interpolation technique. Fig. 6.7 shows one slice of phase variation data in tensor $\hat{\boldsymbol{\Psi}}$, which represents the synchronized phase variation data for all tags sampled by one antenna. As the figure shows, there are still many samples of value 0, indicating that most data are still missing after downsampling, especially for tags 6, 10, 11, and 12. Both HaLRTC and bilinear interpolation techniques are used to interpolate the

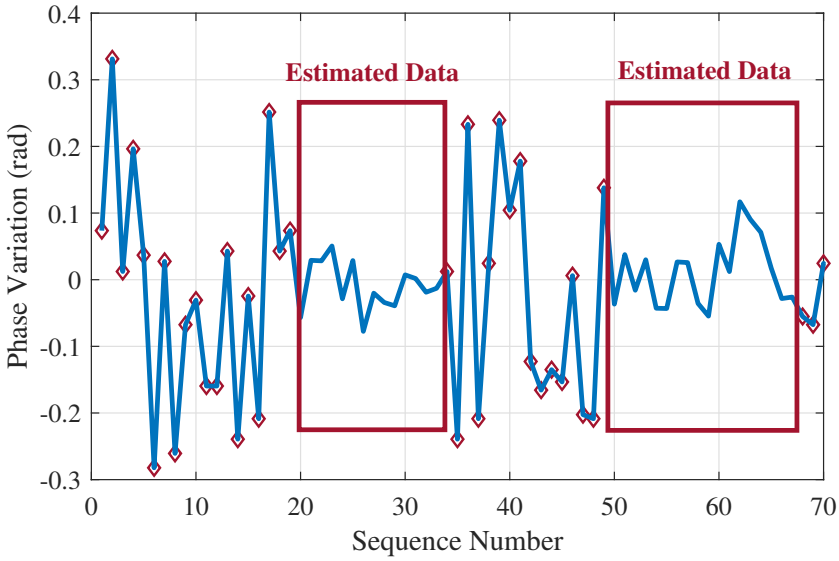


Figure 6.6: The missing data are estimated by HaLRTC.

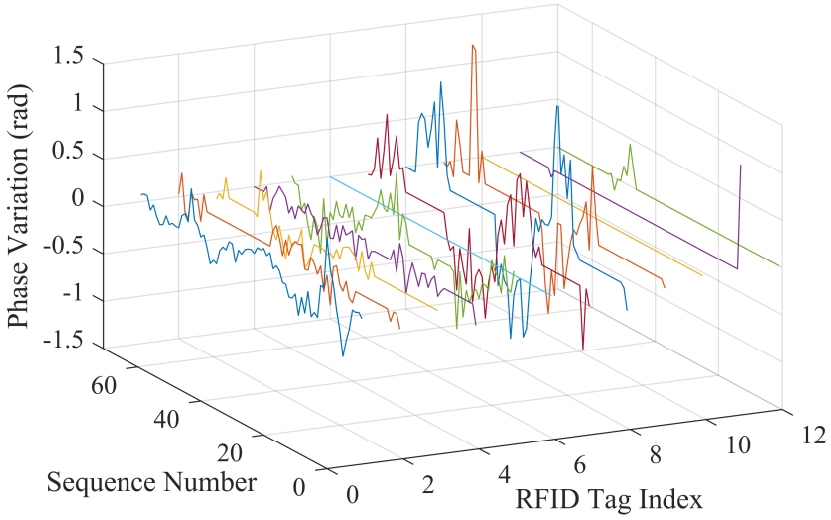


Figure 6.7: Sparse RFID phase variation matrix collected from one antenna.

miss samples, and the results are presented in Figs. 6.8 and 6.9, respectively. From Figs. 6.8 and 6.9, it can be seen that the phase variation data estimated by tensor completion shows high consistency among all tags, while sharp variations are generated by bilinear interpolation. Especially for the tags with high sparsity, e.g., tags 11 and 12, significant distortions have been introduced by bilinear interpolation, which will cause considerable skeleton estimation errors.

The superior performance of tensor completion in data imputation is mainly because the data is not evenly sampled in the RFID system. The sampled data from different tags usually

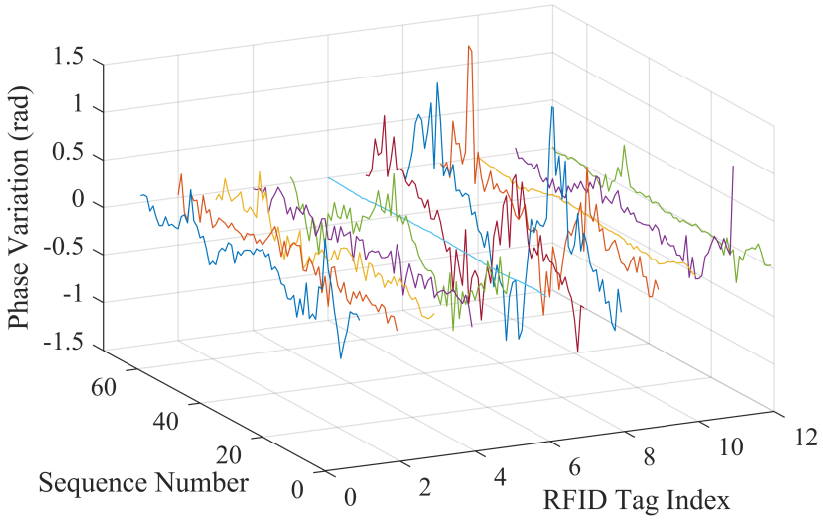


Figure 6.8: Phase variation matrix completed by HaLRTC.

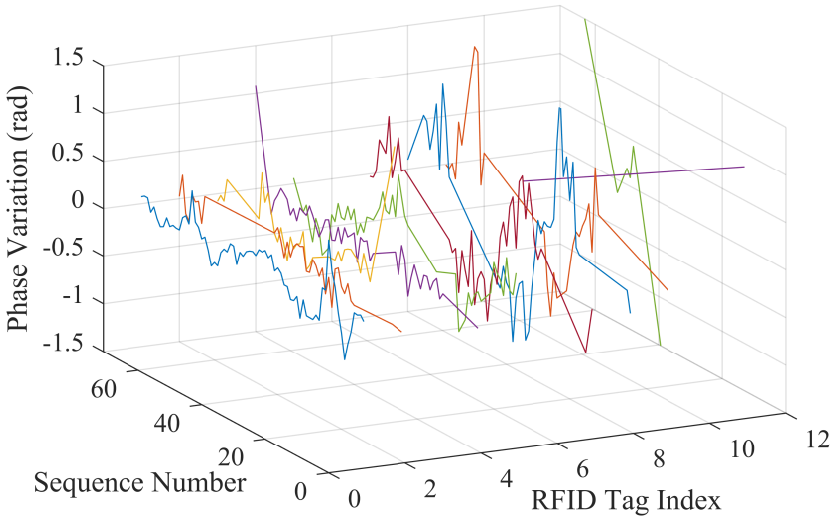


Figure 6.9: Phase variation matrix completed by the bilinear interpolation method.

have highly different sparsity (e.g., tag 1 versus tag 11 or 12 in Fig. 6.7). The traditional interpolation method is not suitable for this significant uneven sparsity situation. However, by solving the optimization problem (4.7), the missing samples can be interpolated based on the low rank components of the tensor data, which indicates the movement of the subject. In addition, the tensor completion process in our system only takes less than 0.1 second to execute because the downsampling has reduced the input tensor size. Thus, HaLRTC is a well-suited method for phase variation data imputation in RFID-Pose.

6.5 Challenges and Solutions: Human Pose Reconstruction with RFID Data

6.5.1 Challenges in RFID-based Human Pose Tracking

Tracking multiple joints of a human subject simultaneously with RFID tags is highly challenging, because the data rate of RFID systems is extremely low comparing to other wireless systems. According to the RFID Gen2 protocol, the medium access control (MAC) in RFID system follows the Slotted ALOHA protocol, which means only one tag can respond to the reader in each time slot. Such a transmission scheme makes the data rate of RFID much lower than other sensing systems such as video camera [173], WiFi [178], and FMCW radar [176]. In these RF-based skeleton tracking system, the human skeleton is extracted from the confidence map of the target joints, which is usually generated by a neural network. The RFID system's sampling rate is about 110Hz for each antenna. In order to generate a 100×100 confidence map to localize the joint positions at a rate of 5 fps(frames/second), only 22 phase data samples can be obtained for each frame. Recovering a map with 10000 data samples with only 22 phase data samples is a severely *ill-posed problem*, which is extremely challenging to solve even with advanced deep learning techniques.

The above ill-posed problem implies that the confidence map method may not be suitable to estimate the 3D pose of the human body. Consequently, the existing RFID based techniques mostly focus on estimating the movement of a particular limb movement, such as the front arm, the front leg, and thighs [134, 135]. Although, theoretically, the entire body movement could be reconstructed by combing all the limb movements, these systems may not be effective for realtime human pose estimation, especially when multiple moving joints need to be tracked simultaneously.

In RF-wear [135], two RFID tag arrays are attached to the two adjacent limbs of the subject, which are then used to estimate the rotation angle of human limbs with good accuracy. However, when tracking multiple limbs simultaneously, every limb should be attached with an RFID array. In this scenario, there will be a large number of tags to be interrogated by the RFID reader. The severe mutual coupling effect and considerable inter-tag collisions will cause a lot of missing samples and some tags may even be hardly sampled by the reader.

Similarly, in the RF-Kinect system [134], the rotation angle of one particular limb is estimated by the RF hologram technique [22]. Unfortunately, since the angle estimation is based on the probability distribution map built on the phase value of all attached tags, the accuracy of angle estimation could be affected when multiple tags are moving together. Moreover, the generation of the probability distribution map for each joint requires phase measurements for all possible rotation angles, which entail heavy calibration work.

Studying existing RFID based pose tracking systems, we found that, although generating the skeleton confidence map is challenging, the rotation angles of all human limbs could be relatively easily estimated from the scarce RFID data. This is because, when the limb's length is known, the system only needs to generate three angle values to reconstruct the particular limb's movement. That is, only $3n$ angle values need to be estimated when tracking n joints, which is considerably less than the number of samples required for confidence map generation, and is highly suited for RFID based sensing systems with constrained sampling rates. Accordingly, our goal is to estimate the rotation angle of each limb and leverage the forward kinematic technique to reconstruct the human skeleton with the estimated rotation angles.

6.5.2 Forward Kinematics

The technique to generate human 3D pose from limb rotation angles is *Forward Kinematics*, which is widely used in robotics and 3D animation [193]. An example of forward kinematic is shown in Fig. 6.10. The left-hand-side figure shows a human skeleton with a "T" pose, and the 12 joints with marked numbers are the target joints to track in our RFID-pose system. In forward kinematics, the 3D position of a joint is generated by (i) the rotation angle of the limb connecting the two joints; and (ii) the length of the limb, (iii) the position of its parent joint, which is defined as the rotation anchor. For example, in Fig. 6.10, the subject puts down his/her arms. Then joints 8, 9, 11, and 12 all move downward. Since joint 7 (i.e., the left shoulder) is the rotation anchor of the left upper arm, it is considered as the parent joint of joint 8 (i.e., the left elbow). The position of joint 8 can be calculated with the length of the upper arm and the 3D rotation angle. Similarly, the locations of joints 9, 11, and 12 can be estimated from their corresponding parent joints 8, 10, and 11, and the 3D rotation angles, respectively.

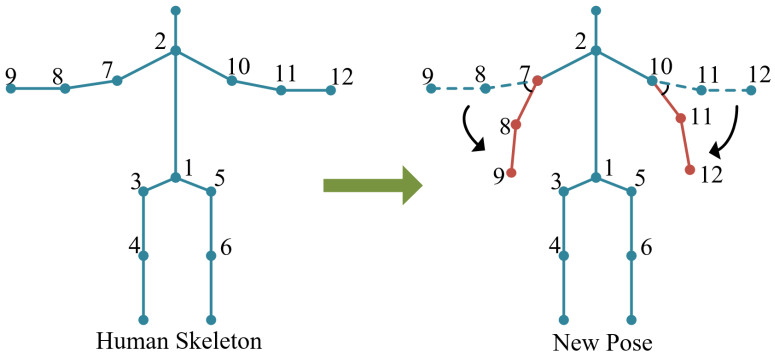


Figure 6.10: Example of limb rotation in the human skeleton.

Accordingly, once the initial skeleton is given (i.e., the original locations of all joints and the lengths of all limbs), each joint can be localized recursively based on the position of its parent joint and rotation angles.

The recursive rotation for the nth joint in time slot T can be expressed as:

$$\vec{P}_n^T = \vec{P}_{parent(n)}^T + \mathbf{R}_n^T \vec{P}_{relative(n)}^0, \tag{6.8}$$

where $\vec{P}_n^T$ represents the position of joint n of time slot T, $\vec{P}_{parent(n)}^T$ denotes the position of joint n's parent joint, $\mathbf{R}_n^T \in SO(3)$ represents the corresponding rotation matrix ($SO(3)$ denotes the 3D rotation group), and $\vec{P}_{relative(n)}^0$ is the 3D offset of joint n relative to its parent joint, given by:

$$\vec{P}_{relative(n)}^0 = \vec{P}_n^0 - \vec{P}_{parent(n)}^0, \tag{6.9}$$

where $\vec{P}_n^0$ and $\vec{P}_{parent(n)}^0$ represent the positions of joint n and its parent joint in the initial 3D skeleton, respectively. From (6.8), we can see that, with the initial skeleton data, all joint positions can be calculated by the corresponding rotation matrix $\mathbf{R}_n^T$.

According to Euler's rotation theorem, a 3D rotation can be represented as a *unit quaternion* in the system with format:

$$\ell + x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}. \tag{6.10}$$

In the unit quaternion ℓ, x, y, and z are real numbers, and $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are quaternion units. Given a 3D position vector represented as $a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}$ and a 3D rotation with unit quaternion $r_\ell + r_x\boldsymbol{i} + r_y\boldsymbol{j} + r_z\boldsymbol{k}$. The rotation matrix $\mathbf{R}$ is derived as:

$$\mathbf{R} = \begin{bmatrix} 1 - 2({r_y}^2 + {r_z}^2) & 2(r_x r_y + r_z r_\ell) & 2(r_x r_z - r_y r_\ell) \\ 2(r_x r_y - r_z r_\ell) & 1 - 2({r_x}^2 + {r_z}^2) & 2(r_y r_z + r_x r_\ell) \\ 2(r_x r_z + r_y r_\ell) & 2(r_y r_z - r_x r_\ell) & 1 - 2({r_x}^2 + {r_y}^2) \end{bmatrix}. \tag{6.11}$$

The new position vector, after the 3D rotation, can be calculated as

$$\begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = \mathbf{R} \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \tag{6.12}$$

The rotation matrix $\mathbf{R}$ is used in the Forward Kinematic (FK) layer of the learning model in the RFID-Pose system, which is to reconstruct the human 3D pose with the initial skeleton and the corresponding spatial rotations.

6.5.3 Deep Kinematic Neural Network

To reconstruct 3D human pose, we leverage a deep kinematic neural network to learn the features of RFID phase variation collected when the subject is moving. The structure of the learning model is illustrated in Fig. 6.11. The offline training goal is to learn the relationship between the RFID phase variation and the rotation of the human limbs. The 3D pose ground truth obtained from Kinect is in the form of 3D coordinates for the human joints. The initial target skeleton is required for each training dataset to transform the estimated rotation angle to the 3D positions through forward kinematic.

As Fig. 6.11 shows, the deep kinematic neural network is mainly composed of two parts, i.e., the Recurrent Autoencoder and the forward kinematic layer. The Recurrent Neural Network (RNN) is suitable for learning the features of phase variation sampled in a time sequence,

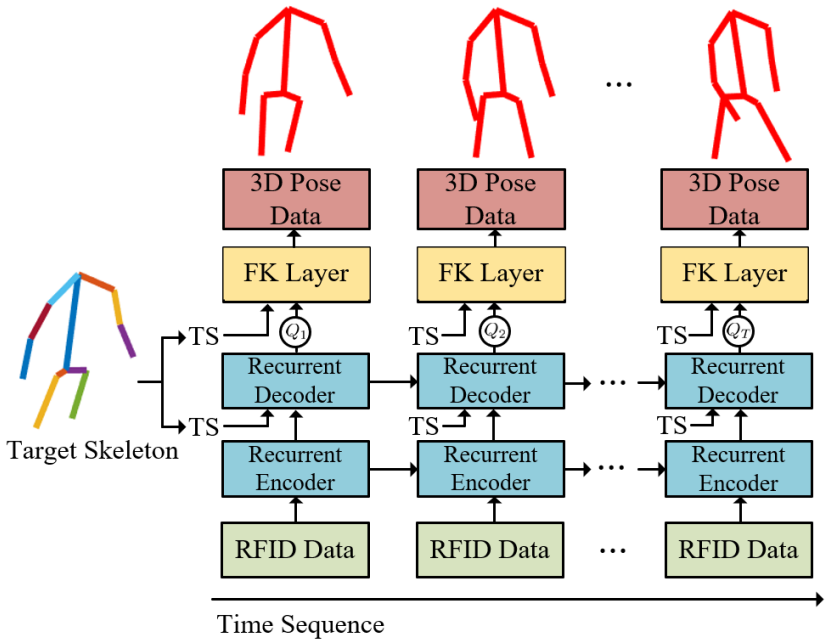


Figure 6.11: The deep kinematic neural network incorporated in RFID-Pose.

while the Autoencoder is a simple but effective learning model to extract the features of RFID phase data [58, 138]. The input training data is the RFID phase variation sequence and the 3D pose data sequence, which are synchronized after data preprocessing (see the previous section).

The Recurrent Autoencoder consists of two key parts, an encoder and a decoder. In each time slot, the features in the input RFID phase data is firstly extracted by the Recurrent Encoder and stored in the hidden layers, which consist of 256 gated recurrent units (GRU). Because of the recurrent structure, the hidden layer outputs in the previous time slot are also fed to the following Encoder. Thus, the Recurrent encoder can extract feature of the RFID phase data from both the current time slot and previous time slots. Then the Recurrent Decoder is leveraged to transfer the extracted feature stored in the Encoder hidden layer to 3D rotation data. Since the limb length data is required for the 3D rotation estimation from extracted RFID feature, the initial human skeleton should be added as another input to the Decoder. Moreover, the recurrent structure also feeds the previous hidden layer outputs to the current Decoder for learning the features in the output data sequence. The unit quaternion Q_T for each joint is obtained by normalizing the Recurrent Decoder output.

Next, with the initial skeleton and Q_T, the forward kinematic layer leverages the rotation matrix $\mathbf{R}$ to generate 3D coordinates for the subject, which are in the same format as the Kinect

ground truth data. With the error calculated between the estimated pose and the ground truth, the weights in the Recurrent Autoencoder will be trained by using error backpropagation.

6.6 Implementation and Evaluation

6.6.1 System Implementation

To evaluate the performance of the RFID-pose system, we develop a prototype system with an off-the-shelf Impinj R420 reader equipped with three S9028PCR polarized antennas. The RFID tags used for tracking human joint movements are ALN-9634 (HIGG-3). The vision data used for training supervision and test accuracy evaluation is collected with an Xbox Kinect 2.0 device. The sampling rate of the RFID phase data is around 110 Hz, and the frame rate of the Kinect 2.0 is 30 fps. All data is downsampled to 7.5 Hz after preprocessing and synchronization. The length of the RFID input tensor N_T is set to 30 during the experiments, which represents 4 seconds motion data.

The setup of the system is illustrated in Fig. 6.12. As the figure shows, we attach RFID tags to the 12 joints of the human body, which are the pelvis, neck, left hip, left knee, right hip, right knee, left shoulder, left elbow, left wrist, right shoulder, right elbow, and right wrist. To each joint, *one* passive RFID tag is attached to monitor the joint movement. The head and feet are omitted in our prototype system because of the limited scanning range of the RFID antenna used. The antennas are placed at different altitude positions to ensure that the antennas can interrogate all the tags. If we want to scan all the joints from head to feet, more antennas should be used in the system. However, the pose with the 12 joints is sufficient to monitor human behavior in most cases.

An MSI laptop with a Nvidia GTX 1080 GPU and an Intel Core i7-6820HK CPU is used as the processor for data training and signal processing. The frequency used by the prototype system hops among 50 channels from 902 MHz to 928 MHz, and it remains on a channel for 0.2 second.

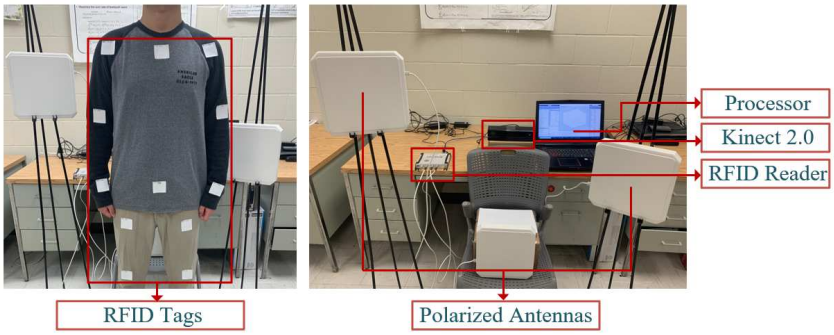


Figure 6.12: Illustration of the system setup for 3D pose estimation.

6.6.2 Performance Evaluation and Results

Overall Accuracy for Different Motions

We train the proposed deep kinematic neural network with different types of motions. The first type of motions is simple motion, which is only involved with the movement of a single-limb. The second type of motions is complicated motion, which is composed of movements of the entire body, such as body twisting, deep squat, boxing, and walking. Two examples of the motions are illustrated in Fig. 6.13. The left-hand-side figure shows a subject simply standing still, and the right-hand-side figure shows the subject is walking. The estimation results for these two examples are presented in Figs. 6.14 and 6.15, respectively, where the estimated pose is marked with red lines, and the Kinect obtained ground truth is marked with blue lines. We also present the estimation results for other complicated motions, including squat, twisting, and kicking, in Figs. 6.16, 6.17, and 6.18, respectively. From these figures, we can see that the estimated poses are all highly close to the ground truth collected by Kinect. These example results show that the RFID-Pose system can adequately estimate the 3D human pose whether the subject is moving or not.

The overall accuracy of human pose estimation is presented in the form of cumulative distribution function (CDF) of estimation errors in Fig. 6.19. The mean error of all the 12 joints for each time slot T is calculated as follows.

$$\epsilon(T) = \frac{1}{12}\sum_{n=1}^{12} ||\hat{P}_n^T - \dot{P}_n^T||, \tag{6.13}$$

Figure 6.13: Illustration of two example poses: (Left) standing still; (right) Walking.

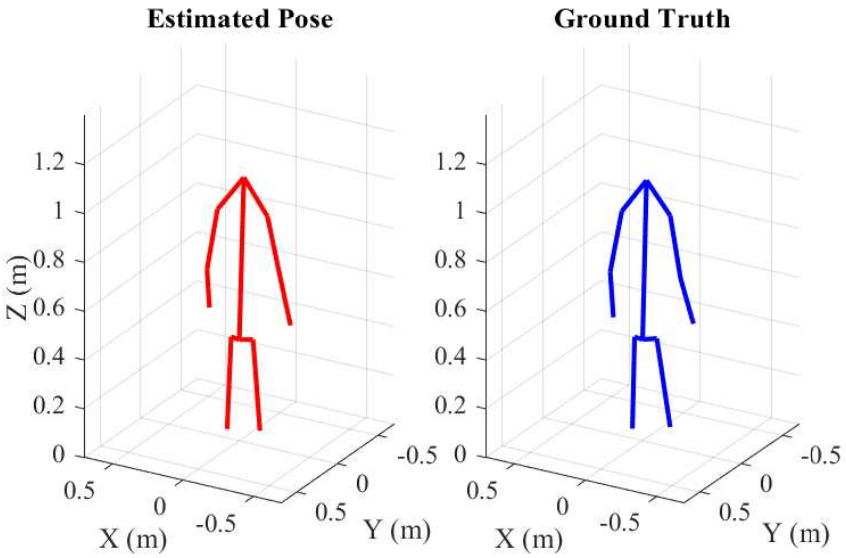


Figure 6.14: Pose estimation when the subject is standing still.

where $\hat{P}_n^T$ denotes the estimated position and $\dot{P}_n^T$ is the ground truth position collected by the Kinect in the 3D space for joint n at time T; and $||\hat{P}_n^T - \dot{P}_n^T||$ is the Euclidean distance between these two 3D vectors. From the CDF curves, we can see that the median estimation error is 2.83cm for the single-limb motion test and 3.75cm for the complicated motion test. The results show that the estimation accuracy of the entire body motion is lower than one-limb motion, because more moving joints need to be reconstructed in the former case. However, RFID-Pose still achieves very high accuracy for all the complicated motions, and the largest error among all the tests is 8.12cm, which is smaller than the maximum estimation error reported in the existing RFID pose estimation system (i.e., 10cm) [134]. The estimation results validates

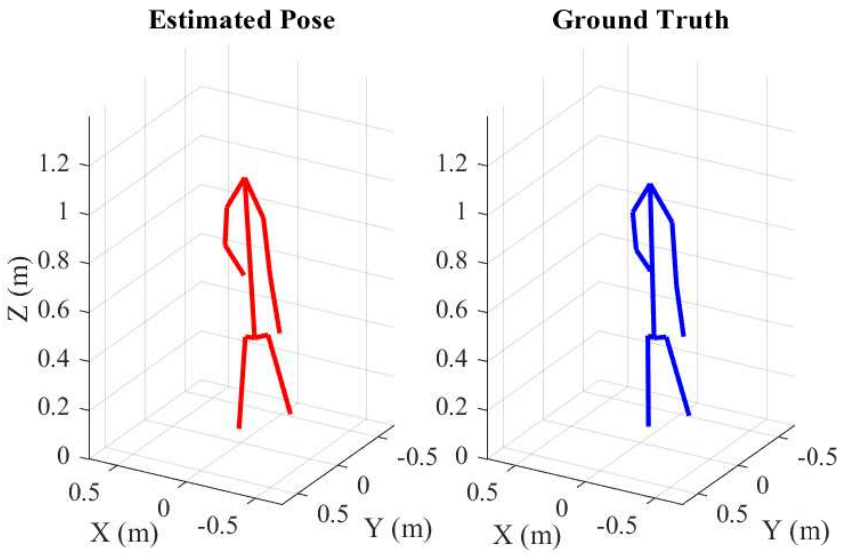


Figure 6.15: Pose estimation when the subject is walking.

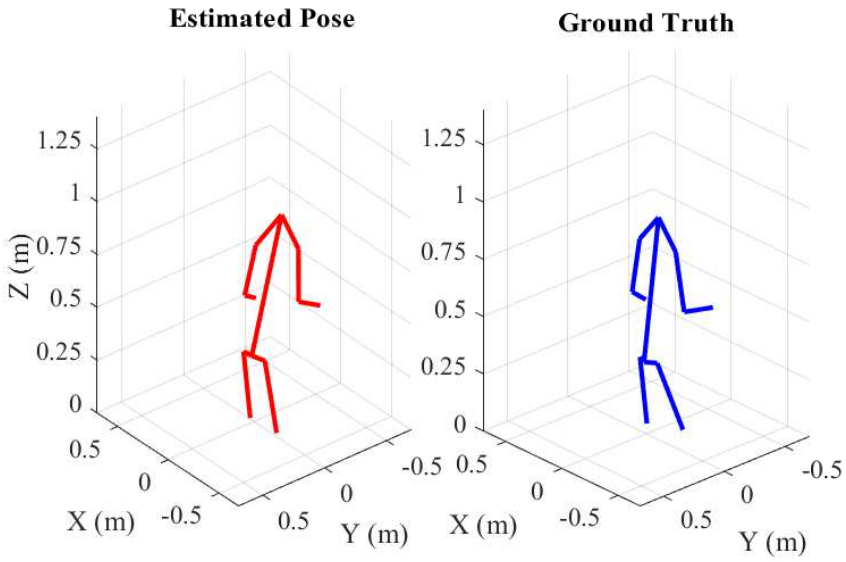


Figure 6.16: Pose estimation when the subject is squatting.

that the proposed RFID-Pose system can estimate the joints position more accurately and can effectively reconstruct the pose of the entire moving body through RFID phase data.

Accuracy for Different Motions

To evaluate the estimation performance for different motions, we plot the accuracy for all the specific movements in Fig. 6.20, including body twisting, squat, waving hands, kicking, walking, boxing, and standing still. As the figure shows, the pose estimation accuracy is different

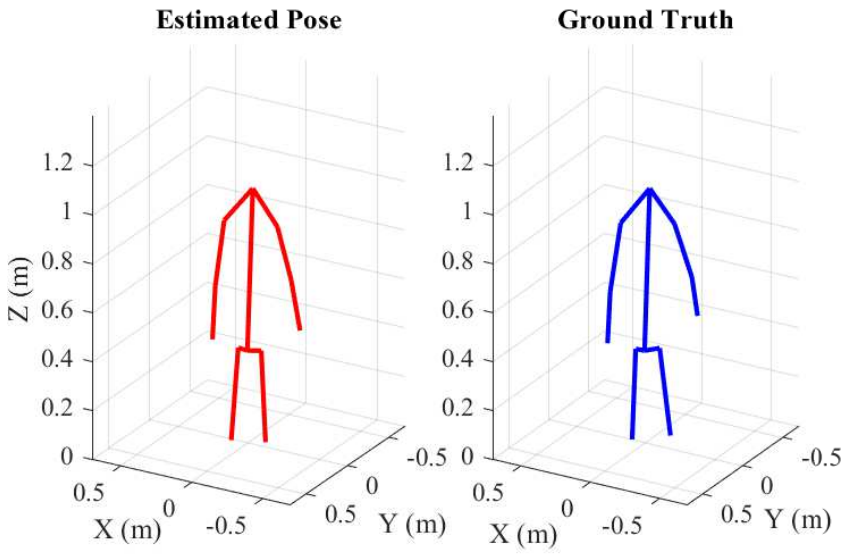


Figure 6.17: Pose estimation when the subject is twisting.

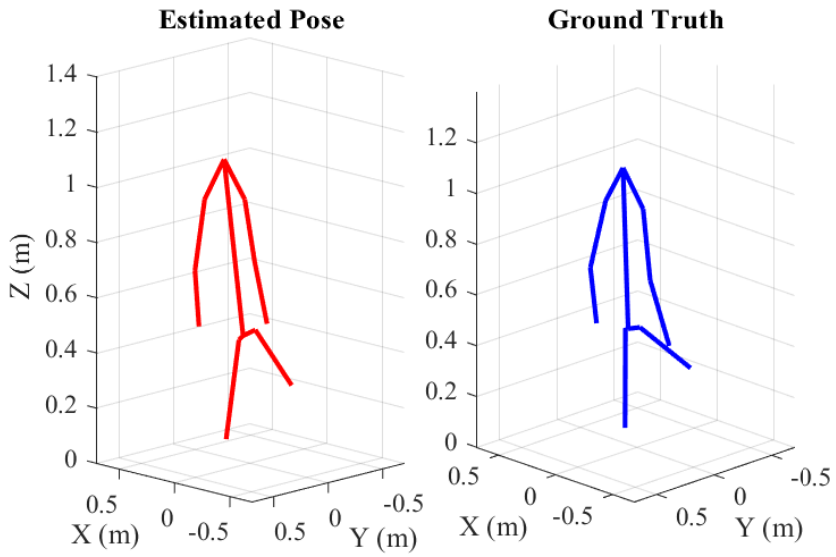


Figure 6.18: Pose estimation when the subject is kicking.

for different motions, where the highest accuracy 1.81cm is achieved when the human is in a stable state (i.e., standing still). This is because no joint is moving when the subject stands still, and thus no joint movements need to be estimated in this case.

We also notice that the squat and walking motions have a worse estimation accuracy than others, which are 5.44cm and 4.12cm, respectively. The pelvis joint position variation is the main cause for the limited performance. Note that our network is designed for learning the spatial rotation of each joint relative to the parent joint. As a root joint of the human skeleton,

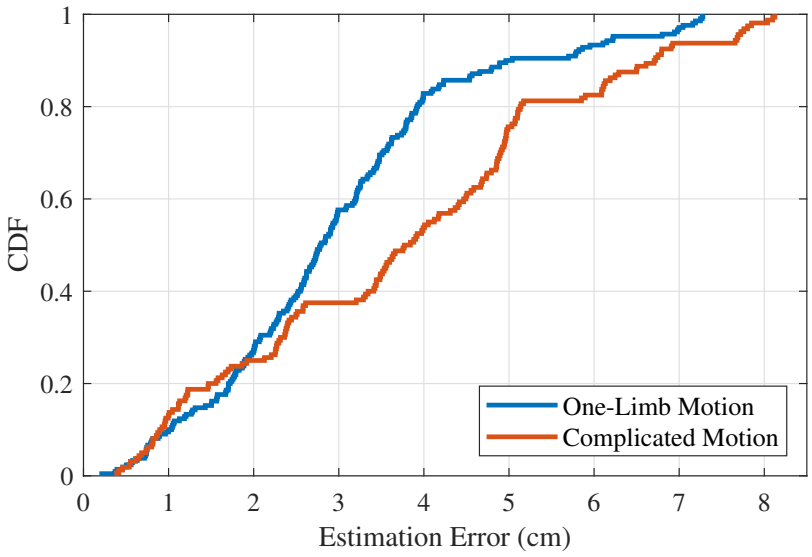


Figure 6.19: Overall pose estimation accuracy in forms of CDF of estimation errors.

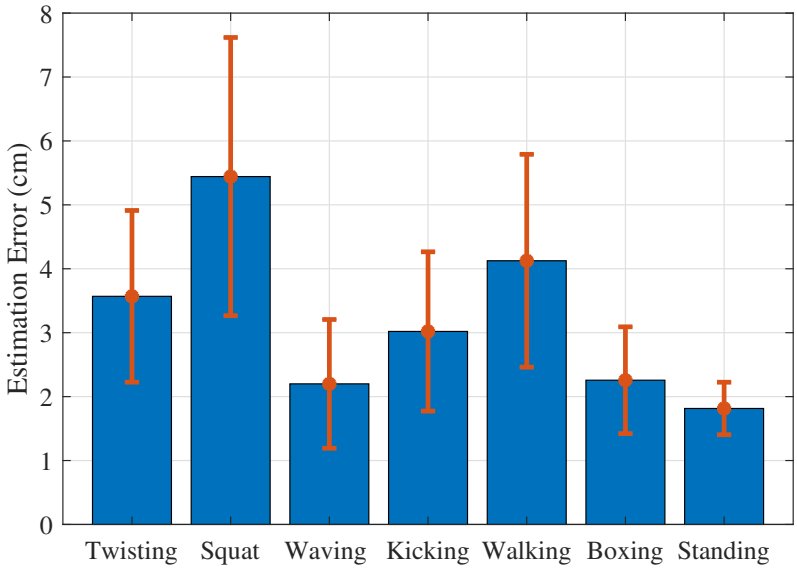


Figure 6.20: Estimation errors for different types of motions.

the pelvis position estimation does not benefit from the forward kinematic layer. Thus, the pelvis joint's position is not as accurate as the rotation angle for each human limb, which also leads to higher errors in all other joints. That is the reason for the lower accuracy when the pelvis joint frequently varies during the monitoring process. Nevertheless, the error 5.44cm is still acceptable for most pose based applications, such as video gaming and motion recognition.

Accuracy for Different Joints

The estimation error for each of the 12 joints is presented in Fig. 6.21. The joint index map is shown in Fig. 6.10. From joint 1 to joint 12, the joints are: pelvis, neck, left hip, left knee, right hip, right knee, left shoulder, left elbow, left wrist, right shoulder, right elbow, and right wrist. As the figure shows, RFID-Pose achieves high estimation accuracy for joints 1, 2, 3, 5, 7, and 10, where the estimation errors are all lower than 3.55cm. The estimation errors for the other joints are all higher than 4.36cm. This is because the joints in the first group are on or close to the human torso, while the other joints are on the limbs (i.e., arms and legs). The relatively worse limbs tracking performance is mainly due to two reasons. First, since the joints of the limbs are tracked based on the torso joints with the forward kinematic technique, the estimation errors of the parent joints on the torso will be accumulated and affect the accuracy of tracking the limb joints. However, the pelvis localization in each time slot is independent, and the estimation error of the pelvis in previous time slots will not be accumulated in the present time slot. Second, since human limbs usually move at a larger extent than the torso joints, there are usually fewer RFID samples for these joints, which leads to a higher estimation error. However, notice that even the wrist estimation error, the highest one, is lower than 5.28cm. Such results prove that the RFID-Post system can accurately estimate the human pose with the vision-aided technique.

6.6.3 More Experiments under Different Scenarios

In addition to evaluating the overall accuracy, we conduct several additional experiments to test the system performance under different scenarios, including different subjects, different environments, and different standing positions in front of the antennas. We also discuss the generalization issue based on the experimental results.

Different Subjects

We conduct experiments with five different subjects to examine the impact of different initial skeletons. The training dataset includes three different subjects, while the other two subjects

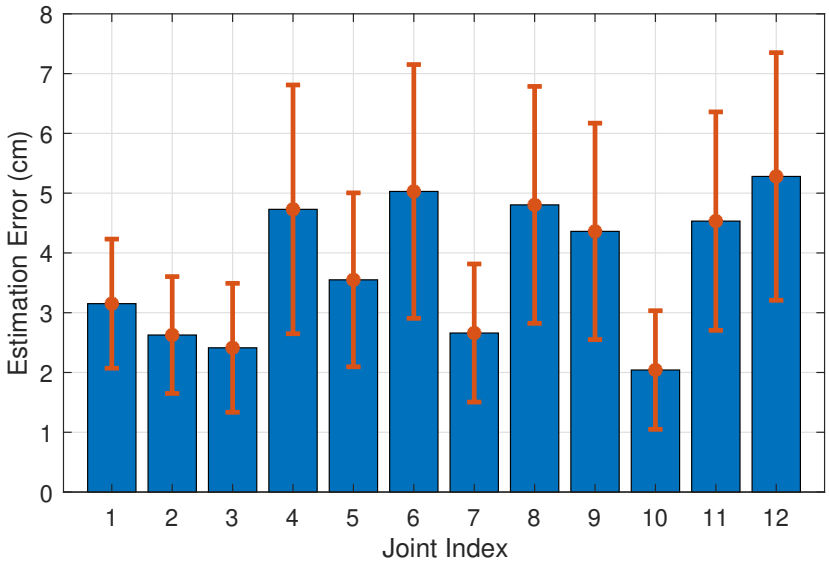


Figure 6.21: Estimation errors for different joints.

Table 6.1: Performance Evaluation For Different Subjects

Subject Index	*Estimation Error*
Subject 1 (trained)	3.72cm
Subject 2 (trained)	4.55cm
Subject 3 (trained)	3.58cm
Subject 4 (untrained)	5.32cm
Subject 5 (untrained)	8.17cm

are not trained but for testing only. The mean estimation errors are presented in Table 6.1. As the table shows, the estimation errors for all the trained subjects are lower than 4.55cm, which means the system can estimate the human skeleton for different subjects. However, when the trained system is used to test the untrained subjects, i.e., subjects 4 and 5, the performance becomes worse but still acceptable. Furthermore, we find that the accuracy for subject 4 is higher than subject 5 because the initial skeleton of subject 4 is similar to trained subject 2. It implies that the performance of testing untrained subject could be improved when the network is trained with more subjects with different skeleton patterns.

Different Environments and Standing Positions

The influence of different environments and standing positions are also investigated. The experiments are conducted in four different environments, including two different locations in the

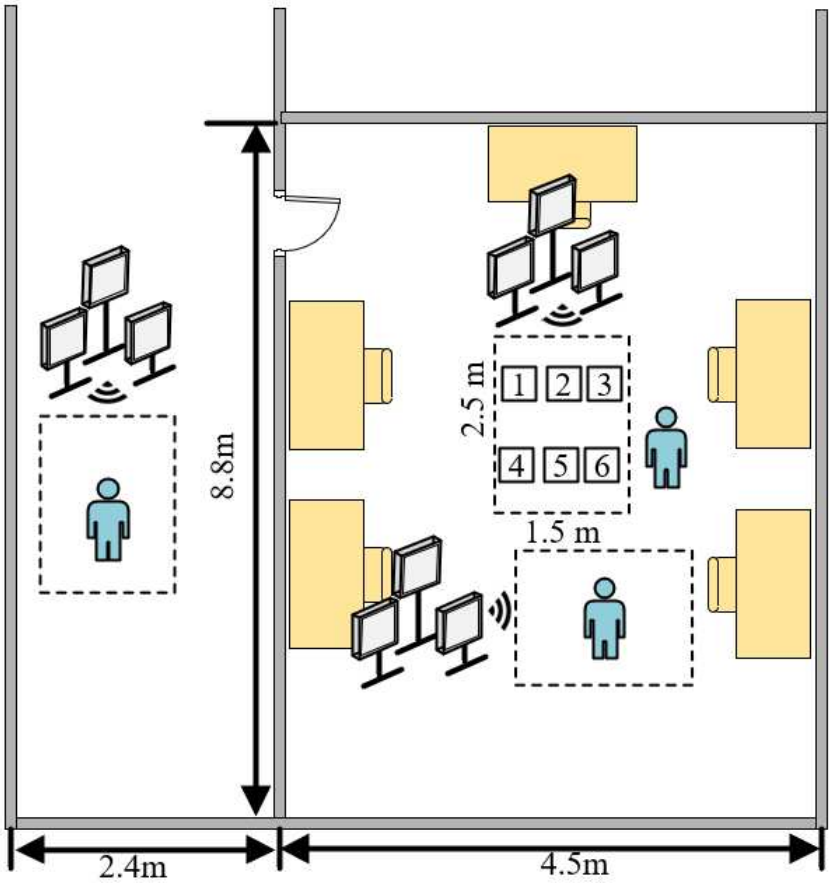


Figure 6.22: Different deployment environments and standing positions.

lab, a corridor, and a living room. The first three environments are illustrated in Fig. 6.22. As the figure shows, the first two locations are selected in the same lab but have highly different deployments, to introduce different environmental interference. The other two locations are selected in the corridor and living room, respectively, which also suffers from quite different multipath effects. As Table 6.2 shows, the estimation error in different environments changes from 3.75cm to 4.03cm, which means the influence of the environments is limited. This is because the received RFID signal is dominated by the line-of-sight component; the other reflected signals are very weak. Thus, the multipath effect from the environment is not strong and does not affect much the performance of RF-Pose.

The interference of different stand positions is also investigated in our experiments. As illustrated in Fig. 6.22, we compare the system performance for six different positions in the 2.5m$\times 1.5$m scanning area in the Lab scenario. Data collected in positions 1, 2, and 3 are used to train the system, while the data collected in positions 4, 5, and 6 are only used for testing.

Table 6.2: Performance Evaluation under Different Environments

Testing Environments	*Estimation Error*
Computer Lab-1	3.83cm
Computer Lab-2	3.90cm
Corridor	4.03cm
Living Room	3.75cm

Table 6.3: Performance Evaluation for Different Standing Positions

Position Index	*Estimation Error*
Position 1 (Trained)	4.53cm
Position 2 (Trained)	3.82cm
Position 3 (Trained)	4.75cm
Position 4 (Untrained)	8.38cm
Position 5 (Untrained)	5.71cm
Position 6 (Untrained)	9.14cm

The estimation errors are presented in Table 6.3. As the table shows, the estimation errors for the three untrained positions 4, 5, and 6 are all higher than 5.71cm, while the errors for the three trained positions 1, 2, and 3 are all lower than 4.75cm. The results show that the estimation accuracy degrades when the subject stands in an untrained position, especially the untrained position near the border of the scanning area. Fortunately, the high accuracy for the trained standing positions shows that the accuracy of untrained positions could be improved by adding more training data sampled from different training positions. Due to limited scanning range of the polarized antennas, six different standing positions for training are sufficient to combat the influence of untrained standing positions.

Remarks on Generalization

Since the initial subject skeleton is needed in the training process, the performance of the proposed system could be affected when testing the subject with an untrained subject or the subject is tested in a different standing position/environment. In RFID-Pose, the initial skeleton is also necessary to address the ill-posed problem caused by the low data rate of RFID systems. This paper is mainly focused on the fundamental problem of transferring sparse RFID data to 3D human skeleton. However, the experiment results shown in Tables 8.1 and 6.3 also

demonstrate that the generalization issue could be mitigated by extending the training dataset for different subjects and standing positions. We will further tackle the generalization problem of RFID based pose monitoring systems in our future work.

6.7 Conclusions

In this paper, we proposed a vision-aided, realtime 3D pose estimation and tracking system named RFID-Pose. A preprocessing module was proposed to effectively mitigate the influence of phase distortion and missing samples in the RFID data. The proposed system then leveraged a deep kinematic network to estimate human postures in realtime from RFID phase data, which was trained with the assistance of computer vision data as labels collected by Kinect 2.0. The RFID-pose system was prototyped with commodity RFID devices. Its high accuracy and real-time operation were demonstrated in our experimental study using Kinect 2.0 as a benchmark.

Chapter 7

Cycle-Pose: Subject-adaptive Skeleton Tracking with RFID

7.1 Introduction

With the rapid development of computer vision, tracking of human poses has become an important problem area in recent years, evolving from 2D [173] to 3D poses [174]. Although camera-based techniques have been shown effective for human pose tracking, such vision-based techniques frequently raise security and privacy concerns. For example, it is reported that millions of wireless security cameras deployed around the world are at risk of being hacked [175]; the video data used for pose tracking could be intercepted and illegally used by hackers. To address this issue, several radio frequency (RF) sensing based schemes have been proposed, such as WiFi [131, 178], Frequency-Modulated Continuous Wave (FMCW) radar [176], and mmWave radar [177].

To this end, radio frequency identification (RFID) provides a promising solution for human pose estimation [157, 179]. Compared with existing contact-free RF sensing systems, RFID tags can be used as wearable sensors because of their small size. Furthermore, the interference caused by the multipath effect is much lower in the RFID system and the cost of RFID systems is lower than the advanced radar based systems. However, because of the low data rate (i.e., the sampling rate) in RFID systems, generating a joint confidence map for all the joints, as in other RF based systems, is highly challenging. Consequently, the existing RFID based pose tracking systems are focused on monitoring the movement of one particular limb using the phase data sampled from multiple tags [134, 135]. When multiple joints are moving

simultaneously, the performance could be affected by the disturbance from other RFID tags (e.g., the mutual coupling effect) or inter-tag collisions.

Subject adaptability is another challenge for RF based human pose tracking. Different people have different skeleton forms, but most of the neural networks incorporated in RF based pose tracking systems are trained with a limited number of subjects [131, 157]. The untrained subjects could be considered as a new data domain in machine learning. When testing in the new domain, the performance of the trained model will be degraded. Transfer learning is a possible solution to address the new domain issues [159, 160], but the trained model needs to be updated by a light-weight training for the new domain. The light-weight training requires new vision data of the untrained subject, which, again, leads to privacy concerns. The domain discriminator proposed in recent works [192, 217] could address the domain-adaptive issue, which, however, only works for the classification problem so far. The model structure may not be suitable for data sequence estimation as in human pose tracking.

In this paper, we address the challenges in human pose estimation using RFID with a novel vision-aided, deep learning based solution. We propose the Cycle-Pose system to track the movements of multiple human limbs in realtime. In Cycle-Pose, RFID tags are attached to the target human joints. The movements of the tags are captured by the phase variations when the tags are interrogated by the reader. A vision-aided solution is proposed to help the deep learning model transform tag phase variations to human limb rotations, rather than localizing them with traditional tag localization techniques [57]. Furthermore, we also proposed a novel *cycle consistent adversarial network model* to achieve subject adaptability. The proposed cycle kinematic network model is trained without the restriction of requiring paired RFID and vision data, such that the network can learn to transform RFID data into a human skeleton for *any* subject. Thus, the system achieves higher subject adaptability than traditional schemes when generating human pose for untrained subjects. In Cycle-Pose, the 3D human pose is reconstructed by estimated rotation angles of human limbs from RFID data and any given initial human skeleton in realtime. A specific benefit is that vision data will not be needed anymore in the inference stage, and thus the user's privacy can be well protected. The main contributions of this paper are summarized as follows.

- To the best of our knowledge, this is the first subject-adaptive 3D human pose estimation system using commodity RFID reader and tags, which can effectively track 3D human pose without vision data in the testing stage.

- We propose a cycle kinematic network model and train the network with self-supervision. The proposed model learns the transformation from RFID data to 3D skeleton for different subjects, to effectively achieve subject adaptability.

- We develop a prototype system with commodity RFID tags/devices and Kinect 2.0, to evaluate the system performance and compare it with the traditional technique RFID-Pose [157]. Our experimental study validates that the proposed Cycle-Pose system can effectively track the human pose for different subjects with subject adaptability.

In the following, we present the system overview in Section 7.2. The challenges and our proposed solutions are discussed in Section 7.3. Our prototype implementation and performance study are presented in Section 7.4. Section 7.5 summaries this paper.

7.2 System Overview

An overview of the proposed Cycle-Pose system is shown in Fig. 7.1, which consists of four main modules: (i) Data Collection, (ii) Data Preprocessing, (iii) Cycle Kinematic Network, and (iv) 3D Skeleton Generation.

(i) *Data Collection:* The Cycle-Pose system aims to generate a 3D human skeleton from collected RFID data. Both RFID data and vision data should be sampled for the training process. The RFID data is collected from 12 RFID tags attached to the human joints by three polarized antennas, and is used as input to the proposed cycle kinematic network. The vision data is sampled by Kinect 2.0 for the same subject and action simultaneously. Kinect 2.0 is a depth camera. It captures 3D human movements by both the RGB camera and infrared sensor. 3D movements of each human joint are generated by processing the Kinect data with MATLAB and stored in the form of 3D coordinates for offline training supervision and used as benchmark in the testing phase.

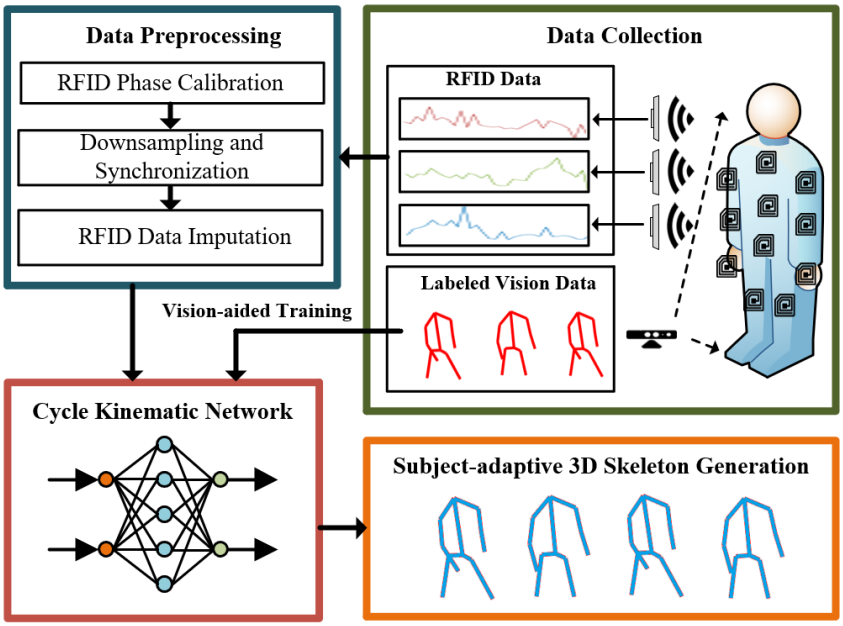


Figure 7.1: The Cycle-Pose system architecture.

(ii) *Data Preprocessing:* However, the collected raw RFID data cannot be directly used for 3D skeleton tracking. It should be calibrated first before analyzed by the proposed neural network model. The collected RFID phase is severely distorted by channel hopping and phase wrapping of the RFID system. It should be firstly calibrated to mitigate such distortion. After that, since the sampling rates of RFID device and Kinect 2.0 are highly different, the sampled RFID data is downsampled and synchronized with the vision data. Because of the slotted ALOHA-like transmissions in RFID systems, the phase data is not evenly sampled; there is at most one sample in each time slot and most RFID phase samples are missing. Thus, we employ the tensor completion technique, High Accuracy Low-Rank Tensor Completion (HaLRTC), to estimate the missing RFID data.

(iii) *User-adaptive 3D Skeleton Generation with the Cycle Kinematic Network:* We propose a cycle kinematic network to generate 3D pose data from calibrated RFID phases. Unlike monitoring only one particular limb's movements as in traditional RFID based pose tracking systems [134, 135], the proposed system estimates the 3D coordinates of all the joints simultaneously. Moreover, the proposed cycle kinematic network achieves subject adaptability, which is missing in prior systems [131, 157]. This is because the cycle kinematic network is trained with unpaired RFID data and vision data, which is sampled from a different moving subject.

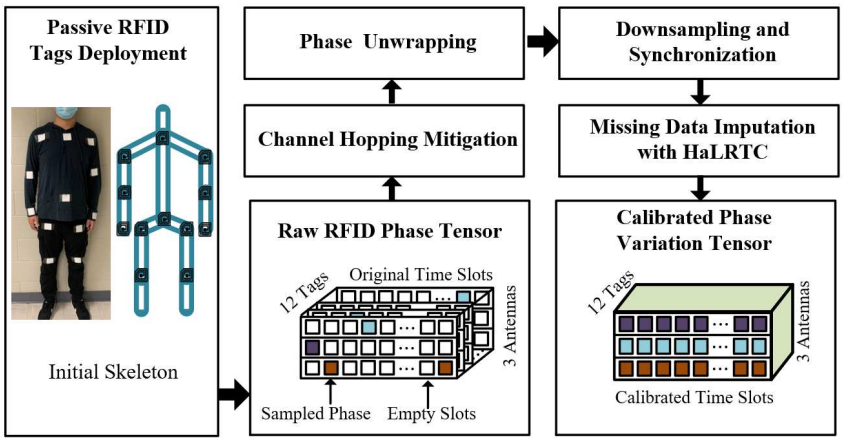


Figure 7.2: Flowchart of RFID data preprocessing.

Thus, the trained network can achieve better adaptability when transforming RFID data to 3D coordinates for a different, untrained subject.

7.3 Challenges and Solutions

7.3.1 RFID Phase Data Calibration

The first challenge in human 3D skeleton generation from RFID data is the poor quality of RFID data. As discussed, the raw RFID data suffers from severe phase distortion and large amount of missing samples. Thus, the RFID data should be well-calibrated before being used to train the neural network. Fig. 7.2 presents the flowchart of the proposed RFID preprocessing procedure. As shown in the figure, the passive tags are attached to 12 joints of the human body. RFID phase data is collected by the reader using the low-level protocol when interrogating the tags [38, 138].

Phase Distortion Mitigation

The collected phase value indicates the distance between the reader antenna and the tag [38]; so the sampled phases captures the movement of the RFID tags attached to the human body. According to the FCC regulation, the frequency of the channel is not fixed but hops among 50 different channels, which generates a different phase offset for each different channel. Consequently, the phase value is also determined by the current channel used for RFID interrogation.

The phase ϕ_s on channel s can be written as [139]:

$$\phi_s = \text{mod}\left(\frac{2\pi 2Lf_s}{c} + \phi_s^0,\ 2\pi\right),\ s = 1, 2, ..., 50, \tag{7.1}$$

where L is the distance between the tag and antenna, c is the speed of light, and f_s and ϕ_s^0 represent the frequency and initial phase offset of channel s, respectively. According to (7.1), if we want to track the variation of the tag-to-antenna distance L, the impact of the channel phase offset ϕ_s^0 should be firstly mitigated. Fortunately, the phase offset ϕ_s^0 is a constant for each channel s. If we use the *variation* between two adjacent phase samples on the same channel, the identical channel phase offset in the two samples will be canceled. The phase variation η_s^n for each channel s is:

$$\eta_s^n = \text{mod}\left(\frac{4\pi(L_s^n - L_s^{n-1})f_s}{c},\ 2\pi\right), \qquad s = 1, 2, ..., 50,\ n = 2, 3, ..., \tag{7.2}$$

where L_s^n represents the tag-to-antenna distance for the nth sample on channel s. It can be seen that the phase offset ϕ_s^0 is removed in (7.2), while the movement $L_s^n - L_s^{n-1}$ remains. The phase distortion caused by channel hopping is effectively mitigated this way.

In addition, phase distortion is also caused by the modulo operation in (7.1) and (7.2). Since the collected phase data ϕ is rounded to the range $[0, 2\pi]$ rad, sharp phase changes are usually generated by the modulo operation when the phase crosses 0 rad or 2π rad. Thus, the calculated phase variation should be unwrapped to remove such distortion. Given the 110Hz sampling rate used in the RFID system, we assume that the phase variation between two adjacent samples, η, should be no larger than π or smaller than $-\pi$. We use the following scheme to unwrap the sampled phase variation data η.

$$\eta' = \begin{cases} \eta - 2\pi\frac{\eta}{|\eta|}, & \text{if } |\eta| > \pi \\ \eta, & \text{otherwise.} \end{cases} \tag{7.3}$$

which automatically determines whether the value should be unwrapped by adding or subtracting 2π. After the unwrapping process, all sharp phase changes will be smoothed out, and the calibrated phase variation data can effectively represent the movements of the RFID tags now.

Data Imputation

In addition to distortion, missing phase samples is another challenge caused by the Slotted ALOHA-like transmission used in RFID communications. In each time slot, only *one* tag can send its EPC and low-level data to the reader. In the Cycle-Pose system, although we attach 12 tags to the human body, only one tag can be sampled at a time. In the input data tensor illustrated in Fig. 7.2, there is only one sample in each slice (12 tags $\times$ 3 antennas). Thus, the sparsity of the phase data tensor is $35/36$, which is way too high for pose estimation. In order to learn the relationship between RFID data and 3D skeleton data obtained from Kinect, we need to (i) deal with the high sparsity issue in the tensor data and (ii) synchronize the phase data (i.e., the input to the deep learning model) and the vision data (i.e., the labels). To solve these problems, we first downsample the RFID data from 110Hz to 30Hz to match the 30fps sampling rate of Kinect. Then, we synchronize the RFID and vision data based on the timestamps when the RFID and vision data samples are simultaneously collected for the same subject. Note that we cannot synchronize the data from different subjects because the RFID data and vision data are not sampled simultaneously in this case.

After synchronizing the two types of data, the input tensor to the deep learning model can be expressed as:

$$\boldsymbol{H}(:,:,t) = \begin{bmatrix} \eta_{t,1}^{1} & \eta_{t,2}^{1} & \cdots & \eta_{t,n_G}^{1} \\ \eta_{t,1}^{2} & \eta_{t,2}^{2} & \cdots & \eta_{t,n_G}^{2} \\ \vdots & \vdots & \vdots & \vdots \\ \eta_{t,1}^{n_A} & \eta_{t,2}^{n_A} & \cdots & \eta_{t,n_G}^{n_A} \end{bmatrix}, \ t = 1, 2, ..., N_t, \tag{7.4}$$

where t means the tth time slot, n_A and n_G are the numbers of antennas and tags, respectively, and $\eta_{t,n_G}^{n_A}$ represents the calibrated phase variation from tag n_G sampled by antenna n_A in time slot t. To address the high sparsity of the tensor, we leverage tensor completion to estimate the

missing samples in $\boldsymbol{H}$. The algorithm used in the system is HaLRTC [139], which can achieve high accuracy in data imputation at a relatively high speed.

7.3.2 Skeleton Generation from RFID Data

Three-dimension human skeleton generation with RFID data is highly challenging also because of the extremely low sampling rate restriction in RFID systems. Most of the existing human pose tracking system is based on the confidence map generated from collected signals, such as camera [173], WiFi [178], and FMCW radar [176]. The human features is first captured and shown on the confidence map, so the human skeleton can be further constructed based on the map. However, this technique is not suitable for RFID based systems, because the sampling rate of the RFID system is much lower than that of video camera, WiFi, and FMCW radar. This is because the RFID system is designed to interrogate RFID tags one at a time, which means no matter how many tags are used in the system, the maximum data rate is fixed by one sample per time slot. If we want to generate a confidence map video with 10fps from RFID data with 110Hz sampling rate, only the 11 phase samples (i.e., sampled at the same time as one video frame or 11 pixels) can be used for map generation. Even if we reduce the map resolution to 100×100, transforming the 11 samples to $10{,}000$ pixels is a severely *ill-posed problem*, which is challenging for training the deep learning model.

In this paper, we employ the forward Kinematic technique to tackle the ill-posed problem, which is widely used in robotics and 3D animation [193]. With a given initial skeleton (i.e., the original locations of all joints and the lengths of all limbs), forward kinematic can estimate the joint position based on the relative space rotation and its parent joint position. For example, when the right elbow position is given, the right-hand position of the subject can be calculated by the length of the front arm and the relative rotation between the hand and elbow. In the proposed cycle Kinematic network, a 3D rotation is represented in a *unit quaternion* format based on Ruler's rotation theorem, expressed as:

$$r + x\vec{\alpha} + y\vec{\beta} + z\vec{\gamma}, \tag{7.5}$$

where r, x, y, and z are real numbers, and $\vec{\alpha}$, $\vec{\beta}$, and $\vec{\gamma}$ are the quaternion units related to the three coordinates, respectively. Given the 3D position of a joint, represented as $a\vec{\alpha} + b\vec{\beta} + c\vec{\gamma}$, and a 3D rotation with unit quaternion $r + x\vec{\alpha} + y\vec{\beta} + z\vec{\gamma}$. The rotation matrix $\mathbf{\Omega}$ can be derived as:

$$\mathbf{\Omega} = \begin{bmatrix} 1-2(y^2+z^2) & 2(xy+zr) & 2(xz-yr) \\ 2(xy-zr) & 1-2(x^2+z^2) & 2(yz+xr) \\ 2(xz+yr) & 2(yz-xr) & 1-2(x^2+y^2) \end{bmatrix}. \tag{7.6}$$

The updated position of the joint, $a'\vec{\alpha} + b'\vec{\beta} + c'\vec{\gamma}$, will be calculated as:

$$\begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = \mathbf{\Omega} \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \tag{7.7}$$

With the forward kinematic technique, the current human pose is determined by the previous human pose and the 3D rotation for each body joint. To estimate the positions of the 12 human joints, only 48 parameters are required. Compared with the traditional approach of generating a $10{,}000$-pixel map, the proposed technique effectively reduces the problem complexity and can improve the accuracy as well.

7.3.3 Dealing with Subject Adaptability

The forward Kinematic technique can effectively address the ill-posed problem. However, the initial skeleton of the subject is still needed, which limits the adaptability of the trained model to different untrained subjects. People have different skeleton forms. To make sure that the deep learning model can successfully generate 3D skeleton for different subjects, the training dataset should include all kinds of human skeletons, which leads to a significantly high cost on collection of labeled training data. If the network is trained with a limited amount of skeletons, the performance of the network could be poor when testing for a subject with a new skeleton

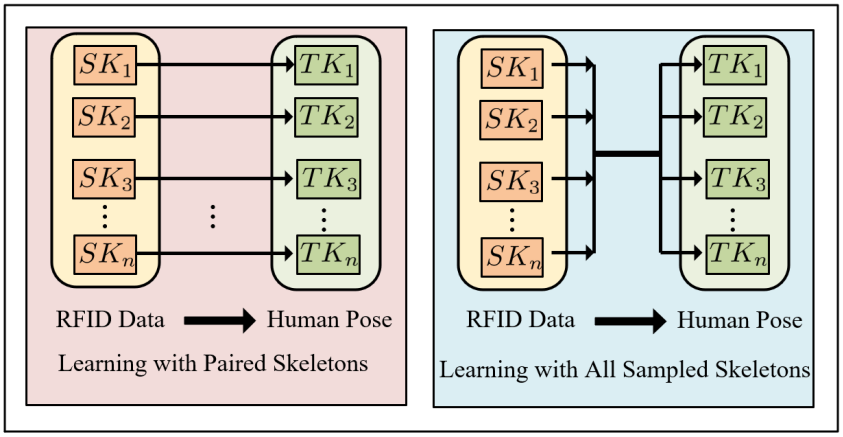


Figure 7.3: Different structures for pose generation training.

not included in the training dataset [131]. This is because the traditional training process is performed with the same source initial skeleton and target initial skeleton, which is illustrated in the left-hand-side graph in Fig. 7.3. In the figure, SK_n represents the source initial skeleton for subject n in the RFID data, while TK_n represents the target initial skeleton in the vision data with $TK_n = SK_n$. The traditional training structure is focused on learning the relationship between 3D skeleton coordinates and the RFID data for the same skeleton. Thus, the training results is suitable for these n specific skeletons included in the training data, but the well-trained model may not perform well when it is used to test a new skeleton.

Cross-skeleton Learning Structure

To improve the subject adaptability of the learning model, the network should be designed to learn the relationship between different source and target skeletons, so that the system could effectively transform RFID data to 3D skeleton no matter the given subject skeleton is included in the training dataset or not. Thus, we propose a new network structure as illustrated in the right-hand-side graph of Fig. 7.3. As shown in the figure, the training is not only focused on learning with paired skeletons, but also leaning with different source skeletons and target skeletons. For example, for a specific movement type such as kicking, all RFID data and vision data are utilized in training, no matter the movement data is sampled from the same subject or not. Thus, the network can learn how to transfer RFID data to human 3D pose with different initial skeletons, such as Sk_1, Sk_2, ..., and Sk_n. Since the network is not trained with a specific

Figure 7.4: Labeled pose data sampled by Kinect for different subjects. The first row is for Subject 1 and the second row is for Subject 2.

initial skeleton, the well-trained model can achieve higher subject adaptability compared with the traditional network structure.

Unfortunately, training with different SKs and TKs is challenging because there is a significant variance in training data between two different subjects. Although performing the same movement, different subjects could have very different speeds and scales, as illustrated in Fig. 7.4. Their limbs have different lengths and it is also hard to guarantee that each RFID tag will be attached at exactly the same location. Fig. 7.4 shows the skeleton obtained by Kinect when two different subjects perform the same action (i.e., arm waving), sampled at the same frame rate. As shown in the figure, both the hand moving speed and the latitude of the arm are very different between the two data sequences shown in the first row (for Subject 1) and the second row (for Subject 2).

The considerable difference shown in Fig. 7.4 indicates that the network should not be directly trained with the position loss between estimated pose and vision pose for different subjects. A *self-supervised network* should be designed for cross-skeleton learning with unpaired initial skeletons.

Cycle Kinematic Network

Cycle-Consistent Adversarial Network is an advanced neural network structure, which is proposed to solve the image-to-image translation with unpaired training datasets [181]. The cycle consistent network has also been employed to solve the temporal video alignment problem of two different video streams [171]. The cycle consistent network can generate fake input data from the output data, so the network can be trained with self-supervision between the real input data and the generated fake input data. Thus, the requirement on paired, labeled data is lower than that in traditional neural networks.

Observing the strength of cycle consistent adversarial network, we propose a Cycle Kinematic Network to deal with the challenges in training for cross-skeleton learning, which is presented in Fig. 8.2. As shown in the figure, the RFID phase variation sequence feature is first extracted by a recurrent encoder termed recurrent encoder-Forward (or, *recurrent encoder-F*). With additional input of the source initial skeleton of the subject, the recurrent decoder-Forward (or, *recurrent decoder-F*) translates the human movement features captured by RFID phase variations to *unit quaternion*, which represents the 3D rotation of the subject's joints. Then, the unit quaternion is employed by the forward Kinematic algorithm to generate 3D human skeleton with a given target initial skeleton. The cycle consistent network is used to recover the RFID data from the estimated quaternion, which is also constructed by the recurrent encoder-Backward and decoder-Backward (or, *recurrent encoder-B* and *recurrent decoder-B*). If the translation from RFID phase variation to 3D limb rotation exists, the inverse transformation could be performed using recurrent autoencoder-B. With the fake RFID data, the network can be trained with self-supervision, and the ground truth of vision data does not need to be strictly paired with the input RFID data.

Loss Function for Training

The loss function used for the training the proposed cycle kinematic network is composed of three parts as illustrated in Fig. 7.5, including the position loss, the adversarial loss, and the cycle consistency loss. When the training step is set to K, we define the calibrated RFID

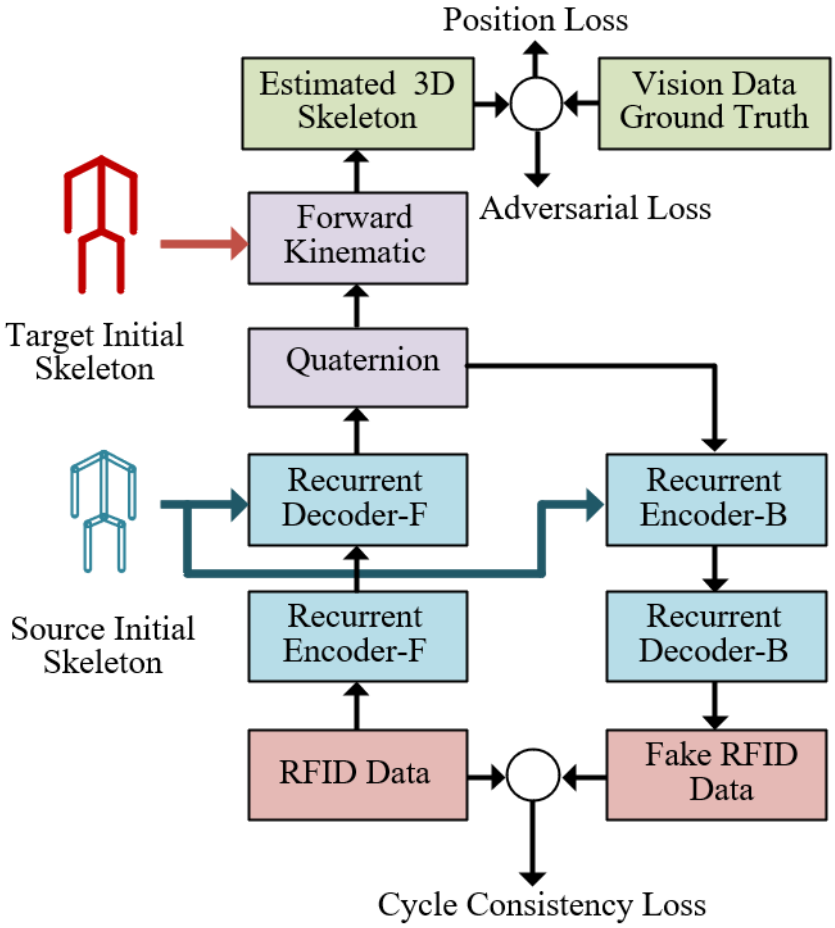


Figure 7.5: Overview of the proposed cycle kinematic network model.

phase variation sequence as $F_{1:K}$ and the reconstructed fake RFID data as $\hat{F}_{1:K}$. The estimated skeleton by the neural network is represented as $\hat{V}_{1:K}$, and the vision data sequence used for supervision is denoted by $V_{1:K}$. The position loss between the estimated 3D skeleton and the ground truth is calculated with the estimated skeleton and the vision skeleton ground truth as:

$$Loss_p = ||\hat{V}_{1:K} - V_{1:K}||_2^2. \tag{7.8}$$

However, for the unpaired training data collected from different skeletons, $Loss_p$ also includes the error caused by the asynchronous training dataset. We can calculate the cycle consistency loss as:

$$Loss_c = ||\hat{F}_{1:K} - F_{1:K}||_2^2. \tag{7.9}$$

Since the cycle consistency loss is calculated with the fake RFID data obtained by the cycle consistent network, the influence of unpaired data can be mitigated by merging the cycle consistency loss and position loss as:

$$Loss_{all} = Q_p \cdot Loss_p + Q_c \cdot Loss_c, \tag{7.10}$$

with suitable positive coefficients Q_p and Q_c, satisfying $Q_p + Q_c = 1$. With the loss function of the generator $Loss_{all}$, the network can be effectively trained whether the RFID data and vision data are sampled from the same subject or not. In this paper, we set $Q_p = 0.6$ and $Q_c = 0.4$.

The adversarial loss is used to determine if the network is well trained or not, which is represented as a realism score calculated by a discriminator network D [193], as:

$$S_D = D(\hat{V}_{2:K} - \hat{V}_{1:K-1}, V_{2:K} - V_{1:K-1}). \tag{7.11}$$

The equation shows that the input of the discriminator is not the position loss but the variation between the previous frame and the current frame in V and $\hat{V}$, respectively. Although V and $\hat{V}$ are unpaired data sequences, the discriminator can determine if the movements performed by the two subjects are the same or not. This is because, for the same movement type, the variations of all the joints between two adjacent data sequences are still similar, no matter the two subject movements are synchronized or not. We set a realism score threshold to balance the discriminator and the generator (i.e., recurrent encoder-F and recurrent decoder-F). When the generator can successfully fool the discriminator, the network will effectively transform RFID data to 3D skeleton data.

7.4 Implementation and Evaluation

7.4.1 Prototype System Implementation

To evaluate the performance of Cycle-Pose, we develop a prototype system with an off-the-shelf Impinj R420 reader equipped with three S9028PCR polarized antennas. The RFID tags used in Cycle-Pose are ALN-9634 (HIGG-3). The vision data, used for training supervision

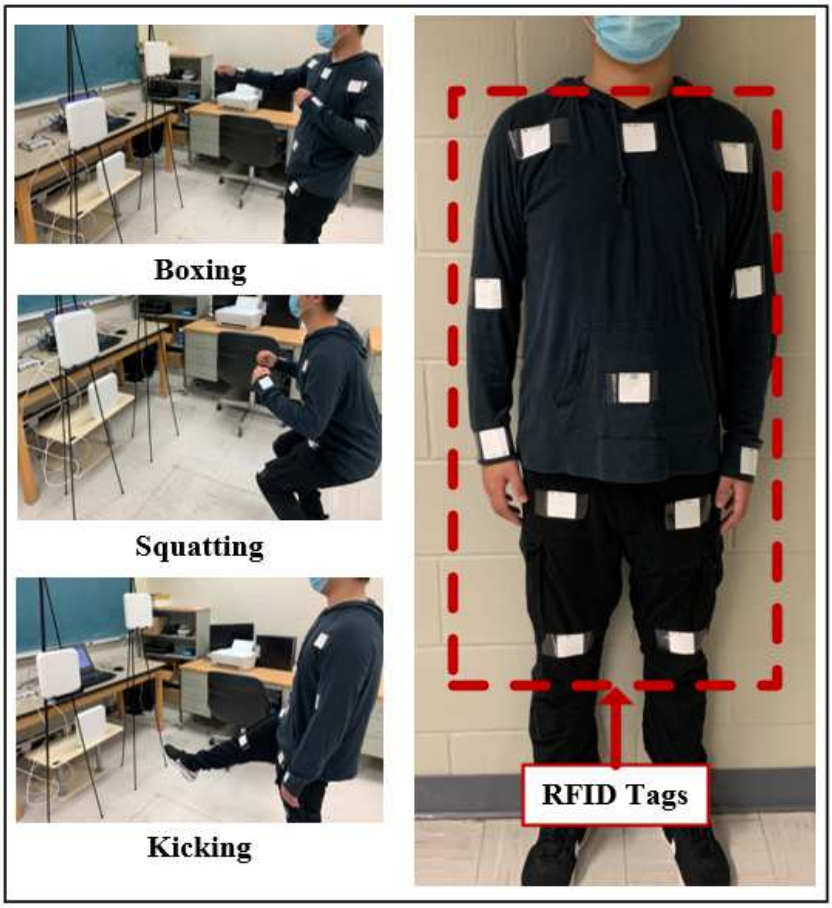


Figure 7.6: RFID tag deployment and motion sampling.

as well as the ground truth for test accuracy evaluation, is collected with an Xbox Kinect 2.0 device.

We attach 12 RFID tags to the human body joints as shown in Fig. 7.6, including the left shoulder, left elbow, left wrist, right shoulder, right elbow, right wrist, neck, pelvis, left hip, left knee, right hip, and right knee. Head and feet are omitted in our prototype system because of the limited scanning range of the RFID antenna used in Cycle-Pose. More antennas can be added to scan the entire body, but the skeleton constructed with the 12 joints is sufficient to monitor human activities in most cases. With the three antennas placed at different altitude positions, every RFID tag can be scanned by at least one of the antennas.

As Fig. 7.6 shows, RFID phase data is collected when the subject is standing in front of the antennas and performing specific motions repeatedly. Different motions are sampled for

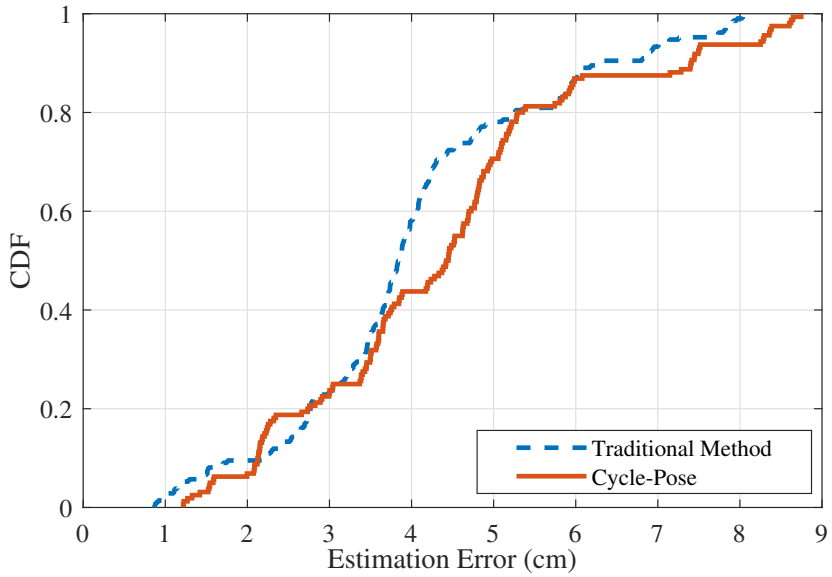


Figure 7.7: Overall accuracy when testing with trained subjects.

training the cycle kinematic network with two different types. The first type includes simple motions, which only involve single-limb movement. The other type includes compound motions, composed of movements of the entire human body, such as boxing, walking, body twisting, and deep squatting.

7.4.2 Experimental Results and Analysis

To evaluate the performance of Cycle-Pose, we compare it with the traditional neural network model used in RFID-Pose [157], which only trains the network with paired RFID and vision data. The dataset used for training and testing is the same for both models, and the overall accuracy is shown in Fig. 7.7 and Fig. 7.10. The overall estimation error $\mathcal{E}_{all}$ used in our evaluation is calculated between the estimated 3D pose data and the ground truth vision data as:

$$\mathcal{E}_{all} = \frac{1}{12}\sum_{n=1}^{12}||\hat{P}_n - \dot{P}_n||, \tag{7.12}$$

where $\hat{P}_n$ denotes the estimated position of joint n, $\dot{P}_n$ is the ground truth position collected by Kinect for joint n in the 3D space, and $||\hat{P}_n - \dot{P}_n||$ is the Euclidean distance between the two 3D coordinates.

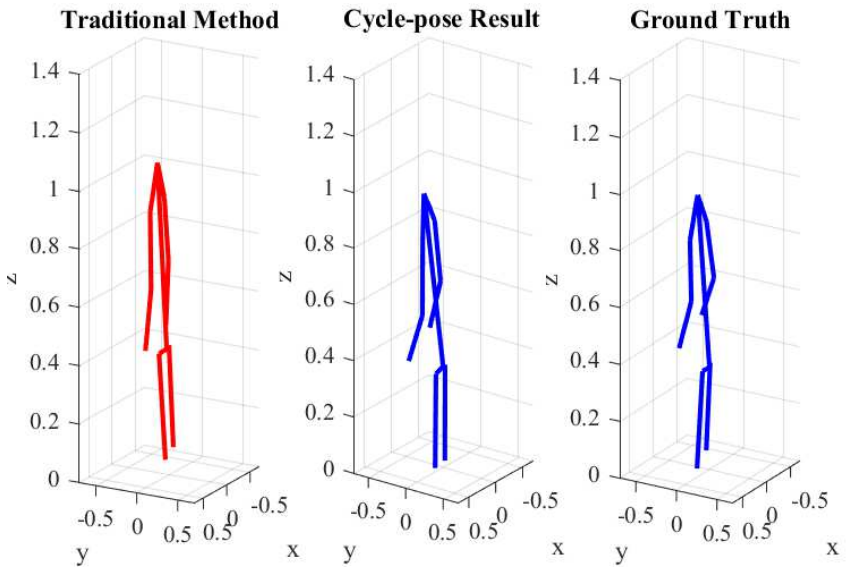


Figure 7.8: Comparison results when the untrained subject is squatting.

Fig. 7.7 presents the cumulative distribution functions (CDF) of the estimation errors for both methods when the testing subjects are involved in the neural network training. The CDF curves show that the median estimation error for the traditional network is 3.83cm, while the median error for Cycle-Pose is 4.44cm. The maximum error for Cycle-Pose is 8.64cm, which is slightly higher than that of the traditional method (i.e., 8.09cm). These results show that the accuracy of Cycle-Pose is slightly lower than the traditional method when testing a trained skeleton. This is because the Cycle-Pose system not only learns the translation from RFID data to 3D skeleton, but also learns the transformation from different source skeletons to target skeletons. The additional learning task affects the system performance for specific skeletons. However, the decrease of the accuracy is acceptable in most skeleton tracking applications, such as video gaming and human motion recognition.

The strengths of the Cycle-Pose system become obvious when testing with untrained subjects. Figs. 7.8 and 7.9 illustrate the comparison between two networks when an untrained subject is squatting and walking, respectively. From the figures, it can be seen that the human poses reconstructed by the Cycle-Pose system are highly similar to the corresponding ground truth, while the skeletons generated by the traditional method show higher estimation errors.

The accuracy results are presented in Fig. 7.10. As the CDF results show, the median estimation error of the Cycle-Pose system is 4.88cm, while the median error of the traditional

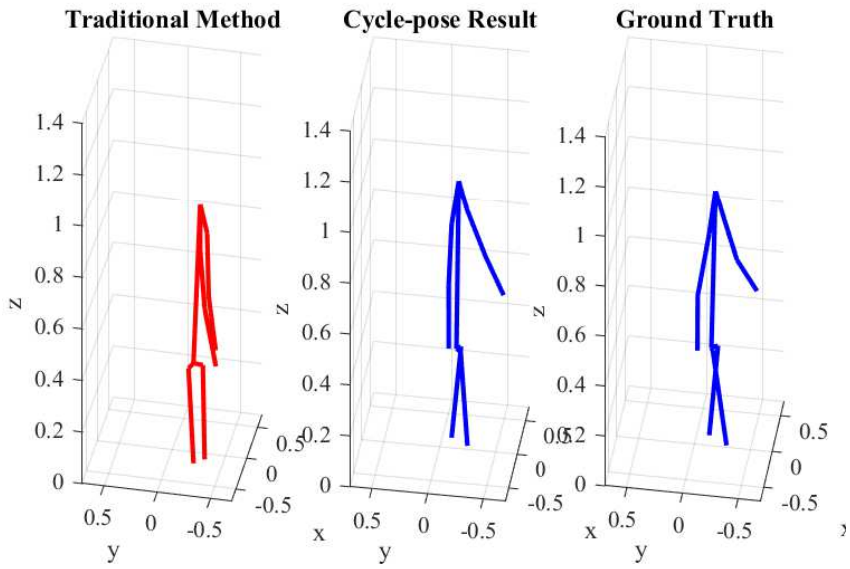


Figure 7.9: Comparison results when the untrained subject is walking.

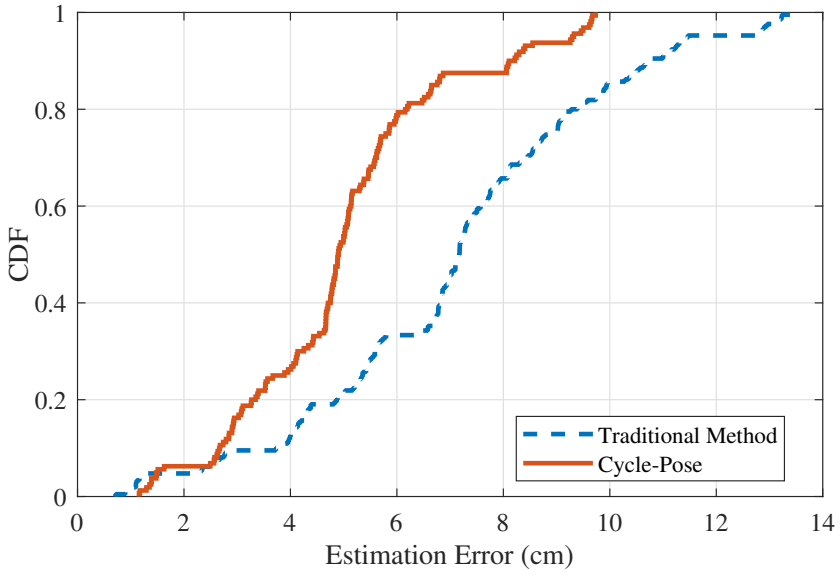


Figure 7.10: Overall accuracy when testing with untrained subjects.

system is 7.66cm, which are both higher than that in Fig. 7.7. The traditional network is only trained by paired RFID and vision data for the same subject. The training domain is restricted to the specific initial skeletons. When testing with an untrained subject with a different initial skeleton, the traditional network exhibits poorer subject adaptability than the Cycle-Pose system. In summary, although the accuracy of the Cycle-Pose system is slightly lower than that of the traditional RFID pose tracking technique when testing with a known subject, the proposed model achieves high subject adaptability when testing untrained subjects.

7.5 Conclusions

In this paper, we proposed a subject-adaptive, realtime 3D pose estimation and tracking system named Cycle-Pose. A preprocessing module was proposed to effectively mitigate the influence of phase distortion and missing RFID data samples. The proposed system then leveraged a novel cycle kinematic network to estimate human postures in realtime from RFID phase data, which was trained with unpaired RFID and vision data sampled from different subjects. The Cycle-Pose system was implemented with commodity RFID tags/devices and compared with a traditional RFID based technique RFID-Pose in our experimental study. Its high subject adaptability ability and accuracy were demonstrated in our comparison experimental study using Kinect 2.0 as ground truth.

Chapter 8

Meta-Pose: Environment Adaptive RFID based 3D Human Pose Tracking with a Meta-learning Approach

8.1 Introduction

Human pose tracking has attracted great interest in recent years, because it is useful for numerous applications such as human-computer interaction, video surveillance, and somatosensory games. The advances in human pose tracking have been mainly driven by the new developments in computer vision, from two-dimensional (2D) systems [173] to three-dimensional (3D) realtime systems [174]. However, The vision-based techniques often raise concerns of security and privacy. For example, many wireless security cameras are easily hacked by malicious users [175]. The collected video data for pose tracking could also be illegally intercepted. Several radio frequency (RF) sensing schemes have been proposed to address the privacy concern in human pose tracking, using various RF sensing techniques such as Frequency-Modulated Continuous Wave (FMCW) radar [176], millimeter wave (mmWave) radar [177], WiFi [131, 178], and RFID [157, 179, 180]. Compared with computer vision-based techniques, RF sensing-based human pose tracking does not require for sufficient lighting, does not require a line-of-sight path between the subject and camera (i.e., capable of getting around obstacles or even through walls), and more important, the privacy of users can be better protected.

In RF sensing-based systems, deep learning has been widely adopted to translate captured RF data to human poses. However, such techniques usually have the generalization problem, i.e., the inference performance usually degrades greatly when applying a well-trained deep learning model to a new, un-trained environment. Since RF signals propagate in the open air, the received RF signal is usually highly sensitive to the specifics of the deployed environment,

such as the placement of the antennas, the layout (e.g., walls) of the room, the obstacles in the surroundings, and the movement of objects and/or subjects nearby. When there are variations in the environment, the same human subject performing the same activity could generate considerably different RF features. It has become a great challenge to develop human pose tracking schemes that are adaptive to the environment.

Researchers have proposed several solutions to address the environment adaptation challenge. The most straightforward approach is to simply increase the size and variety of the training dataset, i.e., to train the deep learning model with vast amounts of data measured in many types of environments. When applying the trained model to a new environment, it is likely that new environment will be similar to an environment that exists in the training dataset, and thus the inference performance would not degrade much. However, this approach requires considerable efforts and incurs high costs in collecting large amounts of training data. In addition, more sophisticated schemes leverage the idea of *adversarial learning* to improve feature extraction [181, 217]. Rather than training using data collected from numerous RF environments, adversarial learning incorporates a *domain discriminator* to distinguish features from different environments (i.e., domains). When the model is trained such that it is capable of fooling the discriminator, the features that are common to all the domains will be extracted. The domain adversarial network can effectively reduce the requirement on training data.

Alternatively, when applying the well-trained (or, pretrained) model to a new, unknown domain, we can fine-tune the model by further training it with new data collected from the new environment, such that the pretrained model can better capture the specific features of the new domain. The fine-tuning technique has been shown effective to address the generalization problem found in other deep learning applications [183]. Fine-turning still requires new data measured from the unknown domain. However, such new data should be as few as possible; otherwise, it will still incur great efforts and a high cost, which hinder the easy deployment of the technique in practice. To this end, meta-learning, a.k.a. "learning to learn" [184], provides an effective solution. Meta-learning optimizes the deep learning model using different learning tasks or datasets [185], so that the model will be appropriately initialized and be amenable to adapt to new domains. When applied to a new RF environment, the meta-learning model will

only require a few training examples from the new environment for fine-tuning (i.e., few-shot fine-tuning), while still achieving a satisfactory performance.

In this paper, we tackle the environment/domain adaptation challenge with a meta-learning approach. We propose a novel environment-adaptive, RFID based 3D human skeleton tracking system termed *Meta-Pose*. As in our prior work RFID-Pose [157], RFID tags are attached to the human body and interrogated by an RFID reader, such that the movements of human joints will be captured by analyzing the phase information in received RFID responses. Meta-Pose is also a vision-assisted scheme, where Kinect captured video data of the same human activity is used for supervised training. Note that the vision data will only be used for training the deep learning model; it will not be needed in the inference stage. Therefore the use of Kinect does not cause privacy concerns. To address the generalization problem, we first analyze the main causes for the divergence of RFID data in different RF environments. Based on the analysis, we then propose a novel Meta-Pose initialization algorithm to pretrain the model with RFID data sampled from a few different environments. With few-shot fine-tuning, the Meta-Pose system will be able to accurately track 3D human skeleton in a new, unknown environment. Extensive experiments are conducted to validate the high environment adaptation ability and high accuracy of the proposed Meta-Pose system.

The main contributions of this paper are summarized in the following.

- To the best of our knowledge, Meta-Pose is the first environment-adaptive system for 3D human pose tracking, which is designed using off-the-shelf RFID reader and tags. Meta-Pose can be easily deployed to estimate and track 3D human poses with RFID data in different RF environments.

- We analyze the divergence of RFID data measured in different propagation environments and identify the main challenges to the generalization problems, including sensitivity divergence of RFID tags and phase distortion in different sampling environments.

- We propose a novel Meta-Pose initialization algorithm based on meta-learning algorithms (i.e., model-agnostic meta-learning (MAML) and Reptile) to pretrain the deep learning model with a limited number of training datasets sampled from several known

environments. We develop the initialization approaches based on both Reptile and MAML. A domain fusion technique is incorporated to generate more synthesized (or, fake) environments for model pretraining, to allow the pretrained model be quickly adapted to a new environment.

- We develop a prototype with off-the-shelf RFID tags and reader, and use Kinect 2.0 to measure the ground truth data for training the model and for performance evaluation. The performance of Meta-Pose is validated with extensive experiments as well as a comparison study with a baseline scheme termed RFID-Pose developed in our prior work [157]. The experimental results show that the proposed Meta-Pose system can accurately track 3D human poses while achieving high environmental adaptability simultaneously.

In the remainder of this paper, Section 8.3 briefly summarizes and contrasts with related works. The background of the proposed system is presented in Section 8.3. Section 8.4 examines the challenges of the domain adaptation problem. Section 8.5 presents our meta-learning based solution to these challenges. Our implementation and experimental study are presented in Section 8.6. Section 8.7 summaries this paper.

8.2 Related Work

In this section, we examine the related work on human pose estimation and tracking, which can be roughly classified into video camera-based schemes, WiFi-based schemes, radar-based schemes, and RFID-based schemes.

8.2.1 Traditional Pose Tracking Systems

A strength of the traditional camera, WiFi, and radar-based systems is that they are "marker-less" methods, which are less intrusive. Video camera was first used to detect human poses in [186, 187]. With deep learning models, such systems localized the coordinates of human joints in the captured video frames, using, e.g., 2D RGB cameras [173, 188] or 3D depth cameras [189]. The most accurate 3D pose tracking performance was achieved, so far, by the Vicon system [190], which has been widely used for production of 3D movies. However, such video

based schemes usually raise privacy concerns, as discussed, and their performance is usually limited by poor illumination, cluttered background, or poor camera angles.

To address the privacy concerns and mitigate the dependency on lighting and background, several RF pose tracking techniques have been proposed. Since such systems record no vision data and the RF data is not visible, user privacy can be better preserved. Furthermore, RF sensing systems perform well in poorly lighted environments and are able to detect human poses through obstacles and walls [177, 191]. FMCW Radar was first utilized to construct both 2D and 3D human poses by incorporating a vision-aided teacher-student deep learning model [176, 191]. As another type of non-intrusive sensor, WiFi channel state information (CSI) has also been analyzed to extract 2D and 3D human poses [131, 178]. Most existing RF sensing systems incorporate a deep learning model with vision data supervised training. Furthermore, due to the relatively wide transmission range of the radio signals, such systems are susceptible to interference from the operating environment. Usually radar-based techniques are more resistant to environmental interference than WiFi-based schemes, but their customized hardware, e.g., the FMCW radar implemented on the Universal Software Radio Peripherals (USRP) platform, usually incurs a higher cost.

8.2.2 RFID-based Pose Estimation Systems

RFID tags can serve as low-cost and light-weight wearable sensors to attach to the human body, which provides a promising solution for human pose estimation. Several RFID sensing techniques have been developed in recent years, such as human vital sign monitoring [27, 88, 89, 139], mechanical vibration sensing [70], user authentication [68], material identification [69], and temperature sensing [52], Furthermore, RFID has also been utilized for indoor localization [21, 34, 57, 91] and drone navigation [53, 54, 90].

Using RFID tags as wearable sensors, such systems are usually more robust to interference from the operating environment than other RF sensing techniques (e.g., WiFi). This feature inspires the development of several RFID based human pose tracking systems as well. For example, RF-Wear [135] and RF-Kinect [134] were developed to track the movements of a single human limb, while RFID-Pose [157] and Cycle-Pose [180] were developed to track 3D

human poses in realtime. However, although the near-field RFID communications are more resilient to environmental interference, the locations of the tags and antennas still have a big impact on how the tags are sampled by the reader, and thus on the performance of the human pose tracking system.

In [192], the authors presented a domain adversarial technique to adapt to changes in the environment by utilizing a domain discriminator, which can constrain the unnecessary feature extraction from different environments. However, the proposed learning model may not be able to obtain the optimal variables when applied in a new RF environment, because all the training variables are determined by the datasets from a limited number of environments.

Inspired by the existing human pose tracking systems, we propose the Meta-Pose system in this paper, which is based on the meta-learning framework for greatly enhanced environmental adaptability. The proposed system incorporates a novel initialization algorithm to pretrain the deep learning model using a limited amount of training data, so that the system can be quickly fine-tuned with a small amount of new data when applied to a new environment, while still achieving a satisfactory performance.

8.3 Preliminaries of RFID-based Human Pose Tracking

The Meta-Pose system is proposed to estimate 3D human pose with RFID data collected from the passive RFID tags attached to the human subject. An overview of the Meta-Pose system is shown in Fig. 9.2. The Meta-Pose system comprises three key components, i.e., (i) RFID phase data collection, (ii) RFID phase preprocessing, and (iii) a deep neural network.

8.3.1 RFID Phase Data Collection and Preprocessing

In the RFID pose tracking system, the human pose is learned from RFID phase data, which is obtained by interrogating the tags attached to the human body using the RFID Low Level Reader Protocol (LLRP) [38]. The received RFID signal on a channel c can be written as [38]:

$$H = \sum_{m=1}^{M} \alpha_m e^{j\{\frac{2\pi 2R_m f_c}{v} + \Theta_c\}}, \tag{8.1}$$

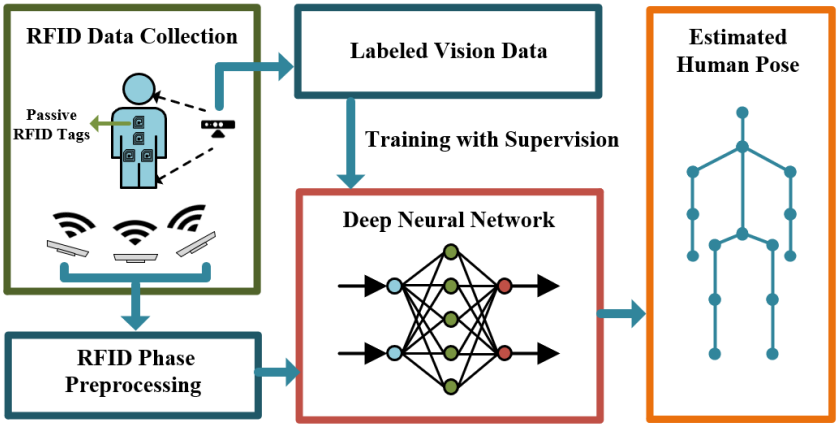


Figure 8.1: Overview of the proposed RFID pose tracking system.

where v represents the speed of light, M is the total number of RF signal propagation paths, α_m and R_m represent the signal strength and distance of each multipath component, respectively, f_c is the current channel frequency, and Θ_c is the initial phase offset caused by the circuit of both the antenna and the tag on channel c. Due to limitation of the Gen2 protocol used in the current commodity RFID systems, only one phase value of H could be directly sampled by the system. Due to the multipath effect, deriving the phase value of the line-of-sight (LOS) component from (8.1) is difficult. The reported phase value may not accurately depict the relationship between propagation distance and received phase. Fortunately, the polarized reader antenna operates as both the transmitter and receiver, and the interference caused by multipath reflections is not strong. Thus, we can assume that each propagation environment has at least one dominant path, and the received phase is given by:

$$\Theta = \frac{2\pi 2Rf_c}{v} + \Theta_c, \quad c = 1, 2, ..., 50, \tag{8.2}$$

where R is the distance of the dominant path between the reader antenna and tag. while the channel index c changes from 1 to 50 for every 200ms on each channel following the FCC regulation [38].

The LOS typically contributes significantly to the received signal for the following two reasons. First, in a passive RFID system, the only source of power utilized to send a response to the RFID reader is the tag antenna. The signal strength from reflection paths is typically

much weaker than that of the LOS path. Additionally, the RFID reader uses a power threshold for packet detection, which means that if there is no LOS path between the antenna and the tags, the interrogation is likely to fail. Second, the RFID phase data shall be prepossessed to mitigate the impact of the randomness in Θ_c on different channels. To this end, using the phase variation Φ between two adjacent samples from the same channel would be effective to cancel most of the randomness, which is given by:

$$\begin{aligned} \Phi(n) &= \Theta(n) - \Theta(n-1) \\ &= \frac{2\pi 2(R(n) - R(n-1))f_c}{v}, \ c = 1, 2, ..., 50, n > 1, \end{aligned} \tag{8.3}$$

where n is the sample index on each channel and $R(n)$ is the propagation distance corresponding to the nth sample on channel c. As (8.3) shows, the impact of the random channel hopping offset Θ_c (see (8.1)) is effectively canceled, except for the first sample on each channel (which is discarded). The phase variation only depends on the changes in the range of the dominant propagation path $\Delta R(n) = R(n) - R(n-1)$. Therefore, the sequence of phase variations $\{\Phi_2, \Phi_3, ..., \}$ can be translated into a sequence of antenna-tag distance variations $\{\Delta R_1, \Delta R_2, \Delta R_3, ..., \}$, which captures the realtime movements of the RFID tags. Consequently, with the RFID phase variations for the attached tags can be leveraged to reconstruct the human skeleton and track 3D human poses in realtime.

8.3.2 Multi-modal Deep Neural Network

Although phase variation can effectively capture the movements of the tags attached to human body, the translation from phase variation data to 3D human pose is still a challenge. In the few existing RFID based human pose tracking systems, the transformation is mostly accomplished using deep learning techniques [157, 180], which is mainly composed of a recurrent autoencoder and a forward kinematic layer. The brief structure of the deep learning model is presented in Fig. 8.2. As the figure shows, the network is designed to generate a sequence of 3D human poses, consisting of coordinate data of the RFID tags extracted from received RFID phase data. Specifically, the recurrent encoder is to extract both long-term and short-term

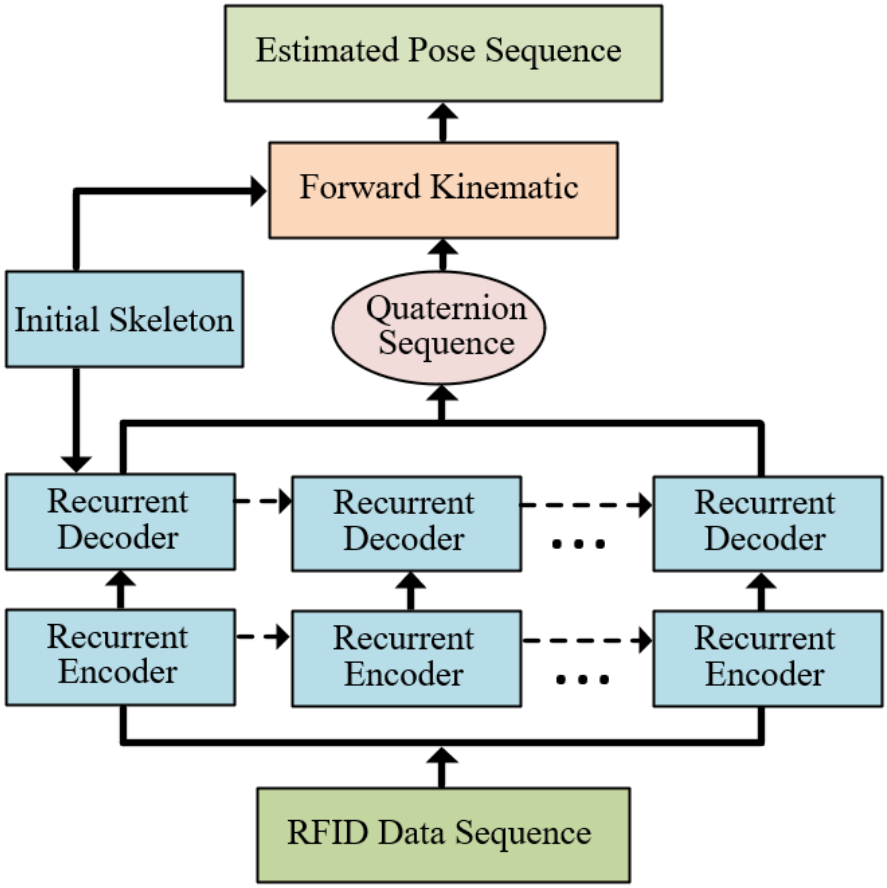


Figure 8.2: Structure of the deep learning model used in RFID based 3D human pose tracking.

features from the RFID phase data sequence, which are then fed into the following recurrent decoder. With a given initial skeleton, the decoder layer will transfer the features of the RFID data sequence to a quaternion sequence. Finally the Forward Kinematics module will construct the human pose sequence using the quaternion sequence, which is a widely used technique in robotics and 3D animation [193].

Rather than using RF signals to generate a confidence map for human skeleton reconstruction as in prior works [173, 178], our RFID-based human pose tracking system is designed to estimate human pose with the forward kinematic technique, which has been widely used in robotics and 3D animation [193]. This is because the information rate (or, the sampling rate) of the RFID system is too low to generate a useful confidence map with an acceptable resolution. However, the forward kinematic technique only requires the quaternions of the human skeleton joints, which indicates the 3D rotation angle of each human limb. Compared with AoA based localization techniques, the output human pose does not contain the global position of each

human joint, but the location relative to the root joint (pelvis). The ambiguity in AoA based localization techniques is not an issue because continuous human pose estimation mainly focuses on monitoring the relative movement of human limbs. As a tradeoff, the additional constraint is that the initial human skeleton should be the input to the system, which contains the length of each human limb, so that the the precise relative joint location can be estimated based on rotation angles.

As in RFID-Pose [157] and Cycle-Pose [180], vision data collected by a Kinect 2.0 camera is used as labels for supervised training of the deep learning model. The model is trained with a loss function that computes the difference between the estimated pose and the labeled vision data sampled simultaneously when the RFID data is collected, so the well-trained network can effectively transform RFID data sequence to a sequence of 3D human poses [157].

8.4 Challenges in Domain Adaptation

RF-based systems can better protect users' privacy and do not require sufficient lighting, compared to vision based approaches. However, they also bring about several unique challenges. Unlike vision data, the RF signal is usually sampled with unrelated noise from the system itself and the environment, which is hard to mitigate. The same test subject performing the same activity could generate very different RF data when being sampled in different environments, making it hard to reconstruct poses using a model well trained offline. To improve the adaptability of the system to different environments, generalization of the deep learning model is a big challenge needs to be addressed.

To analyze the influence of the environment, we use the term *data domain* to denote a specific wireless propagation environment in this paper. Since the tags are attached to the human body, a different data domains could be generated by the following ways. First, we fix the antennas and change the position of the subject. Second, we fix the subject position but change the antenna deployment. Finally, we could change the surrounding around the subject or the antenna. The wireless propagation environment depends on the characteristics of all the propagation paths, which could be significantly different in a different data domain.

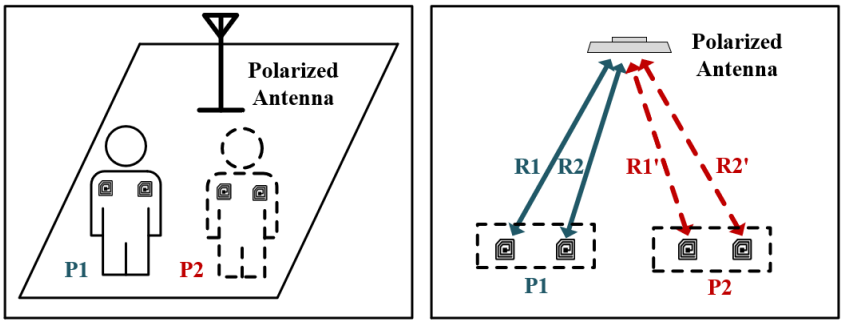


Figure 8.3: Illustration of data domains in RFID sensing systems.

However, the passive tags are merely powered by the incident signal from the reader, while the reader has a threshold for received power needed for a successful tag interrogation. Following FCC regulations, the effective radiated power of RFID should be less than 1watt. Therefore, we usually need to ensure that all tags are within 5 meters from the antenna. Compared to other long-range systems, e.g., WiFi and Radar, the RFID system is considered as a near-field system, and the environmental interference is relatively weaker. The data domain of RFID sensing systems is mainly determined by the relative position between the subject (i.e., the tags) and the reader antenna. As shown in the left plot in Fig. 8.3, when the subject stands at different positions, i.e., P_1 or P_2, the sampled phase data will come from two different data domains denoted by D_{p1} and D_{p2}, respectively.

The divergence of different data domains is mainly caused by: (i) the divergence in successful interrogation probability, and (ii) the distortion in RFID phase data, which are analyzed in the rest of this section.

8.4.1 Successful Interrogation Probability Divergence

The first cause of data divergence in different domains is the variation in Successful Interrogation Probability. When multiple tags are scanned by a reader, some tags are more likely to be detected, while some others may hardly be scanned. We define Successful Interrogation Probability as the probability for a tag to be successfully interrogated by the reader, which mainly depends on the received power strength from the tag.

Following the Friis transmission model, the received power S_r from a passive RFID tag can be written as [194]:

$$S_r = G_{An} G_{Tag} \gamma \left(\frac{\lambda_c}{4\pi R} \right)^4 S_t, \tag{8.4}$$

where S_t is the reader's transmit power; G_{An} and G_{Tag} are the gains of the transmit antenna and the tag, respectively; γ represents the aggregated attenuation coefficient, accounting for the losses incurred in the antenna cable and polarization, etc. during the transmission process; λ_c is the wavelength of the current channel c; and R is the LOS path range as in (8.2). Eq. (8.4) shows that with the same antenna and tag, the received power strength is degraded by an increased LOS path distance R and the attenuation loss γ, as:

$$S_r = K_c \gamma \left(\frac{1}{R} \right)^4 S_t, \tag{8.5}$$

where K_c is the product of all other coefficients other than γ and R, which takes different values in different tag and antenna deployment scenarios.

For example, see Fig. 8.3. When the subject is in the P_1 position, the LOS path distances for Tag 1 and Tag 2 satisfy $R_1 > R_2$. Because of the limited scanning range of the polarized antenna, the polarization loss γ of Tag 1 is also higher than that of Tag 2. Referring to (8.5), on the same channel c, the received power from Tag 1, denoted by S_r^{Tag1}, should be smaller than that from Tag 2, denoted by S_r^{Tag2}. However, when the subject is sampled in position P_2, we will have $S_r^{Tag1} > S_r^{Tag2}$. In RFID systems, tags with a higher S_r are more likely to be successfully interrogated than tags with a lower S_r, especially when multiple tags are scanned by a single antenna. Consequently, when multiple tags are attached to the human body, the sensitivity of the tags could be very different in different data domains.

The influence of Success Integration Probability divergence in different data domains could be considerable for RFID based pose tracking using an deep learning model. Since the tags with a higher sensitivity are more likely to be sampled, the training dataset will be mostly composed of the data from such tags. Thus, the training variables in the deep learning

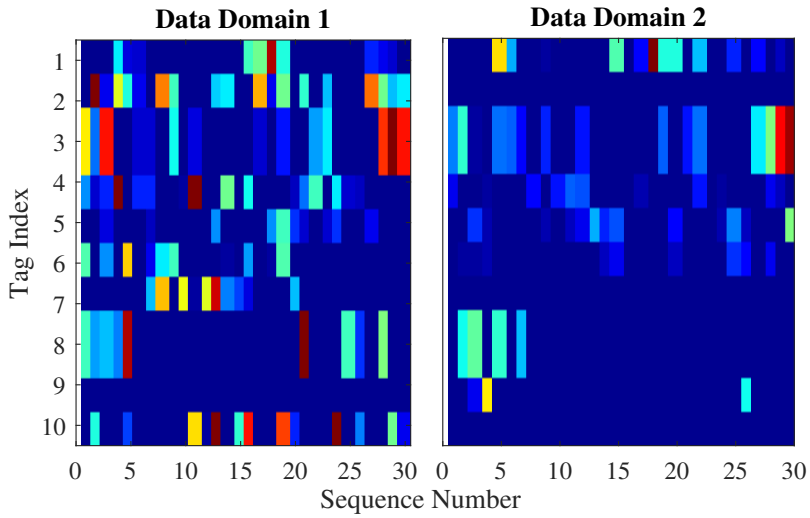


Figure 8.4: Phase distortion in RFID data collected in two different data domains.

model will be mostly trained by the data from the tags with higher sensitivity. When applying a trained deep learning model to a different data domain, the inference performance could be poor, since the Success Integration Probability in the new data domain could be very different from that where the model was originally trained.

8.4.2 Phase Distortion in Different Data Domains

The second cause of data domain divergence is the phase distortion caused by different antenna deployment scenarios. As (8.2) shows, the phase data of each tag is determined by the LOS propagation path distance R, which is the length of the space vector $\vec{R}$. For the tags attached to a moving human body, we can consider the overall space vector as the sum of two subspace vectors as: $\vec{R} = \vec{R}_s + \vec{R}_d$, where $\vec{R}_s$ is the *static vector* determined by the deployment scenario and $\vec{R}_d$ is the *dynamic vector* generated by the subject's movements.

Fig. 8.4 plots 30 sequentially received phase data from 10 RFID tags attached to the human body, where the phase value is represented by different colors. Two antennas are used to interrogate the tags simultaneously. We change the antenna deployment positions to create two different data domains. From the figure, we can observe considerable divergence in the collected phase values from the two data domains.

According to (8.2), the sampled phase Θ is affected by both $\vec{R}_s$ and $\vec{R}_d$ as:

$$\Theta = \frac{2\pi 2|\vec{R}_s + \vec{R}_d|f_c}{v} + \Theta_c,\ c = 1, 2, ..., 50. \tag{8.6}$$

Even if we have an identical $\vec{R}_d$ in the two data domains (i.e., the same subject and the same movement), the sampled phase could still be very different when the antennas are deployed differently (which leads to a different $\vec{R}_s$). Consequently, different antenna deployment scenarios will have an impact on the RFID phase distortion, causing considerable divergence between the datasets sampled from different environments.

Unlike Success Interrogation Probability divergence, a change in the operating environment usually causes considerable phase distortions in all sampled phase data. Thus, the model variables in the deep learning network should be trained and optimized to combat such phase distortion. Given all kinds of possible deployment environments, it is a big challenge to generate a well-optimized deep learning model, which is generalizable to all different environments.

8.5 Meta-learning based Solutions

8.5.1 Meta-learning for Domain Adaptation

The adaptation problem to new data domains or new tasks has been investigated in prior works. On one hand, researchers try to optimize the model variables, so that the network can achieve good adaptation in different data domains. The most straightforward approach is to train the model using datasets from more and more data domains. However, to achieve a good generalization performance, the training data should cover numerous data domains, incurring an overly high cost on obtaining labeled training data. To address this issue, the adversarial learning approach has been proposed to improve network adaptability by training using a limited number of data domains with the generative adversarial network (GAN) model [181, 217]. A domain discriminator is leveraged to constrain the loss function of the neural network, in order to combat the unrelated features from different data domains. The advantage of this approach is that the network does not need to be trained again when applied to a new data domain, but

the network variables are not well optimized when only considering the specific known data domains.

On the other hand, the network variables can be fine-tuned in the new data domain. Rather than addressing the data divergence issues in the well-trained network, this approach relies on additional training data for fine-tuning. The purpose is to let the network be further optimized in the specific new domain with a small amount of new training data sampled from the new domain. For typical pose tracking applications, e.g., video gaming or long-term pose monitoring, such light calibration is usually acceptable. Therefore, fine-tuning has been recognized as a promising way to improve generalization. With this approach, the network variables should first be well initialized in the pretraining stage, and then the fine-tuning process will be performed quickly with only a few additional data from the new data domain.

Meta-learning has been proved to be an effective technique for model pretraining so that a pretrained model can be quickly adapted for a new data domain [184]. In the case of the RFID based pose tracking, when data is sampled from an untrained domain, the performance of the previously trained model will usually degrade. New training data sampled from the new data domain is necessary to fine-tune the model for the new domain. The MAML algorithm is a representative meta-learning algorithm to pre-train the model for a satisfactory initialization before fine-tuning [185]. The Reptile algorithm [195] is another representative meta-learning algorithm for model pretraining, which has been shown to achieve a similar performance as MAML but at a lower computational complexity. We leverage these two algorithms in Meta-Pose to adapt the model to a new, unknown environment with few-shot fine-tuning using a few new training data.

In the Meta-Pose system, we implement both meta-learning algorithms for network initialization, and then fine-tune the pretrained network for a new data domain using only a few data examples.

These three key components are presented in the remainder of this section.

8.5.2 Meta-learning Framework with Domain Fusion

The objective of meta-learning is to determine the satisfactory initial model variables through network initialization, which can then be adapted to a new data domain with a few training examples. With appropriately trained initial variable x, the network loss for data domain D should be minimized after few steps of fine-tuning. Thus, the optimization problem for network initialization can be formulated as:

$$\min_{X} \quad \mathbb{E}_D[L(U_D^k(X))], \tag{8.7}$$

where $L(\cdot)$ denotes the loss function of the network, and $U_D^k(X)$ denotes the gradient descent operation that updates variables X for k times using the data sampled from D, which is the Adam algorithm.

Equation (8.7) shows that, the meta learning algorithm considers the gradient descent process as optimization target. Thus, rather than the normal training process based on the gradient of the loss function $\Delta L(X)$, meta-learning calculates the gradient of the gradient descent $\Delta L(U_D^k(X))$ in each training step. From the equation we can see that the performance of meta-learning can be determined by the amount of training data domain in D. However, it is highly costly to directly sample a large amount of human pose data from numerous data domains. Therefore we develop a domain fusion based meta-learning algorithm for model pretraining. The domain fusion algorithm randomly select samples from the four known domains to form new domains, in order to increase the number of known domains for pretraining.

Figure 8.5 represents the brief structure of the training procedure of the proposed Meta-Pose system, which consists of network initialization and fine-tuning in a new domain. As shown in the figure, the deep learning model is first pretrained using datasets from a few (e.g., four) known data domains, which are sampled when the subject is standing at four different positions. The the network is pretrained with two different meta-learning algorithms. Since the second-order gradient $\Delta L(U_D^k(X))$ is hard to calculate in practice, we leverage the first-order approximation instead to update the training variables. Based on the divergence in the first-order approximation, we develop two different initialization approaches based on Reptile and

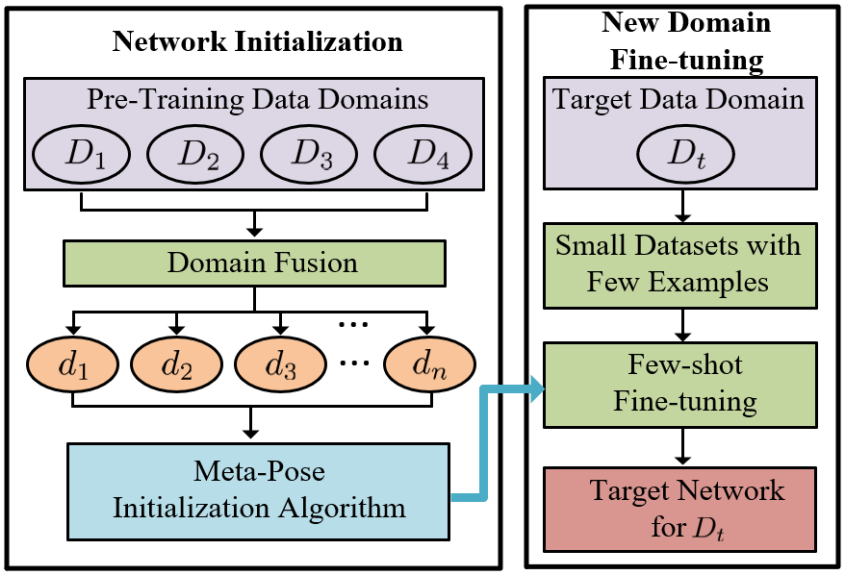


Figure 8.5: Training framework of the proposed Meta-Pose system.

Algorithm 2: Reptile based Initialization Algorithm

Input: Sampled data sets from the four known data domains (denoted by D_1, D_2, D_3, and D_4);
Output: Optimally initialized variables X_t for the pretrained network.
Randomly initialize the training variable as X;
for $i = 1 : n$ **do**
    Generate d_i by randomly sampling from D_1, D_2, D_3, and D_4;
    Randomly sample k batches from d_i;
    Set the inner loop training variables: $X_{in} \leftarrow X$;
    for $j = 1 : k$ **do**
        Update the variables in X_{in} with loss function L as: $X'_{in} = U^1_{d_i}(X_{in})$, $X_j = X'_{in} - X_{in}$, $X_{in} \leftarrow X'_{in}$;
    end
    Calculate the outer loop gradient as: $\hat{W}_i = \sum_{j=1}^{k} W_j$;
    Update the outer loop variables X as: $X \leftarrow X + \epsilon \hat{W}_i$;
end
Set $X_t \leftarrow X$;

MAML algorithms, respectively. When transferring the learning task to a target data domain D_t, we only need to collect very few examples in the target domain to fine-tune the generalized network.

8.5.3 Reptile-based Network Initialization

In the Reptile-based algorithm, we first fuse the four data domains (i.e., D_1, D_2, D_3, and D_4) into a larger number of fused data domains (i.e., $d_1, d_2, ..., d_n$). Specifically, each d_i contains

40 batches of data randomly sampled from D_1, D_2, D_3, and D_4. To solve the optimization problem (8.7), we need to find the gradient of any fused data domain $\Delta L[U_{d_i}^k(X)]$, so the gradient descent algorithm can be applied to find X by recursive updating. With the Reptile learning algorithm [195], we first calculate $\Delta L[U_{d_i}^1(X)]$ for each iteration in the inner loop as:

$$\begin{aligned} \Delta L[U_{d_i}^1(X_{in})] &= U_{d_i}^1(X_{in}) - X_{in} \\ &= X'_{in} - X_{in}, \end{aligned} \tag{8.8}$$

where X_{in} is the set of variables used in the inner loop. In the algorithm, denote the one-step gradient $\Delta L[U_{d_i}^1(X_{in})]$ as W_j. The overall gradient after k iterations is calculated as:

$$\Delta L[U_{d_i}^k(X)] = \sum_{j=1}^{m} W_j. \tag{8.9}$$

$\Delta L[U_{d_i}^k(X)]$ is denoted as $\hat{W}_i$ for each data domain d_i. In the algorithm, we set $k = 8$ for effective training in each data domain. With gradient $\hat{W}_i$, we solve problem (8.7) by recursively training variable X in the outer loop iterations as:

$$X \leftarrow X + \epsilon \hat{W}_i, \tag{8.10}$$

where ϵ is the learning rate, which is set to 0.1 in the system. We repeat the updating process for $5{,}000$ times (i.e., setting $n = 5{,}000$), so the final training result X_t could satisfy the initialization requirement of problem (8.7).

8.5.4 MAML-based Network Initialization

With MAML based initialization, we leverage the same method to generate fused data domains d_i for each iteration in the outer loop updates. Unlike the Reptile algorithm, $\Delta L(U_D^k(X))$ is approximated by two-step training [185]. Thus, for each iteration, we firstly sample two batches of data B_1 and B_2 from the fused data domain d_i and update the variables X_i with one-step gradient descent using batch data B_1 to obtain X'_{in}. In the MAML-based learning

Algorithm 3: MAML based Initialization Algorithm

Input: Sampled data sets from the four data domains (denoted by D_1, D_2, D_3, and D_4);
Output: Optimally initialized variables X_t for the pretrained network.
Randomly initialize the training variables as X;
for $i = 1 : n$ **do**
    Generate d_i by randomly sampling from D_1, D_2, D_3, and D_4;
    Randomly sample 2 batches B_1 and B_2 from d_i;
    Set the inner loop training variables: $X_{in} \leftarrow X$;
    Update the variables in X_{in} with loss function L and dataset B_1 as: $X'_{in} = U^1_{d_i}(X_{in})$;
    Update the variables in X'_{in} with loss function L and dataset B_2 as: $X''_{in} = U^1_{d_i}(X'_{in})$;
    Calculate the outloop gradient as: $\hat{W}_i = X''_{in} - X'_{in}$;
    Update variables X as: $X \leftarrow X + \epsilon \hat{W}_i$;
end
Set $X_t \leftarrow X$;

algorithm, we set $k = 1$ to reduce the training complexity. So the outer loop gradient can be approximated by:

$$\begin{aligned} \Delta L[U^1_{d_i}(X)] &= \Delta L[X'_{in}] \\ &= U^1_{d_i}(X'_{in}) - X'_{in}. \end{aligned} \tag{8.11}$$

We next update X'_{in} by one more step of gradient descent using another batch data B_2 and generate $X''_{in} = U^1_{d_i}(X'_{in})$. Accordingly, the outer loop gradient is estimated as the gradient of the second step training, which is calculated by $X''_{in} - X'_{in}$. With the outer loop gradient found by the MAML based algorithm, the training variables X is initialized by recursively updating it in the outer loop following (8.10), where the learning rate ϵ is also set to 0.1. The update is also iterated for $5{,}000$ times, so a large number of fake data domains will be used in model pretraining. After initialization training, the network will be able to be quickly fine-tuned using a few shots of data sampled from a new data domain.

8.5.5 Few-shot Fine-tuning

After an appropriate initialization of X, the fine-tuning process only requires a very small dataset from the new data domain. Since the training data are all in the form of data sequences, including RFID phase data and Kinect vision data [180], the data shots are defined specifically

in the Meta-Pose system. We divide the data sequence into small segments during the training process, each consisting of 30 consecutive data samples sampled within a window of 6s. We consider one such data batch as a shot in Meta-Pose, and less than 5 shots of data from the new data domain will be leveraged for fine-tuning. We also find that the type of activities also affects the fine-tuning performance and will discusses this further in Section 8.6.

8.6 Implementation and Evaluation

8.6.1 System Implementation

To evaluate the performance of Meta-Pose, we develop a prototype system using an off-the-shelf Impinj R420 reader, which is equipped with three S9028PCR polarized antennas, as shown in Fig. 8.6. ALN-9634 (HIGG-3) RFID tags are used in Meta-Pose operating in the Ultra High Frequency (UHF) band. The vision data, used for training supervision as well as ground truth for evaluating the precision of inference, is collected using an Xbox Kinect 2.0 device. As shown in the figure, we attach 12 RFID tags to the 12 joints of the subject, including neck, pelvis, left hip, left knee, right hip, right knee, left shoulder, left elbow, left wrist, right shoulder, right elbow, and right wrist. With the three reader antennas placed at different positions with different heights, every RFID tag can be interrogated by at least one of the antennas.

Environment adaption is validated using RFID data collected from eight different data domains, which are generated by specific deployments of the subject and antennas as illustrated in Fig. 8.7. Seven data domains are sampled in a computer lab, and the eighth data domain is sampled in an empty corridor. Each domain is a 0.6×0.6 m^2 square area, where the subject shall stand inside performing certain activity during data collection. With the 900 MHz frequency and 0.33 m wavelength used in the proposed RFID system, a 0.6 m interval is sufficient to generate considerable divergence to create different data domains.

Among these domains, D_1 to D_4 are used for model pretraining, where 70% of the sampled data from each domain is used for training, and the rest 30% is used for testing. D_5 to D_8 are

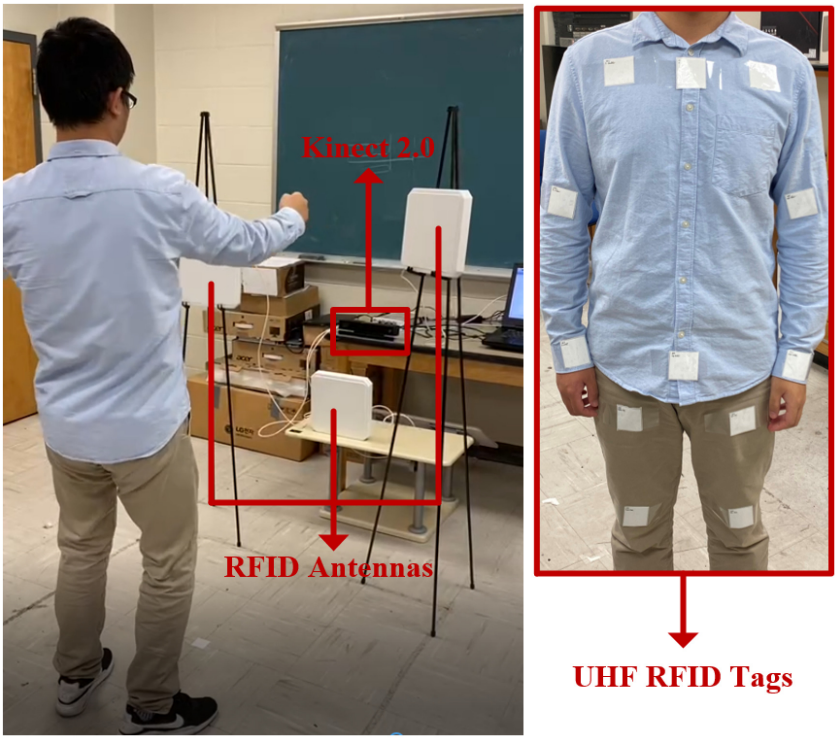


Figure 8.6: Experiment configuration of the Meta-Pose system.

considered as new data domains for evaluating the generalization performance, where 50% of the data from each of these domains is used for fine-tuning, and the rest 50% is used for testing.

RFID phase data is collected when the subject stands in front of the antennas and repeatedly performs specific activities. Different types of activities are sampled in all the data domains, such as walking, body twisting, deep squatting, and moving a single limb. Five subjects participate in the experiments for sufficient data diversity, including one female and four males. The sampling rate of the antenna is 110 Hz. However, due to the collision avoidance protocol, when the reader is interrogating multiple tags, only one randomly chosen tag could respond to the reader at a time. The sampling rate for each tag of a multi-tag system is not even nor constant, depending on the relative location of each tag, interference, and the mutual coupling effect [57]. To deal with the low sampling rate and sparse RFID data, we firstly construct a tensor with the sampled raw data, and then leverage tensor completion to interpolate the missing data. Finally, the calibrated data fed to the neural network is downsampled to 5 Hz for processing in realtime.

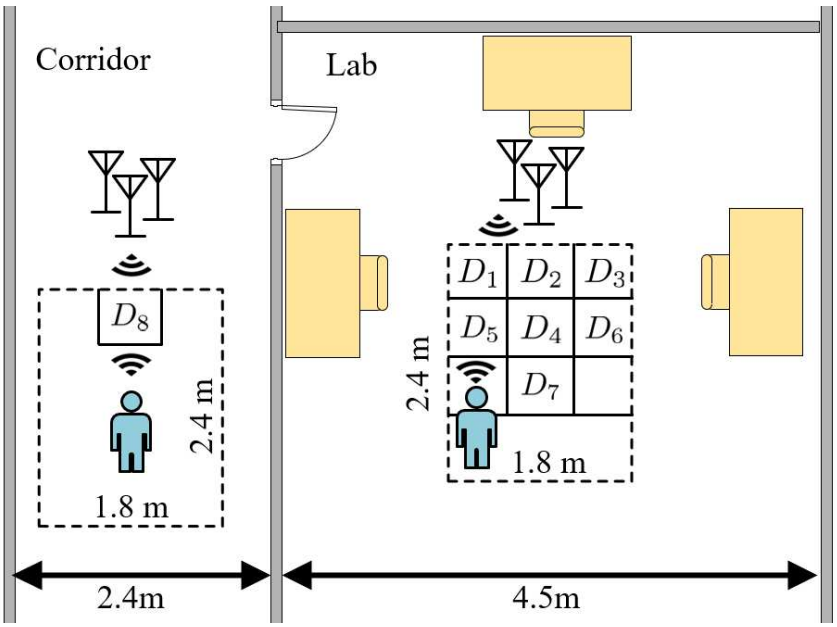


Figure 8.7: Illustration of the data domains used in the Meta-Pose experiments.

Table 8.1: Performance Evaluation for Different Subjects

Subject Index	*Estimation Error*
Subject 1	3.62cm $\pm$ 0.24cm
Subject 2	4.35cm $\pm$ 0.34cm
Subject 3	3.78cm $\pm$ 0.22cm
Subject 4	5.12cm $\pm$ 0.37cm
Subject 5	4.17cm $\pm$ 0.31cm

8.6.2 Overall Performance Evaluation

To demonstrate the overall system performance, we use the 3D human pose data collected by Kinect 2.0 as ground truth. For each video frame, we calculate the mean error Ψ_{all} of all the 12 joints as:

$$\Psi_{all} = \frac{1}{12}\sum_{n=1}^{12} ||\hat{T}_n - \dot{T}_n||, \tag{8.12}$$

where $\hat{T}_n$ represents the estimated 3D position of joint n, $\dot{T}_n$ is the ground truth, and $||\hat{T}_n - \dot{T}_n||$ is the Euclidean distance between the two 3D positions.

The overall performance (i.e., mean error) of the fine-tuned network for all the eight data domains is presented in Fig. 8.8. Recall that only the first four data domains are used for model

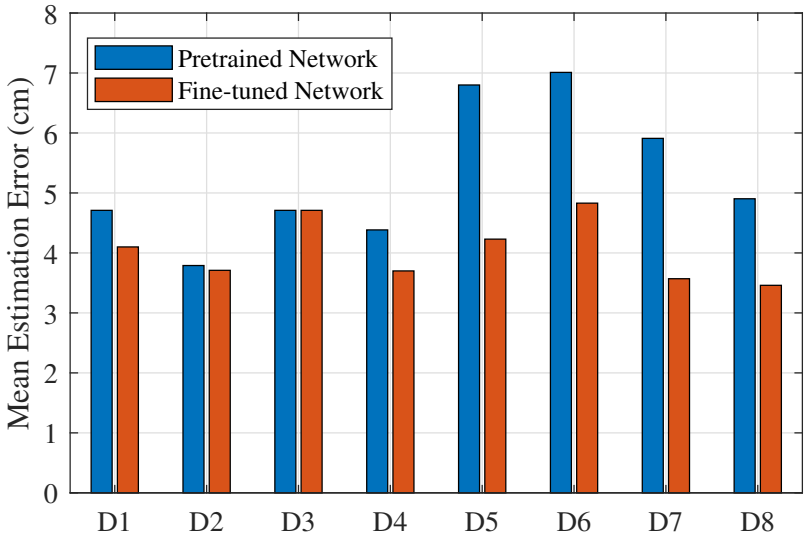


Figure 8.8: Overall performance in terms of mean estimation error in the eight different data domains.

pretraining, while the other four domains are used for testing. In addition, we also present the accuracy of the pretrained network in the figure (i.e., without fine-tuning using additional data from the new data domain). As shown in the figure, the maximum error of the fine-tuned network is 4.83 cm achieved in D_6, while the minimum error is 3.46 cm achieved in D_8. The minimum pretraining error for the new data domain (i.e., D_5 to D_8) is 4.91 cm in D_8, which is higher than that of all the pretrained domains (i.e., D_1 to D_4). The higher pretained errors imply the large divergence between the known and new data domains. However, with few-shot fine-tuning, the mean error for all the four new data domains is reduced to 3.98cm, which is very similar to that of the known data domains. The considerable error reduction in D_5, D_6, D_7, and D_8 is due to the Meta-Pose initialization algorithm. With the well optimized training variables, the deep learning model can be effectively fine-tuned for new data domains. Compared to the height of the subject and range of motions, the 3D human pose estimation errors are all small and negligible. These results demonstrate the high adaptability of the Meta-Pose system.

8.6.3 Fine-tuning for the Two Pretrain Algorithms

For most effective fine-tuning, we conduct experiments to investigate the impact of the number of shots and the type of activities on different initialization algorithms. Fig. 8.9 illustrates the accuracy of human pose tracking in the four new data domains with Reptile initialization, which

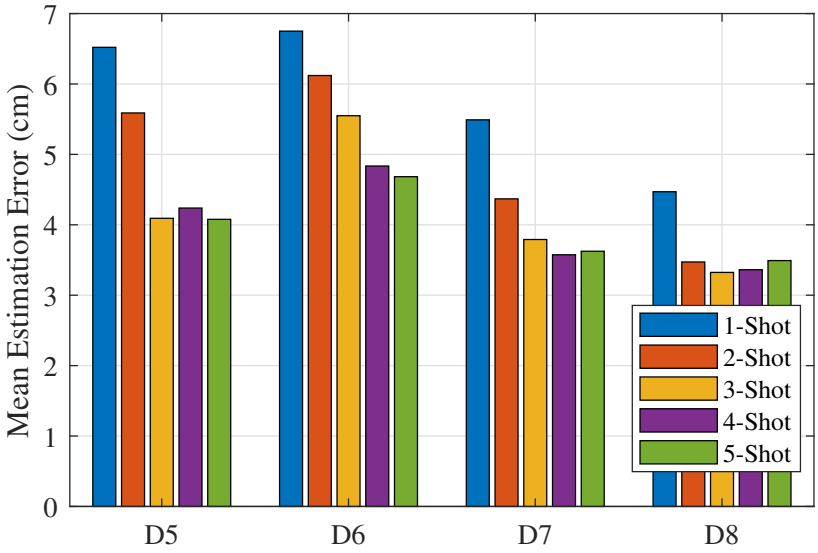


Figure 8.9: Fine-tuning performance of Reptile based initialization using different shots of new data.

are fine-tuned with different numbers of data shots ranging from 1 to 5. Fig. 8.10 shows similar fine-tuning results but with MAML based initialization. As defined earlier, one-shot of data in Meta-Pose is defined as a consecutive data sequence within a time window of 6 s. It can be seen that, after 5-shot fine-tuning after Reptile initialization, the minimum error 3.49 cm is achieved in D_8, while the error in D_6 is the highest (i.e., 4.68 cm). For MAML based initialization, the minimum error 3.53 cm is achieved in D_7, and the max error is 4.34cm achieved in D_6. From the performance of different data domains shown in Figs. 8.9 and 8.10, it can be seen that both Reptile and MAML are able to compute satisfactory initial learning variables. Both models can be adapted to different new data domains within five shots of fine-tuning.

In addition, although the final estimation accuracy is different in the four data domains, the performance of fine-tuning is generally improved as more data shots are used. However, as the figure shows, the improvement becomes not obvious beyond four shots of data for both algorithms. Thus, four-shot fine-tuning will be sufficient when the Meta-Pose system is transferred to a new environment.

We also examine the impact of different types of activities based on the accuracy of tracking different types activities. In Fig. 8.11, we present the n-shot fine-tuning results of Reptile based initialization in the specific data domain D_5 with different types of activities, including

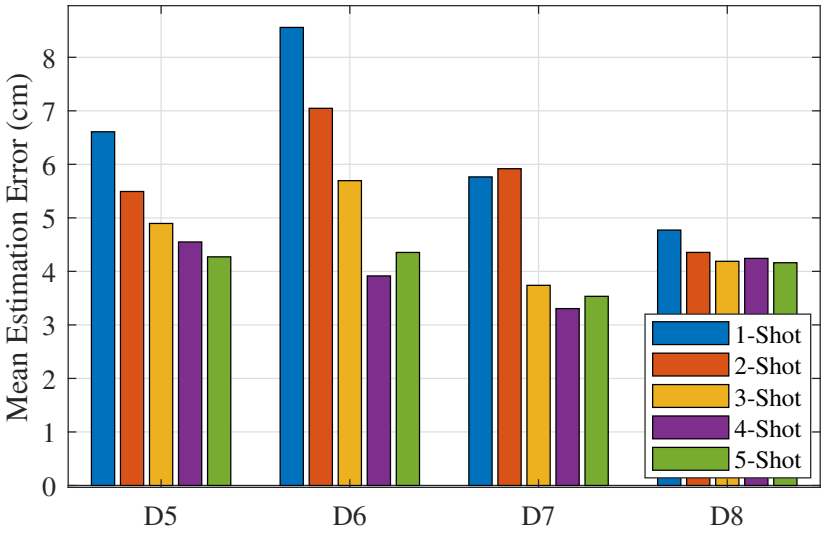


Figure 8.10: Fine-tuning performance of MAML based initialization using different shots of new data.

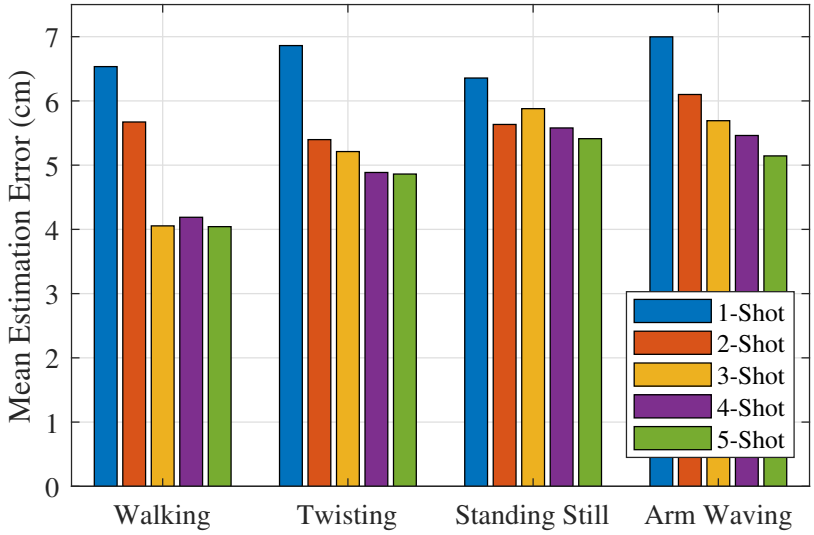


Figure 8.11: Fine-tuning performance of Reptile based initialization for different activities in new data domain D_5.

walking, body twisting, standing still, and arm waving. we also provide the n-shot fine-tuning result of MAML initialization in Fig. 8.12 for comparison purpose.

Figure 8.11 shows that, after 5-shot fine-tuning, the minimum error 4.04 cm of Reptile based initialization is achieved when the system is fine-tuned for the walking activity, while standing has the maximum error of 5.41 cm. For MAML based initialization, the the minimum error 3.94 cm is also achieved by the walking activity. We also find that fine-tuning is not as effective for arm-waving and standing for both initialization algorithms. This is because simple

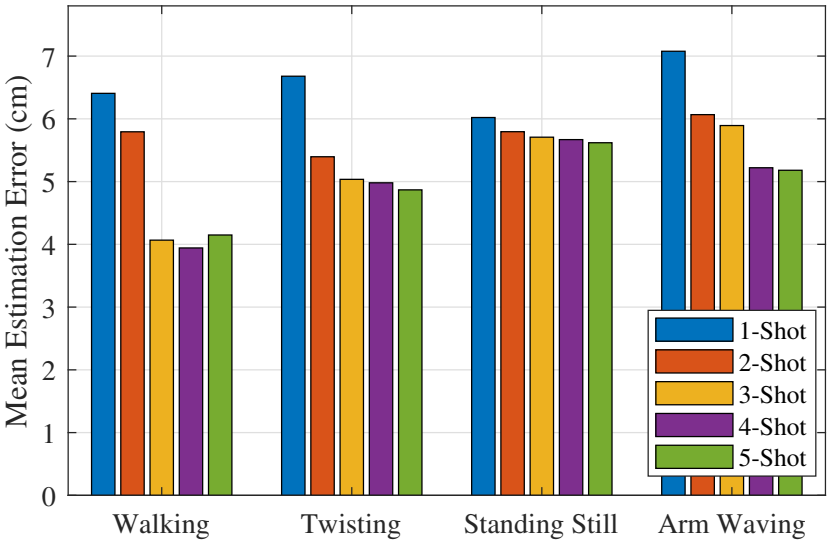


Figure 8.12: Fine-tuning performance of MAML based initialization for different activities in new data domain D_5.

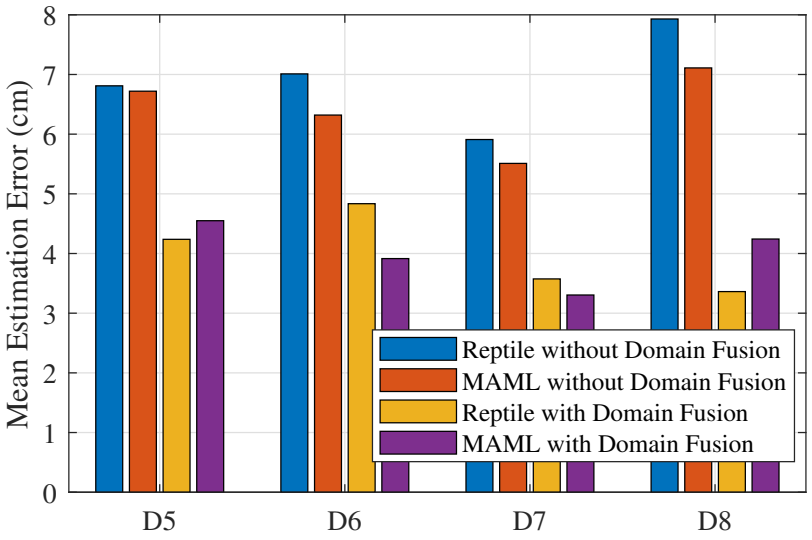


Figure 8.13: Fine-tuning performance of the domain fusion algorithm and typical meta-learning algorithm.

activities, such as standing and arm waving, contain less information than the more complicated activities, such as walking. Generally, fine-tuning will be more effective when more information is carried in the new data shots. Thus, we conclude that fine-tuning is more effective for more complicated activities, no matter which algorithm is used for network pretraining.

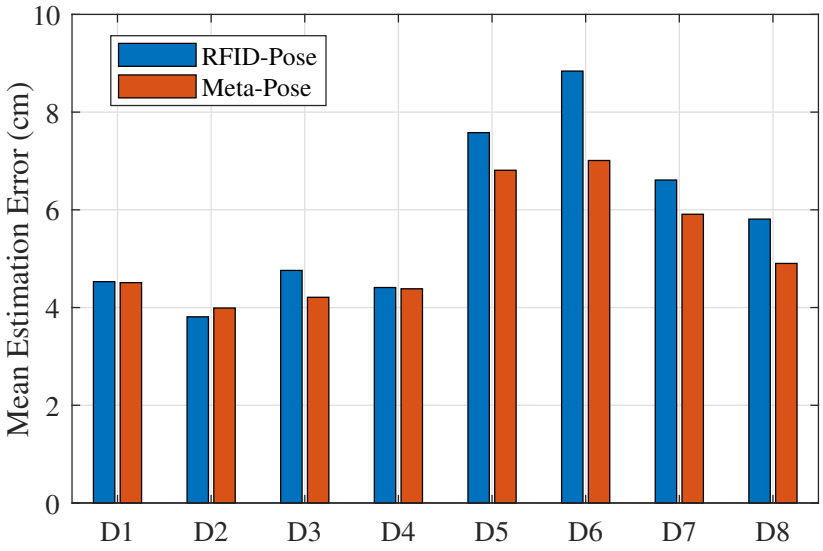


Figure 8.14: Pretraining comparison with the baseline method RFID-Pose [157] without fine-tuning.

8.6.4 Effect of the Domain Fusion Algorithm

The superiority of the domain fusion algorithm used in meta-learning based pretraining is demonstrated by the next experiment. As shown in Fig. 8.5, we randomly sample training data from the four known data domains to generate more virtual data domains, i.e., the d_i's, to enhance the performance of the meta-learning algorithms. Fig. 8.13 illustrates the fine-tuning performance of the domain fusion algorithm and the two representative meta-learning algorithms. The figure presents the four-shot fine-tuning results with different initialization algorithms for all the four untrained data domains. As the figure shows, without the domain fusion algorithm, the minimum estimation error is 5.51 cm, and the maximum estimate error is 7.93 cm, which are quite high for 3D human pose tracking. In contrast, with the domain fusion algorithm, the minimum error is reduced to 3.36 cm while the maximum error is only 4.83 cm now. Thus, the greatly reduced errors prove that, the domain fusion algorithm could effectively enhances the model pretraining and reduce the cost of obtaining training data.

8.6.5 Comparison with a Baseline Scheme

Finally, we conduct a comparison study using our recent RFID based pose tracking system RFID-Pose as a baseline scheme [157]. As in Meta-Pose, we leverage the same training dataset

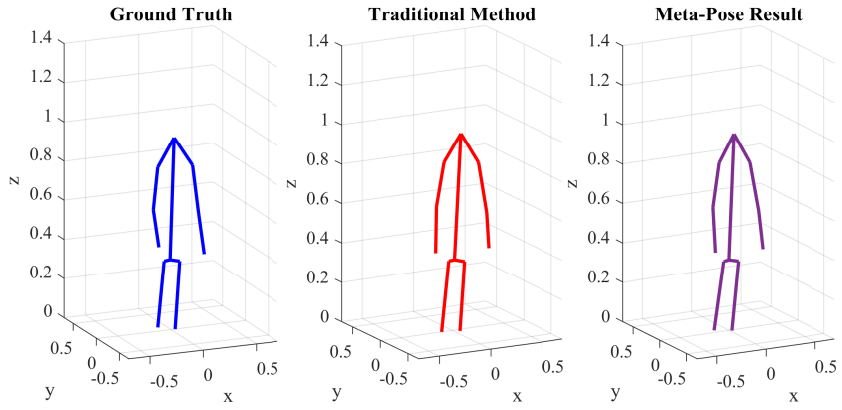


Figure 8.15: Comparison results for a pretrained data domain D_4.

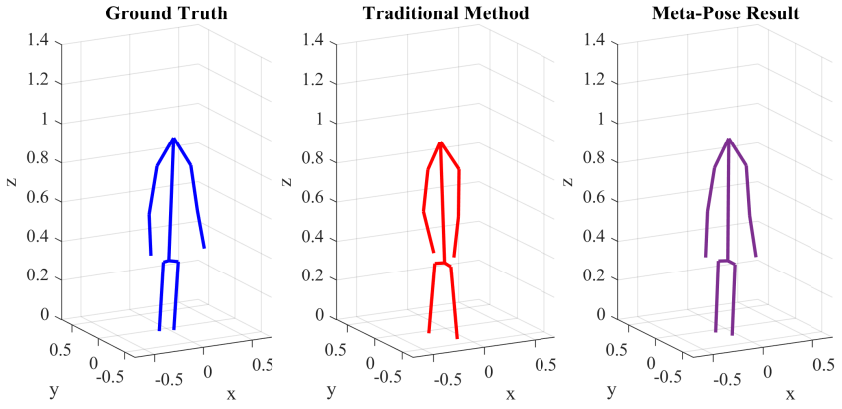


Figure 8.16: Comparison results after four-shot fine-tuning for a new data domain D_5.

collected from D_1 to D_4 to pretrain the RFID-Pose model. The estimation error for all the domains are presented in Fig. 8.14 without fine-tuning. The figure validates that both systems achieve a good, comparable performance for the known domains (i.e., D_1 to D_4). However, RFID-Pose has relative larger errors when applied to the four unknown domains (i.e., D_5 to D_8). The maximum error of RFID-Pose is 8.84 cm and the mean error of all the new data domains is 7.21 cm. In contrast, the mean error of Meta-Pose for the unknown domains is 6.12 cm without fine-tuning. These results indicate that the Meta-Pose initialization algorithm finds better initial model variables for the new data domains than RFID-Pose.

The superiority of the Meta-Pose initialization algorithm is further demonstrated with the fine-tuned results. Fig. 8.15 illustrates examples of estimated poses obtained by Meta-Pose and RFID-Pose for a pretrained data domain D_4. The ground truth shown on the left is generated by Kinect. The middle and the right poses are estimated by RFID-Pose and Meta-Pose after

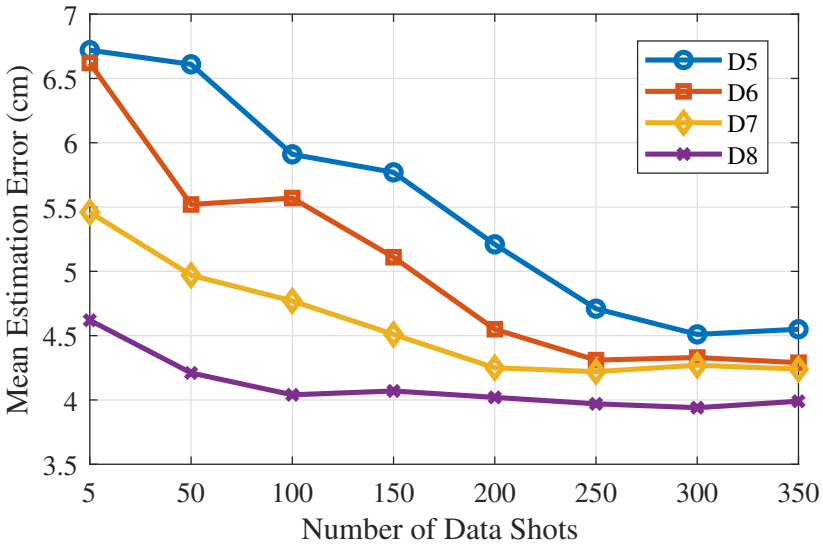


Figure 8.17: Fine-tuning performance of the baseline method RFID-Pose in different data domains.

pretraining, respectively. Fig. 8.16 depicts the estimated poses for an unknown data domain D_5 following a four-shot fine-tuning. As demonstrated in these two examples, for a pretrained data domain, the predicted human poses by RFID-Pose and Meta-Pose are both similar to the ground truth. However, for the new data domain, the Meta-Pose generated pose is still close to the ground truth, while the traditional method generated pose looks obviously different from the ground truth.

Fig. 8.17 illustrates the performance of RFID-Pose for the untrained data domains ($D_5 \sim D_8$), while different numbers of data shots are used for fine-tuning. The figure shows that the traditional system requires considerably more data for adaptation to new environments. For example, at least 300 data shots are needed for fine-tuning when adapting RFID-Pose to new untrained data domain D_5, while D_6 and D_7 require 250 and 200 data shots, respectively. However, as illustrated in Fig. 8.9 and Fig. 8.10, four data shots are sufficient for Meta-Pose. As defined before, one-shot data consists of 6s of consecutive data samples, and so four data shots mean the system needs to collect 24 seconds of training data for a new data domain. However, with 200 training data shots for D_7, the traditional system requires at least 20 minutes of new training data for domain adaptation, while D_5 and D_6 need 30 and 25 minutes of new training data, respectively. The large difference in the amount of training data show that Meta-Pose effectively reduces the expense of new environmental adaption.

Table 8.2: Performance Comparison after Fine-tuning

Domain	*RFID-Pose*	*Meta-Pose (Reptile)*	*Meta-Pose (MAML)*
D_5	6.72cm	4.23cm	4.55cm
D_6	7.62cm	4.83cm	3.91cm
D_7	5.46cm	3.57cm	3.30cm
D_8	4.62cm	3.36cm	4.24fcm
D_{all}	6.27cm	3.97cm	4.03cm

The Cumulative distribution functions (CDF) of the estimation errors of the two systems are plotted in Fig. 8.18. The figure presents the estimation error after four-shot fine-tuning for all untrained data domains. The figure shows that the median estimation error of RFID-Pose is 3.94cm, whereas the median error of Meta-Pose is 6.87cm. Furthermore, we observe that the overall estimation error of RFID-Pose is considerably higher than the Meta-Pose system.

Table 8.2 presents the mean estimation error for each untrained data domain. As the table shows, the mean error of RFID-Pose for all the new data domains is 6.27 cm, while the mean errors of Meta-Pose with Reptile and MAML based pretraining are 3.97 cm and 4.03 cm, respectively. We find that the RFID-Pose error is also reduced by fine-tuning, but its estimation error for new data domains is still quite high.

The experiments show that larger datasets sampled in the new environments are needed for RFID-Pose to achieve a satisfactory fine-tuning performance, which considerably increases the training data collection effort and cost. In contrast, the error of Meta-Pose can be effectively reduced by few-shot fine-tuning, because the meta-learning-based algorithms have suitably initialized the model variables based on the known data domains. Meta-Pose is able to quickly optimize its training variables for untrained data domains with a few data examples. Through these experiments, we demonstrate that Meta-Pose can better adapt to unknown environments compared with the baseline scheme. Thus it can be easily deployed in practice in different application environments.

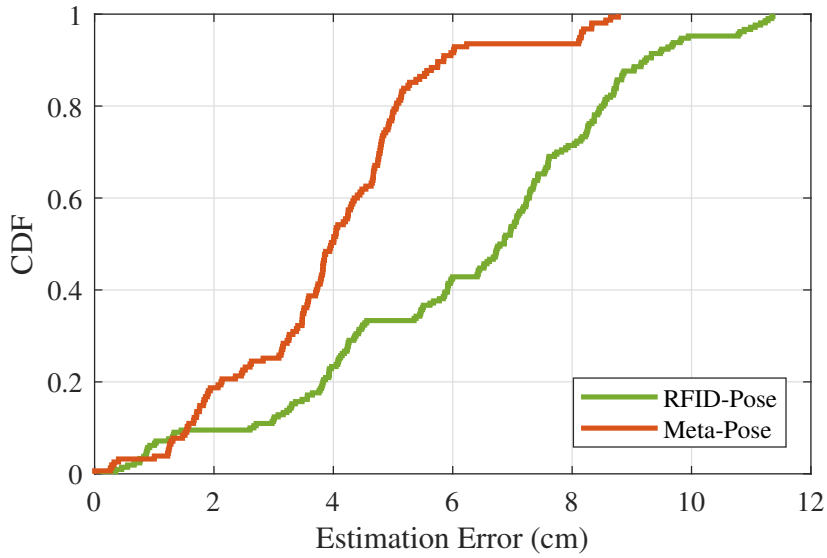


Figure 8.18: The CDF curves of the four-shot fine-tuning results of RFID-Pose and Meta-Pose.

8.7 Conclusions

In this paper, we proposed an RFID based realtime 3D pose tracking system, termed Meta-Pose, that is environment-adaptive. A novel Meta-Pose initialization algorithm was proposed to pre-train the network with several known data domains, and few-shot fine-tuning was then utilized to adapt to unknown data domains. The Meta-Pose system was developed with two different meta-learning algorithms, i.e., Reptile and MAML. The Meta-Pose system was implemented using off-the-shelf RFID reader and tags. Extensive experiments were conducted with ground truth provided by Kinect 2.0 vision data. Meta-Pose's high accuracy and adaptability to new environments were demonstrated by our experimental results and a comparison study with a state-of-the-art baseline scheme.

Chapter 9

TARF: Technology-agnostic RF Sensing for Human Activity Recognition

9.1 Introduction

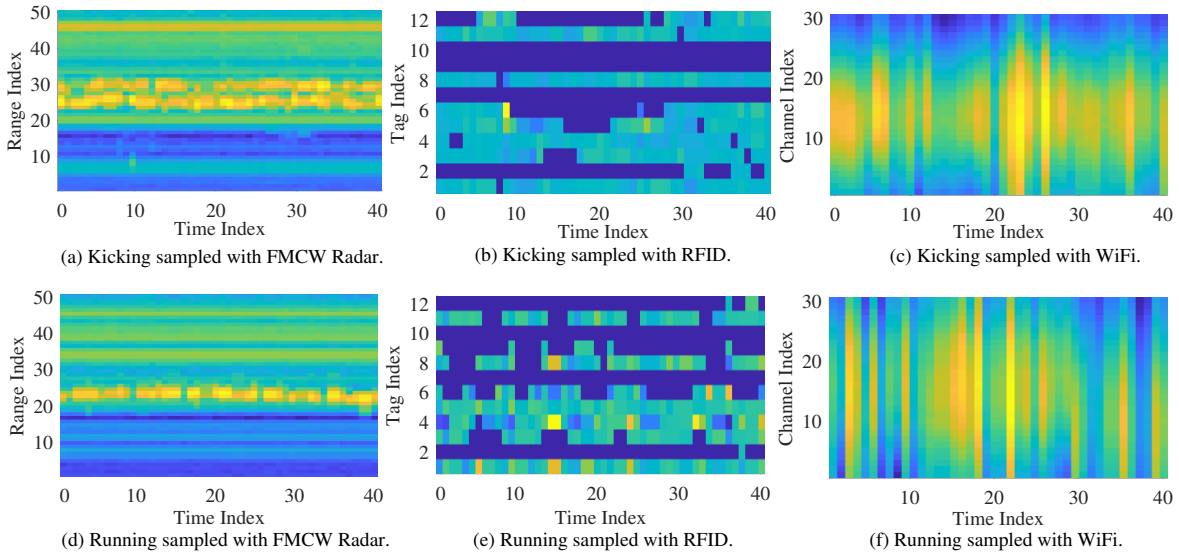
Human activity recognition (HAR) has been recognize as one of the most important technology for many Internet-of-Things (IoT) applications, such as smart homes, safety surveillance, and health-care monitoring [196]. Video cameras and wearable sensors, such as smart watches and gyroscopes embedded in smartphones, are mostly used in traditional HAR solutions [197]. However, vision based HAR is usually constrained by the lighting condition and interference from the background, and may raise privacy concerns, while wearable sensors are uncomfortable for prolonged usage. As a result, several RF-based HAR solutions, such as WiFi [19, 198], Radio Frequency Identification (RFID) [192], and various types of radars [199], have been developed to overcome such constraints. By incorporating deep learning algorithms, these RF-based HAR approaches were shown effective to distinguish various types of human activities.

However, the existing solutions are each closely designed and tailored for a specific RF technology or platform. In the growing trend towards smart IoT systems, various RF sensing technologies are emerging. The limitation of being tied up with a specific technology or platform will hinder the development of large-scale, and easy-to-deploy HAR systems. It will be highly desirable to develop a HAR solution that can work with different types of RF technologies. First, such a *technology-agnostic solution* will greatly reduce the cost and overcome the barrier of wide deployment of HAR systems. For example, an existing RFID-based solution, e.g., [192], does not work with WiFi or radar. However, a user that does not have access

to RFID can still make use of a technology-agnostic system with whatever RF sensing platform that is available, e.g., WiFi, without needing to acquire an RFID system. In addition, a technology-agnostic system can be used both in the lab, where both radar and RFID are available, and in the home, where there is only WiFi. Such a technology-agnostic solution will be of great value to users in such scenarios. Second, due to the different frequency band, wireless communication protocols, and hardware design, various RF platforms have their unique strengths and weaknesses in specific deployment environments. WiFi-based techniques, for example, can cover a large area, but are also susceptible to interference from the surrounding environment. RFID, on the other hand, is more resistant to interference from the environment, but is restricted by the shorter interrogation range, and the collisions and the mutual coupling effect induced by crowded tags. A technology-agnostic approach that utilizes the diverse, and sometimes complementary RF sensing technologies will lead to generalized and more robust HAR systems.

Obviously, this is a highly challenging problem given the variety of frequency bands, protocols, and hardware used in different RF sensing systems. The same propagation environment will become very different wireless channels and the same human activity will be transformed into very diverse RF representations. For example, Fig. 9.1 presents the RF data captured by Frequency-Modulated Continuous Wave (FMCW) radar, RFID, and WiFi for the same kicking and running activities, respectively. The same human activities are translated into very different RF data, due to the different channel frequency, Physical layer (PHY) protocols, and hardware used in these three RF sensing technologies. Not only the metrics used to describe such measurements are very different, but also the characteristics as indicated by the measurements for the same human activity exhibit a high degree of diversity. It is a great challenge and open problem to develop a technology-agnostic system to detect the original human activity from such diverse representations.

In this paper, we propose a novel generalized, **t**echnology-**a**gnostic **RF** sensing system, termed TARF, for flexible and accurate human activity recognition utilizing a wide range of



(a) Kicking sampled with FMCW Radar. (b) Kicking sampled with RFID. (c) Kicking sampled with WiFi.

(d) Running sampled with FMCW Radar. (e) Running sampled with RFID. (f) Running sampled with WiFi.

Figure 9.1: Raw RF signals sampled by different RF devices.

different RF sensing technologies. We first investigate the causes of the barriers between various RF technologies and find that the diversity is mostly caused by three factors: metric diversity, measurement sensitivity, and distinct translation of human activity to RF features. To address these problems, we first calibrate the RF data from different RF sensing technologies to represent them in a unified format. We then propose a signal preprocessing module that uses the Short Time Fourier Transform (STFT) to generate a generalized RF feature tensor, which can limit the interference of metric diversity and sensitivity diversity of different RF sensing technologies. In addition, we propose a Domain Adversarial Neural Network (DANN) to compensate for the discrepancy in RF signal translation. The domain discriminator of the DANN is to optimize the training variables in the feature extractor, thus allowing the network to concentrate on learning the generalized motion features, and ignore the technology-specific features for HAR.

The main contributions of this paper include the following.

- To the best of our knowledge, the TARF system is the first technology-agnostic human activity identification system capable of performing generalized and accurate HAR using various RF sensing platforms.

- We investigate the challenges in technology-agnostic HAR and show that they are caused by three main factors: metric disparities, heterogeneous sensitivity distributions, and diverse motion feature translations.

- A universal RF data preprocessing module is proposed to reduce the disparity between different RF sensing technologies. The sensitivity diversity is addressed by remapping the signal strength measurements, and generalized tensor data is constructed using STFT. The DANN is utilized to categorize different types of human activities, which further mitigates the interference from diverse RF domains.

- We develop a prototype of TARF to demonstrate the robustness of human activity recognition when data collection and testing are conducted using four different RF sensing technologies, including FMCW radar, WiFi in 2.4GHz and 5GHz bands, and RFID. The proposed system is compared with the traditional Convolutional Neural Network (CNN)-based technique, and the results validate that the proposed TARF system is resilient to technology-agnostic human activity recognition.

The remainder of this paper is organized as follows. We first review the related work on RF-based HAR and adversarial domain adaptation in Section 9.2, We then introduce the preliminaries and the problem statement in Section 9.3. Section 9.4 provides an overview of the proposed TARF system, and Section 9.5 presents the detailed design of the key TARF components. Section 9.6 presents the experimental evaluation of the TARF system and Section 9.7 concludes this paper.

9.2 Related Work

The prior works on human activity recognition can be roughly categorized as camera-based, sensor-based, and wireless-based techniques [200]. In this paper, we mainly focus on RF-based human activity recognition, including Radar-based, WiFi channel state information (CSI)-based, and RFID-based methods. We will review such related work, as well as the recent works on adversarial domain adaptation for wireless human activity recognition in this section.

9.2.1 RF-based Human Activity Recognition

Several radar systems have been utilized for human activity recognition, such as the Frequency-Modulated Continuous Wave (FMCW) radar, Doppler radar, and Ultra Wide-band (UWB) radar [201]. FMCW radar was first employed for human activity monitoring, e.g., through-wall monitoring [202], 3D passive human tracking [203], and vital sign monitoring [80], by measuring the distance and velocity of body movement. However, these works require special hardware (e.g., Universal Software Radio Peripheral (USRP)) to implement the RF sensing system (usually operating at 5.46–7.25 GHz), thus incurring a higher cost. Commodity mmWave radars (e.g., IWR1443BOOST from Texas Instruments) operating at 77 GHz have also been utilized for various RF sensing tasks, such as human activity recognition [199], user authentication [204], and vital sign monitoring [205]. In [206], vision data captured by the Vicon motion capture system was used for supervised training of the deep learning model, which constructs 3D human meshes from sparse point clouds. Doppler radar can detect the velocity and direction of the subject, and has also been utilized for human activity recognition [201]. Low-cost UWB devices have been shown useful for vital sign monitoring [207] and human activity recognition [208], where meta-learning was used to adapt to different deployment scenarios.

As a dominant wireless communications technology, there has been great interest in utilizing WiFi for human activity recognition. Several open-source tools have been developed to extract channel state information (CSI) from the Orthogonal frequency division multiplexing (OFDM) channel, such as for Intel 5300 cards [209], the ESP32 WiFi microcontroller [210], the nexmon CSI Extractor for Broadcom and Cypress WiFi chips [211], the Atheros CSI tool [212], and the openwifi tool [213]. CSI amplitude and phase difference data have been used in applications for activity recognition, vital sign monitoring, and gesture recognition [19, 198]. In addition, deep learning techniques have great potential for achieving high recognition accuracy [10]. For example, long short-term memory (LSTM), a recurrent neural network (RNN) architecture, outperformed the traditional model-based method on human activity recognition

using WiFi CSI amplitude data [214]. Generative adversarial networks (GAN) have been utilized to augment training data for human activity classification [215]. As in [206], high precision 3D skeletons captured by the Vicon motion capture system were used to supervise the training of the deep learning model that works with WiFi CSI data [131].

RFID is a near-field communication system originally developed for identifying tags attached to objects. Low-cost and lightweight RFID tags can be attached to the human body as wearable sensors for activity monitoring. Commodity RFID readers (e.g., the Impinj R420 reader) can extract RF phase angle, Doppler frequency, and Peak RSSI from received signals, which can be used to estimate the range between the tag and the reader antenna. RFID-based sensing is usually more resilient to environmental interference than other RF sensing methods (e.g., radar and WiFi) due to the short range. RFID sensing techniques have been developed vital sign monitoring [139], driver fatigue detection [138], activity recognition [192], and 3D human pose estimation [157].

9.2.2 Adversarial Domain Adaptation for RF Sensing

Although deep learning has a great potential for RF sensing applications, it still faces great challenges for real-world applications. This is because different deployment environments, different users, or different wireless devices will lead to different data distributions, i.e., domain (or distributional) shift will occur between the source domain and target domain. A well trained deep learning model may fail when applied to unseen data. To address this challenge, generative adversarial network (GAN) [215], meta-learning [208,216], and adversarial domain adaptation [217] have been proposed to adapt a trained deep learning model in the source domain to the new RF data in the target domain. In the following, we review several related works on adversarial domain adaptation methods for human activity monitoring.

Adversarial domain adaptation comprises feature learning, classifier learning, and domain adaptation, where adversarial training is leveraged to address the domain adaptation problem. The goal is to obtain an effective feature representation, which is discriminating for the learning tasks but invariant for the domain classifier. For example, the conditional domain adaptation architecture was used for radar-based sleep stage classification in different indoor scenarios,

which was focused on supervised tasks [218]. In [217], unlabeled data was used in adversarial training for human activity classification, where four wireless devices were adopted to remove the environment and subject specific information. Furthermore, multi-view deep learning was introduced to improve the classification accuracy in different environments by fusing different wireless data [219], while multi-adversarial domain adaptation was proposed for WiFi based in-Car activity recognition [220]. All the above related works were focused on environment and user adaptation using adversarial domain adaptation. Unlike the related works, in this paper, we develop a novel technology-agnostic human activity recognition framework utilizing different wireless techniques such as radar, WiFi, and RFID.

9.3 Towards Technology-agnostic Generalization

9.3.1 Preliminaries of the Wireless Technologies

To develop the generalized technology-agnostic approach for RF-based human activity sensing, we first present the preliminaries of RF sensing with different RF technologies.

FMCW Radar

Frequency-Modulated Continuous Wave (FMCW) radar is a useful technology to provide both distance and velocity measurements. With the FMCW radar, the transmitted signal is modulated in the form of chirps [221], whose frequency keeps on increasing periodically. In each period, the signal frequency f_M is modulated as:

$$f_M(t) = f_0 + \frac{Bt}{T_c}, \quad 0 \leq t \leq T_c, \tag{9.1}$$

where f_0 is the starting frequency, B is the bandwidth of the channel, and T_c is the duration of each period. The reflected chirp signal is received by the radar and fused with the transmitted signal. Based on (9.1), the fused signal $S_{FMCW}(t)$ at time t is given by [222]:

$$S_{FMCW}(t) = A \exp\left\{-j2\pi\left(f_0\tau + \frac{B\tau t}{T_c} - \frac{B\tau^2}{2T_c}\right)\right\}, \tag{9.2}$$

where A is the gain of the signal and τ represents the propagation delay of the backscattered signal. Eq. (9.2) indicates that the frequency of the fused signal, $B\tau/T_c$, is determined by the propagation delay of the signal τ. By multiplying the speed of light, τ will be translated to the distance between the reflecting object and radar. Taking Fourier Transform on the fused signal, the power spectrum can therefore reveal the reflected signal strengths from various distances, which is referred to as the range profile [222]. When a person moves inside the radar detecting range, the gathered range profile will change with the body movements. As a result, the received range profile can be leveraged to distinguish between different movement types.

Commodity 2.4GHz and 5GHz WiFi

The WiFi technology has also been explored as a promising solution to non-intrusive RF sensing of human activity. Compared to FMCW radar, WiFi is quite accessible due to the wide deployment of the infrastructure and the low-cost commodity devices. Recent WiFi sensing techniques mostly utilize the Channel State Information (CSI) extracted from the device driver, which provides a fine-grained representation of the orthogonal frequency-division multiplexing (OFDM) channel.

Considering the multipath effect of signal propagation, the CSI of a channel c can be written as [212]:

$$S_{WiFi}(c) = \sum_{n=1}^{N_c(t)} A_n \exp\left\{-j2\pi(f_c\tau_n + \phi_c)\right\}, \tag{9.3}$$

where $N_c(t)$ is the total number of propagation paths, f_c is the central frequency of channel c, ϕ_c is the phase offset of channel c, and A_n and τ_n are the gain and the propagation delay of the nth path, respectively. It can be seen that the channel offset largely determines the received CSI for all propagation paths. Human activity is captured in the CSI because, as a part of the propagation environment (or, the WiFi channel), moving human body parts can create considerable variations in most propagation paths, such as the gain A_n, the propagation delay τ_n, and even the total number of paths $N_c(t)$. As a result, both CSI amplitude and phase can be leveraged for learning human activity.

RFID

RFID devices have also been utilized in recent years for human activity monitoring. As wearable sensors, RFID tags are more resistant to environmental interference than broadband systems such as WiFi. Furthermore, RFID systems' low power consumption makes them a suitable RF sensing technology for the Internet of Things (IoT). The line-of-sight (LOS) path usually contributes to the dominant component in the received signal, and hence the received signal on a channel c can be written as:

$$S_{RFID} = A_c \exp\left\{-j2\pi(f_c\tau + \phi_c)\right\}, \tag{9.4}$$

where A_c, f_c, and ϕ_c are the gain, frequency, and phase on channel c, respectively, and τ is the propagation delay. With the Low Level Reader Protocol (LLRP) [38] used in RFID systems, the phase value of signal S_{RFID} can be extracted for sensing of human activities. By attaching tags to the human body, the propagation delay τ of each tag changes along with the movements of body parts. Thus, human activities can be captured by the variations in phase values of the attached tags.

9.3.2 Problem Statement

Developing the technology-agnostic approach for human activity recognition is highly challenging. So we need to formulate the problem and figure out the mechanism to integrate the RF based techniques. Following the Fourier expansion, we can write the source signal $S(t)$ of body movement as a combination of different periodical components, as

$$S(t) = \sum_{n=1}^{M} K_n \cos(W_n t + \phi_n), \tag{9.5}$$

where W_n is the frequency of the sinusoidal signal, K_n denotes the coefficient, ϕ_n represents the initial phase of each component, and M is the total number of periodic components. The set of parameters, i.e., $\{W_n, K_n, \phi_n, M\}$, represent the unique features of the corresponding human activity. The received signal reflected from the human body is dynamically distorted by

the moving human body, and the distortion is mainly due to the prorogation path changed by the activity $S(t)$. Therefore, the reflected signal can be written as

$$R(t) = A_T \exp\left\{-j\left(2\pi\frac{D+S(t)}{\lambda} + \phi_T\right)\right\}, \tag{9.6}$$

where D is the average distance of the propagation path, λ denotes the wavelength of the transmitted signal, and A_T and ϕ_T represent the amplitude and the initial phase of the signal, respectively. Eq. (9.6) indicates that the human activity introduces an offset on the phase component $\angle R(t)$. The relationship between $\angle R(t)$ and the source signal $S(t)$ can be further investigated in the frequency domain. Taking Fourier transform on $\angle R(t)$, we have

$$\begin{aligned}\Gamma(\omega) &= \int_{-\infty}^{\infty}\left\{\left(2\pi\frac{D+S(t)}{\lambda} + \phi_T\right)\right\} e^{-j\omega t}\, dt \\ &= \phi_c\delta(0) + \frac{\pi}{\lambda}\sum_{n=1}^{M} K_n[e^{j\phi_n}\delta(\omega - W_n) + e^{-j\phi_n}\delta(\omega + W_n)],\end{aligned} \tag{9.7}$$

where $\delta(\omega)$ is the Dirac function, and ϕ_c is a constant given by $\phi_c = 2\pi D/\lambda + \phi_T$. The expression of $\Gamma(\omega)$ illustrates the mapping from the source signal $S(t)$ to the phase of the reflected signal $\angle R(t)$. The phase is determined by the unique features of the human activity, i.e., $\{W_n, K_n, \phi_n, M\}$.

The challenge in many RF sensing applications is, accurate phase angle of the reflected signal is usually hard to obtain due to the multipath effect. The mapping from the human activity signal to the received phase sample becomes highly complex. Traditional RF HAR based systems employ various signal preprocessing techniques to combat the interference caused by the complex mapping, to extract useful features for motion classification. Unfortunately, a specific preprocessing method developed for one RF technology is usually not applicable to other RF technologies, due to their different frequency bands, different communication protocols, and different types of hardware. To address this challenge, the primary objective of our technology-agnostic approach TARF is to learn the generalized features of the human activity signal $S(t)$ from various RF technologies, which will facilitate the accurate classification of different human activities.

9.4 System Overview

9.4.1 Main Challenges

Each existing RF sensing based activity recognition system is closely tailored for the specific technology and has its unique advantages and certain limitations. Such a system designed for one RF technology usually does not work for a different technology (e.g., FMCW radar vs. RFID). Given the availability of various RF technologies in our daily lives, a technology-agnostic approach would be highly desirable to achieve better adaption to different sensing scenarios, as well as greatly reduce the barrier to deploying the system.

However, pursuing a generalized approach that works with very different RF technologies is a great challenge due to two main reasons. First, different RF technologies are established on different frequency bands. For example, the Ultra High Frequency (UHF) RFID systems operate on the 900 MHz band, WiFi works on the 2.4GHz or 5GHz bands, and the FMCW radar used in our experiments is on the $76 \sim 81$ GHz millimeter wave (mmWave) band. Even deployed in the same environment, different RF technologies see different propagation channels and different signal characteristics. Second, due to their different physical layer and medium access control layer protocols, as well as different device drivers that are available, the RF data collected by different RF devices are highly diverse. It is extremely challenging to develop a generalized approach to accurately detect human activities from such diverse RF data.

9.4.2 System Architecture

To address the above challenges, we design a novel system TARF that is generalizable to diverse RF data measured with different RF sensing technologies to perform technology-agnostic human activity recognition. Fig. 9.2 provides an overview of the system architecture of TARF, which is composed of three main components, including (i) RF signal collection, (ii) generalized RF signal preprocessing, and (iii) domain adversarial deep neural network based activity recognition. In the RF signal collection module, raw RF signals are sampled by several different RF sensing platforms. According to (9.7), we are interested in the phase angle $\angle R(t)$ of the collected signals. Thus, we use the phase signal from the RFID system, and the phase

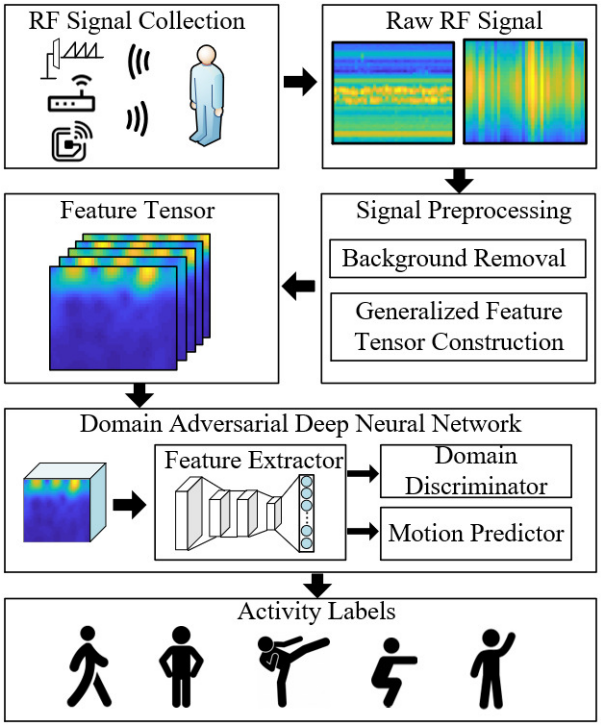


Figure 9.2: Architecture of the proposed technology-agnostic RF sensing system TARF.

difference signal from 2.4GHz and 5GHz WiFi systems. For FMCW radar, phase is not a good indicator of human activity because of the modulated frequency. Instead, we leverage the range profile from the FMCW radar as an input signal, which is indicative the propagation distance of the signal and is readily available from the device.

In the proposed technology-agnostic TARF system, the signal will be treated using the same generalized signal preprocessing module, no matter which RF technology is used for sensing. Generalization to multiple RF technologies should begin with a standardized RF data format. We propose a generalized RF data preprocessing module for different RF sensing systems, where the input signal is treated as a group of different observations of the same source signal $S(t)$ and is converted to a generalized input data matrix. Then the general background removal is implemented with Hampel filters, where the interference from the common static background is removed. Finally, the observations are reordered according to their signal strengths to mitigate the sensitivity diversity across different RF devices.

Then, generalized feature tensors will be constructed and fed into a domain adversarial deep neural network for activity classification. In comparison to a traditional convolutional

neural network, the domain adversarial neural network is able to acquire more generalized features of diverse human activities by combating characteristics gained from other domains. The details of the signal preprocessing and the domain adversarial deep neural network structure will be elaborated in Section 9.5.

9.5 Design of the Technology-Agnostic System

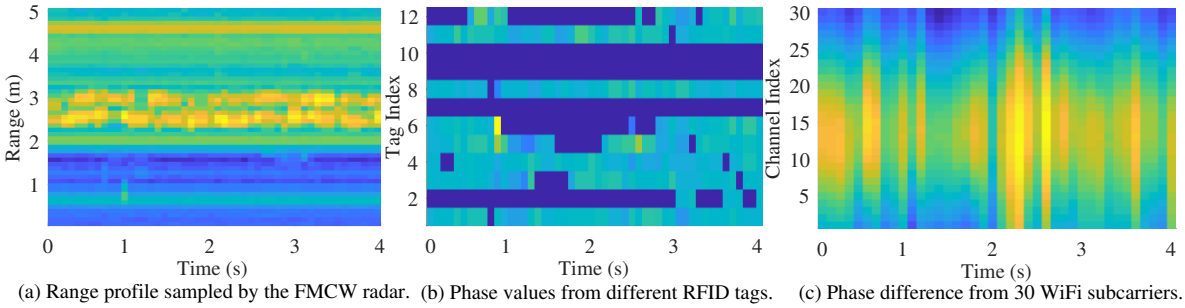
In this section, we present the detailed design of the proposed TARF system. We will examine the challenges in diverse RF data, generalized feature remapping, and activity recognition with domain adversarial neural networks, and then present our proposed solutions to address these challenges.

9.5.1 Metric Generalization

Challenge: Diversity in Measured Data

The first challenge of generalization arises from the use of very different types of channel data. Fig. 9.3 illustrates the raw data collected for the same human activity sampled by three different RF sensing platforms over a 4-second period. For convenience, we have normalized the data from each platform. In these "temperature" plots, a lighter color, e.g., yellow, indicates larger values or higher strengths, while a darker color, e.g., dark blue, represents smaller values. Each plot in Fig. 9.3 presents a different type of sampled data. Specifically, Fig. 9.3(a) shows the raw range profile sampled by an FMCW radar, where the 2-dimensional data (or, matrix) consists of the signal strengths sampled over time for different ranges from 1 m to 5 m. The RFID data, plotted in Fig. 9.3(b), comprises the phase values sampled from 12 RFID tags. The WiFi data, plotted in Fig. 9.3(c), consists of the phase differences, i.e., the difference of phases from a pair of WiFi antennas, sampled from 30 subcarriers.

It is obvious that such RF data are very different, each with their unique features, making it extremely hard to handle with a generalized model. In particular, for the three RF technologies, the range profile of FMCW radar is measured in decibel (dB) typically ranging from 20 dB to 120 dB. The phase values sampled from the RFID tags, as well as the phase differences from



(a) Range profile sampled by the FMCW radar. (b) Phase values from different RFID tags. (c) Phase difference from 30 WiFi subcarriers.

Figure 9.3: Raw data sampled by different RF technologies for the same human activity over a 4-second period.

the 30 subcarriers of the WiFi channel, vary from 0 to 2π rad. The different metrics incur considerable diversity in the scale of data measurement. Moreover, such RF data with different metrics cannot be directly generalized into a normalized format.

Proposed Solution

The first step in data preprocessing is to remove the disparity in metrics of different RF hardware platforms. We begin by defining the generalized data matrix in order to gather raw data from different kinds of RF platforms. The generalized data matrix has the following format:

$$\mathbf{S}_G = \begin{bmatrix} F_1^1 & F_2^1 & \dots & F_{N_t}^1 \\ F_1^2 & F_2^2 & \dots & F_{N_t}^2 \\ \vdots & \vdots & \ddots & \vdots \\ F_1^{N_F} & F_2^{N_F} & \dots & F_{N_t}^{N_F} \end{bmatrix}, \tag{9.8}$$

where F represents a RF data measurement for human activities and N_F denotes the total number of measurements. The integer N_t denotes the number of time frames captured by the RF device. The sampling frequency of all RF platforms is set to 10 Hz in (9.8), therefore the length of the x-axis N_t is given by $N_t = 10 \times t$. Meanwhile, the amount of measurements taken by different RF platforms determines the size of the y-axis dimension. Each measurement is regarded as an observation of the source signal $S(t)$, which is generated by the human activity.

With FMCW radar, the subject performs activities within its range and the range profile in the form of power spectrum is obtained by 256-point fast Fourier transform (FFT). Therefore

we have $N_F = 256$ for the FMCW platform. With WiFi, a transmitter sends packets to a receiver, with the subject in the middle. The receiver has three antennas and each can extract phase data from 30 subcarriers, which results in 90 RF measurements for each received packet. For RFID based sensing of human activities, we attach 12 tags to the joints of the subject, including neck, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, pelvis, left hip, right hip, left knee, and right knee. Three RFID readers are used to interrogate the tags, while phase data is collected from received responses. Unfortunately, because of the Slotted-Aloha-like collision avoidance protocol, only one phase measurement can be collected by the reader at a time. We employ an effective tensor completion based data interpolation method [157] to augment the sparse RFID phase data. After tensor completion, each time frame has 36 phase samples (i.e., from 3 antenna and 12 tags). With the generalized data matrix employed in the system, we can remove the diversity in metrics of the diverse RF platforms.

9.5.2 Generalized Feature Remapping

Challenge: Diversity in Sensitivity

No matter which RF technique is employed, a number of measurements are gathered at the same time. However, the sensitivity of these measurements to human activity may be highly different. For example, in Fig. 9.3(a), the signal strength around 2.5 m is more sensitive to the human movements, because this is the average distance between the subject and the FMCW radar. As a result, measurements taken at a distance of 2.5 m should contribute more to the correct extraction of motion features. When it comes to the RFID technology, however, the situation is completely different. The sensitivity, as shown in the measurements, is strongly dependent on the limb where the RFID tags are attached, since the received phase value is determined by the movements of the RFID tags. The high-sensitivity data should be be more emphasized for accurate activity detection. Such diverse sensitivity of different RF platforms also poses a challenge to the development of a general technology-agnostic approach.

Proposed Solution

To deal with the diverse sensitivity in measured RF data, we should remap the RF measurements from different wireless technologies into a generalized order, so that the same human action introduces a comparable distribution of measurements. Since the wireless propagation environment has a great impact on the measured signals, the environment influence should be firstly removed before the remapping process. Thus, we measure component corresponding to the impact of the static background, and eliminate it from the sampled signal. Such a background removal operation is performed on each row of matrix $\mathbf{S}_G$ using two separate Hampel filters. The first filter has a window size of 4 for thermal noise reduction, while the other has a window size of 15 for extracting the background component. To extract the component corresponding to the human activity, we subtract the signal filtered by the Hampel filter with the larger window size from the signal filtered by the Hampel filter with the smaller window size. After removing the background component, we remap all the rows in $\mathbf{S}_G$ according to the signal strength.

The signal strength for row i, denoted by P_i, is computed by the variance of the time sequence as follows:

$$P_i = \frac{1}{N_t} \sum_{t=1}^{N_t} (F_t^i - \mu_i)^2, \tag{9.9}$$

where μ_i represents the mean value of each row. Since the background component has been removed, the variance of the signal indicates its strength. The new matrix is sorted in descending order of signal strength. We always choose the first N_P rows of data for human activity recognition. The reordered matrix $\mathbf{S}_R$ is given by:

$$\mathbf{S}_R = \begin{bmatrix} F_1^1 & F_2^1 & \dots & F_{N_t}^1 \\ F_1^2 & F_2^2 & \dots & F_{N_t}^2 \\ \vdots & \vdots & \ddots & \vdots \\ F_1^{N_P} & F_2^{N_P} & \dots & F_{N_t}^{N_P} \end{bmatrix}, \tag{9.10}$$

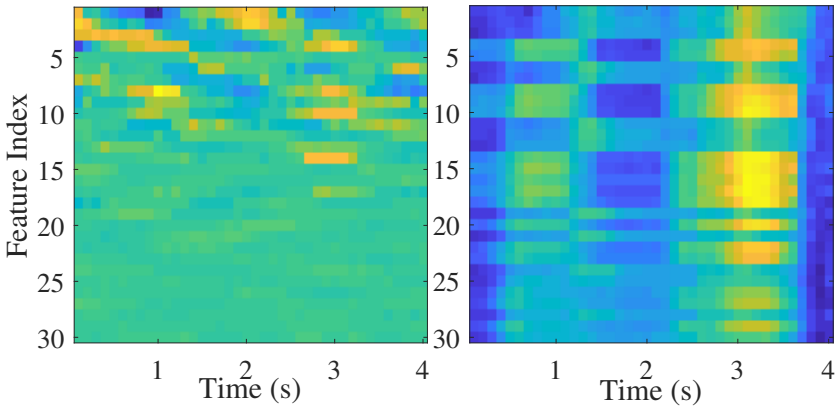


Figure 9.4: Examples of the calibrated generalized feature matrix $\mathbf{S}_R$ measured from the kicking activity. Left: sampled with FMCW radar; Right: sampled with 5GHz WiFi.

where N_P is the number of the most powerful signals chosen for activity recognition.

In Fig. 9.4, we present the examples of the reordered matrix $\mathbf{S}_R$ of the data collected by FMCW radar and 5GHz WiFi devices, where $N_P = 30$ for a period of four seconds. It can be seen from the figures that the dimensions of the two signals are equivalent for the same sampling period. Furthermore, since the background component has been eliminated and the samples are reordered according to their strength, the overall sensitivity distributions of the two different RF technologies are now similar to each other.

With the metric generalization and the generalized feature remapping process, regardless of the physical meaning of the raw sampled data, all RF data measurements could be converted to a generalized data tensor for human activity recognition. Thus, the TARF system could implement the same signal preprocessing framework for different wireless technologies, which means the system is also adaptable to RF technologies other than the four technologies illustrated in the paper.

9.5.3 Activity Recognition with Domain Adversarial Neural Network

Challenge: Diversity in Motion Feature Translation

Since different RF technologies utilize different protocols and frequency bands, the translation from received RF measurement to the target activity is highly diverse. Although the same source signal is generated by the same activity, it is transformed into very different RF data by

the different protocols, frequency bands, and hardware. For example, with the RFID system, the human activity directly changes the positions of the RFID tags attached to the body. Following (9.4), the change of tag position will introduce significant variation in the propagation delay τ. However, with the WiFi system, the human activity only affects part of the propagation paths of the OFDM channel as shown in (9.3). Furthermore, Eqns. (9.3) and (9.4) show that the different channel frequencies used in RFID and WiFi also generate large diversity in measured RF data. The considerable frequency diversity (i.e., 900 MHz in RFID and 2.4 GHz and 5 GHz in WiFi) causes large variation on the motion feature transformation in the raw sampled signals. Since such translation diversity is highly complicated and nonlinear, we propose a two-step solution to deal with this challenge, i.e., (i) time-frequency (TF) domain transformation and tensorization, and (ii) a domain adversarial neural network model.

Proposed Solution Step 1: TF Domain Transformation and Tensorization

Given that human activity can be seen as a mixture of distinct periodic components [223], the characteristics recorded in the frequency domain are more universal than that in the time domain. To extract the generalized features from the remapped matrix $\mathbf{S}_R$, we perform Short Time Fourier Transform (STFT) on the remapped matrix to convert each row to a Time-Frequency (TF) domain matrix, and then construct a TF tensor with N_P slices. The TF domain data incorporates both frequency domain properties and variations over time. Fig. 9.5 presents one slice of the generalized TF tensor data, when the window size used in STFT is set to 16. The examples show that, although sampled by two different RF technologies, i.e., FMCW radar and 5GHz WiFi, the TF domain data is well generalized with the proposed approach. Thus the same human activity will produce similar features as shown in the generalized feature tensor. Deep learning models can then be applied to classify different activities using such generalized TF tensor data.

Proposed Solution Step 2: Domain Adversarial Neural Network

We propose to use a domain adversarial deep neural network [224] to recognize human activities using the generalized feature tensors. Compared with the traditional CNN models, the

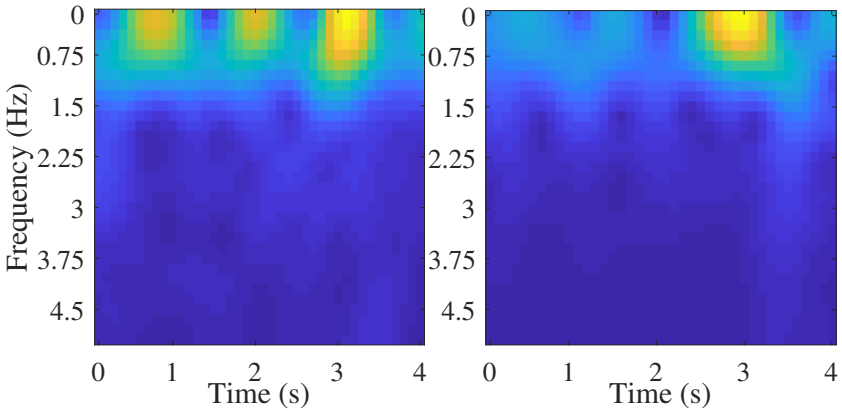


Figure 9.5: Examples of one slice of the generalized feature tensor for the kicking activity. Left: sampled with FMCW radar; Right: sampled with 5GHz WiFi.

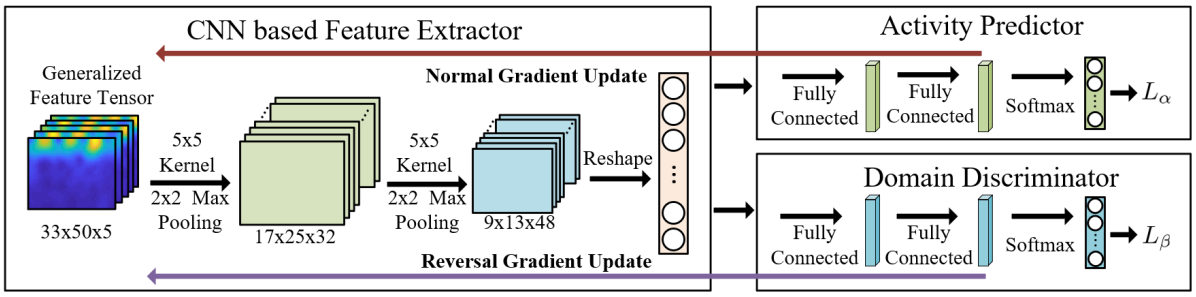


Figure 9.6: Structure of the domain adversarial deep neural network used in the TARF system.

domain adversarial neural network can further optimize the feature extractor with the domain discriminator. The network structure used in TARF is shown in Fig. 9.6, which is composed of the CNN based feature extractor, the activity predictor, and the domain discriminator.

a) Feature Extraction with CNN: The feature extractor used in the deep neural network is based on CNN. As a classic neural network structure, CNN could effectively extract the feature from all the slices in the generalized tensor. As Fig. 8.2 shows, the CNN feature extractor consists of two convolution layers, where the convolutional kernels used for feature extraction all have a size of 5×5. Each convolution layer is connected to a 2×2 max pooling layer to downsample the extracted feature.

The final feature is formalized as a one-dimensional vector, which is used for the following activity predictor and domain discriminator. The generalized tensor used as the input is sampled every five seconds and transformed by 64-dot STFT. We only use data in the positive

frequency domain, including 0 Hz, so the dimension of each slice is 33×50. The slice number is determined by N_p, which is equal to 30. We find the CNN-based feature extraction for all data slices may generate too many training variables, making the training time-consuming. Thus, we downsample the data tensor from 30 slices to 5 slices to reduce the complexity of network training. After the two feature extraction layers, the final feature is reshaped into a vector of 5616 elements.

b) Motion Identifier with Domain Discriminator: After extracting features using the CNN, the activity predictor and domain discriminator are applied, which consist of two fully connected layers. The final classification probability is calculated by the Softmax function. The loss function of the activity label predictor, denoted by L_α, is calculated by the cross entropy between the Softmax output and the activity label as:

$$L_\alpha = \frac{1}{N_b} \sum_{b=1}^{N_b} \sum_{k=1}^{N_a} \hat{y}_k^b \log \left(y_k^b \right), \tag{9.11}$$

where N_b is the number of training data in a batch, N_a is the number of classes of human activities, $\hat{y}_k^b$ denotes the estimated probability for class k with data sample b, and y_k^b is the class label which is either 0 or 1. L_α represents the accuracy of human activity prediction, and the deep neural network is trained by minimizing L_α using the gradient descent algorithm.

In addition to the activity predictor, the domain adversarial neural network also employs a domain discriminator to combat the diversity between different domains, i.e., different RF technologies. The loss function of the domain discriminator, denoted by L_β, is calculated similarly as L_α.

$$L_\beta = \frac{1}{N_b} \sum_{b=1}^{N_b} \sum_{q=1}^{N_d} \hat{y}_q^b \log \left(y_q^b \right), \tag{9.12}$$

where N_d indicates the number of RF technologies for data sampling and $\hat{y}_q^b$ denotes the estimation probability for the qth RF technology in the bth sample in the batch. Unlike the normal gradient descent learning algorithm used for maximizing L_β, the domain adversarial neural network performs a reversal gradient update for minimizing L_β, and the training variables of

the network are updated as [224]:

$$\hat{X_\gamma} = X_\gamma - \xi \left(\frac{\partial L_\alpha}{\partial X_\gamma} - C_r \frac{\partial L_\beta}{\partial X_\gamma} \right) \tag{9.13}$$

$$\hat{X_\alpha} = X_\alpha - \xi \frac{\partial L_\alpha}{\partial X_\alpha} \tag{9.14}$$

$$\hat{X_\beta} = X_\beta - \xi C_r \frac{\partial L_\alpha}{\partial X_\beta}, \tag{9.15}$$

where X_γ denotes the training variables in the feature extractor; X_α and X_β represents the training variables for the label predictor and the domain discriminator, receptively; ξ denotes the learning rate; and C_r is the combating rate. The training goal for the feature extractor is to maximize L_β and minimize L_α, hence the feature extractor will be trained to ignore the domain-related features. Accordingly, the network will learn the generalized human activity related features and abandon the features associated with different RF technologies.

9.6 Implementation and Evaluation

9.6.1 Experiments Setup

Hardware Platforms

To evaluate the proposed technology-agnostic human activity recognition system, we develop a prototype using several RF technologies, including FMCW radar, 2.4GHz WiFi, 5GHz WiFi, and the UHF RFID system. The hardware platforms are shown in Fig. 9.7. The FMCW radar employed in the system, as shown in the figure, is an IWR1843 BOOST single-chip FMCW mmWave sensor that operates at $76 \sim 81$ GHz. The WiFi devices are integrated with a standard Intel 5300 network interface card (NIC), which operates at either 2.4 GHz or 5 GHz. The RFID platform consists of three S9028PCR polarized antennas, one Impinj R420 reader, and ALN-9634 (HIGG-3) passive RFID tags. An MSI laptop with an NVIDIA GTX 1080 GPU and an Intel Core i7-6820HK CPU are used for signal processing, model training, and inference.

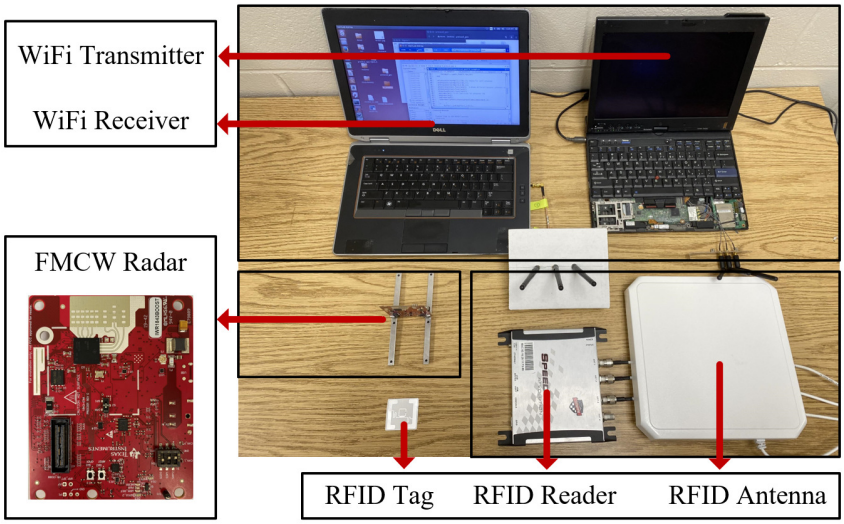


Figure 9.7: RF platforms used in our implementation and experiments.

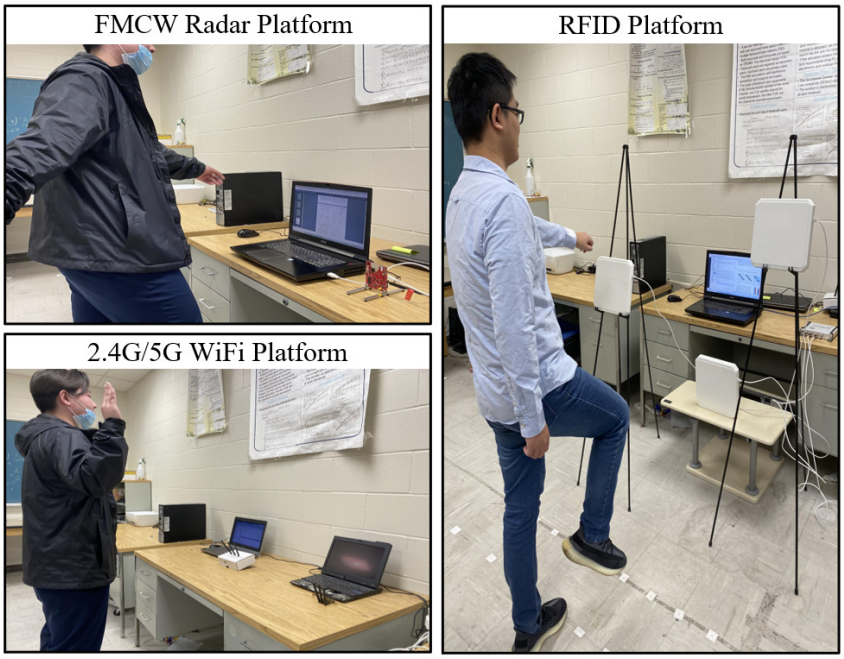


Figure 9.8: Human activity data sampling for different RF platforms.

Dataset Collection

RF data has been collected by sampling activities performed by a subject in front of the RF sensing platforms. The individual conducts seven types of different activities, including standing still, walking, running, squatting, body twisting, kicking, and hand waving. The data is sampled when the subject continuously repeats the activities. During the data acquisition using WiFi devices, the WiFi transmitter is set to the *injection* mode while the receiver is set to the

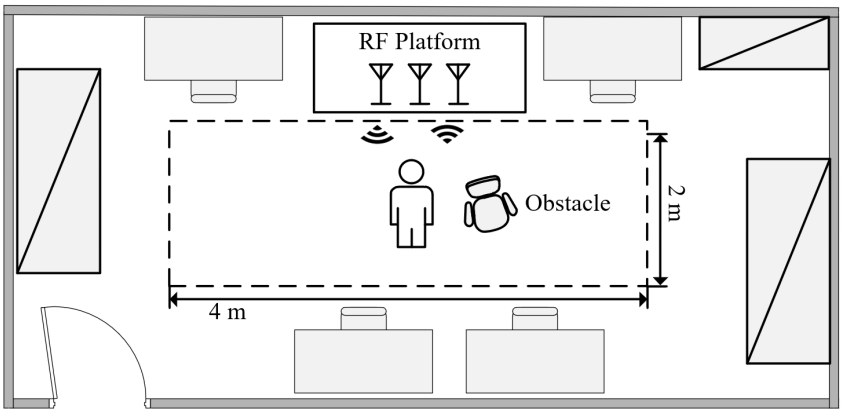


Figure 9.9: The environment of Human activity data sampling

monitor mode [212]. Two industrial, scientific and medical (ISM) bands, 2.472 GHz and 5.3 GHz, are used for the WiFi system, which allows us to examine the impact of different bands with the same WiFi protocol. RFID-based sampling is carried out with 12 passive RFID tags attached to the 12 joints on the subject's body, including neck, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, pelvis, left hip, right hip, left knee, and right knee. Three polarized antennae are used to interrogate these tags to ensure that each RFID tag is covered by at least one antenna. The FMCW radar employed in the investigations is a well-developed commodity mmWave sensor that produces range profiles for the scanned area.

Each of these four RF technologies can independently sample human activities and the data is processed by the proposed TARF framework.

The detailed deployment for different platforms and experiments configuration are illustrated in Fig. 9.8 and Fig. 9.9. As the figure shows, the experiments are conducted in an office lab. The subject performs different activities in the selected $2m \times 4m$ scanning area illustrated in Fig. 9.9. In the experiments, we emulate three different usage scenarios for the RF-based HAR demonstration application, which includes LOS monitoring, NLOS monitoring, and dynamic environment monitoring. The LOS monitoring is implemented with cleaning space in the scanning area, while the NLOS monitoring is conducted by adding chairs between the subject and the RF platforms. Although the LOS propagation is not entirely eliminated, the obstacle, like chairs, we used in the experiments could effectively mitigate the signal strength of the LOS component. In addition, the dynamic environment is introduced by moving people around the

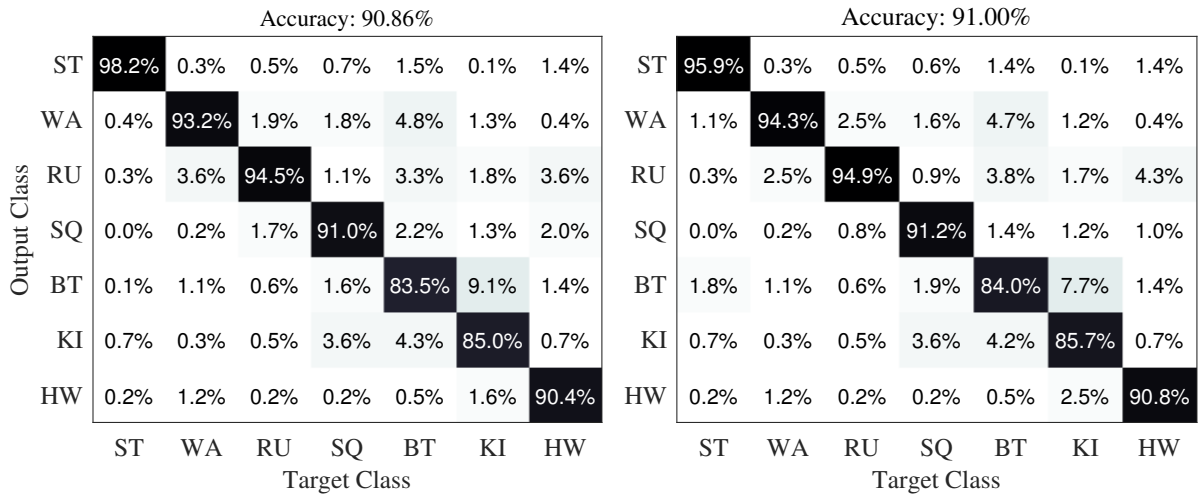


Figure 9.10: Confusion matrix of human activity recognition with a single RF technology (i.e., the FMCW radar). Left: the CNN baseline scheme; Right: TARF.

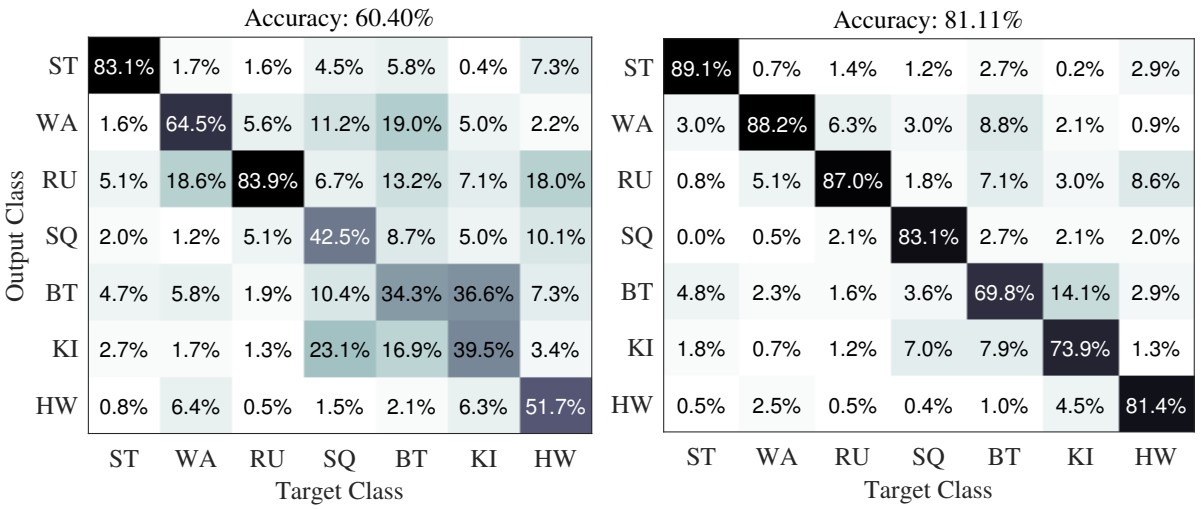


Figure 9.11: Confusion matrix of human activity recognition obtained using TARF.

testing subject when the RF data is being collected. In our experiments, 90% of the sampled data is used for the network training, and the remaining 10% is used for testing.

9.6.2 Performance with Different RF Technologies

To analyze the experimental results, we define the number of correctly classified data samples as the true positive number (TP), and the number of mistakenly recognized results as the false negative number (FN). The true positive rate (TPR) and false negative rate (FNR) are calculated

as:

$$TPR = \frac{TP}{TP + FN} \tag{9.16}$$

$$FPR = \frac{FN}{TP + FN}. \tag{9.17}$$

The overall evaluation result is presented in the confusion matrix format, which is composed of the TPRs and FPRs for all the seven types of activities. The overall accuracy η is calculated by:

$$\eta = \frac{\sum_{i=1}^{7} TP_i}{\sum_{i=1}^{7}(TP_i + FN_i)}, \tag{9.18}$$

where TP_i and FN_i denotes the true positive number and false negative number for target activity i, respectively. For convenience, we label different activities with the following acronyms: standing still–*ST*, walking–*WA*, running-*RU*, squatting–*SQ*, body twisting–*BT*, kicking–*KI*, and hand waving–*WH*.

To demonstrate the performance of the TARF system, we evaluated it using different combinations of the RF platforms, ranging from using a single RF technology to using all four RF technologies. The baseline for comparison is the traditional CNN based classification network [225]. The CNN network is composed of the same feature extractor and activity predictor as in TARF, but without the domain discriminator or reversal gradient update. Fig. 9.10 presents the confusion matrices obtained with a single RF technology (i.e., the FMCW radar). The left confusion matrix is the result obtained by the traditional CNN based approach, and the right confusion matrix is generated by the proposed technology-agnostic TARF system. The overall accuracy is 90.86% for the baseline CNN method and 91.00% for the proposed TARF approach. These results demonstrate that both CNN and TARF can effectively distinguish the seven types of human activity. This is because, when there is only one data domain, the influence of the domain discriminator could be ignored. Therefore, the performance of domain adversarial deep neural network is equivalent to that of the traditional CNN model.

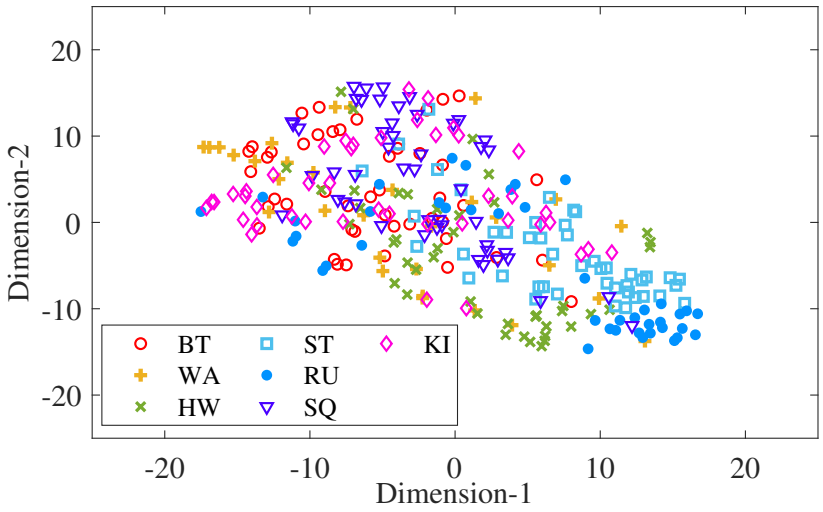


Figure 9.12: T-NSE illustration of human activity recognition from four RF technologies obtained using CNN baseline scheme.

We next examine the case when all the four RF sensing technologies are used for data acquisition. Fig. 9.11 presents the confusion matrices when all four technologies are utilized for human activity recognition. The confusion matrix on the left is obtained with the CNN baseline method, whose accuracy is significantly reduced from 90.86% in Fig. 9.10 to 60.40% here. The *TPR* of identifying body twisting and kicking is mostly affected, which drop to 34.3% and 39.5%, respectively. The confusion matrix shows that, with the data sampled with four different RF technologies, the CNN method fails to effectively learn the generalized features of different human activities.

The confusion matrix obtained with the proposed TARF approach is presented on the right side of Fig. 9.11. In contrast, TARF still achieves an overall accuracy of identification of 81.11%, although still affected by using the more diverse RF data collected from four different platforms. Such robustness to diverse RF data is achieved by the domain discriminator used in TARF. The domain discriminator can prevent the network to learn the domain-related features, and thus the technology-agnostic learning approach is quite effective to adapt to different RF technologies.

We also perform T-distributed Stochastic Neighbor Embedding (T-SNE) on the tested data for the CNN baseline scheme and proposed TARF system. T-SNE is an effective approach for visualizing high dimensional data introduced by the feature extractor, which could illustrate all

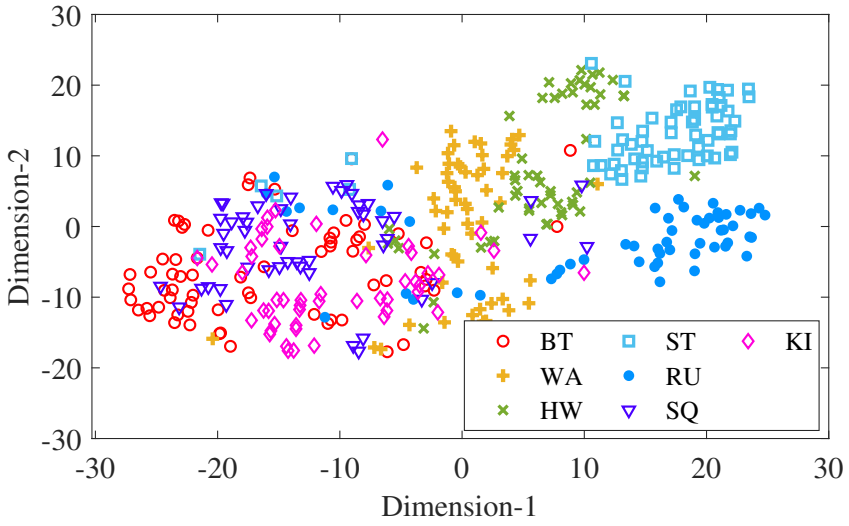


Figure 9.13: T-NSE illustration of human activity recognition from four RF technologies obtained using TARF .

tested RF data on a two-dimensional map. As illustrated in Fig. 9.12 generated by CNN baseline, the data collected same activity is not grouped effectively. Except for a few activities like standing still and running, the feature of other human activities are not satisfactorily extracted due to interference from various RF technologies. In contrast, as shown in Fig. 9.13 obtained by TARF , the data from different human behaviors are better grouped in the 2D map. Although there is some overlap between data groups from body-twisting, squatting, and kicking, the RF data sampled from different human activities are effectively grouped. The visualization data obtained from T-SNE could intuitively demonstrate that the domain discriminator could optimize the activity feature extractor by mitigating the interference from RF technology diversity.

The superiority of the TARF system is further demonstrated by Fig. 9.14, which shows the variations in accuracy as the number of RF technologies is increased from one to four. The figure shows that both CNN based scheme have similar classification accuracy when single RF technology is involved in the system, which are 90.86% and 91.00% respectively. The figure also illustrates that using more RF technologies will introduce increased diversity of the acquired data, which affects the classification performance. However, compared with the CNN baseline, the proposed TARF is effective in combating such diversity and adapting to different data domains.

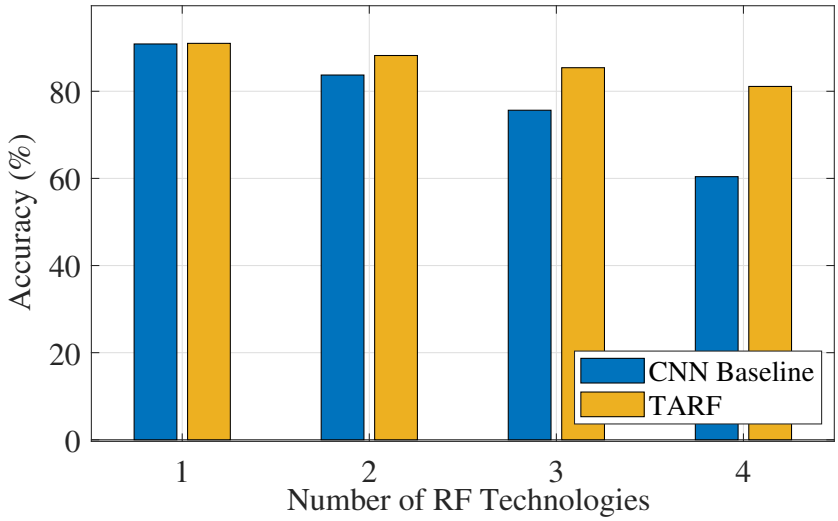


Figure 9.14: Accuracy performance with different combination of RF technologies (1: RFID only; 2: RFID and 2.4GHz WiFi; 3: RFID, 2.4GHz WiFi, and 5GHz WiFi; 4: RFID, 2.4GHz and 5GHz WiFi, and FMCW radar).

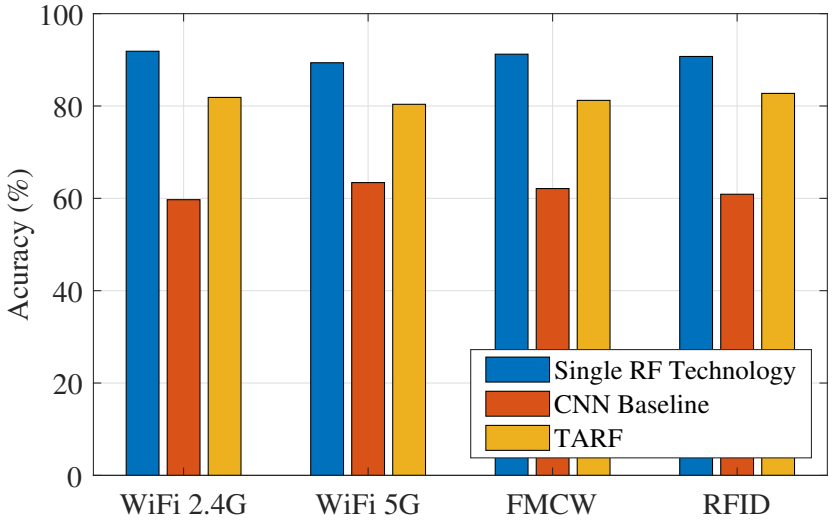


Figure 9.15: Accuracy performance in LOS testing scenario.

9.6.3 System Evaluation with Different Scenarios

We also evaluate the proposed TARF system in different scenarios and compare it with the HAR system trained by RF data from single RF technologies. In the experiments, we intend to investigate system performance with three different usage scenarios, such as LOS testing scenario, NLOS testing scenario, and dynamic RF environment testing scenarios. The NLOS environment is emulated by adding obstacles between the subject and the RF platforms, and the dynamic RF environments are emulated by moving people around. We trained the network

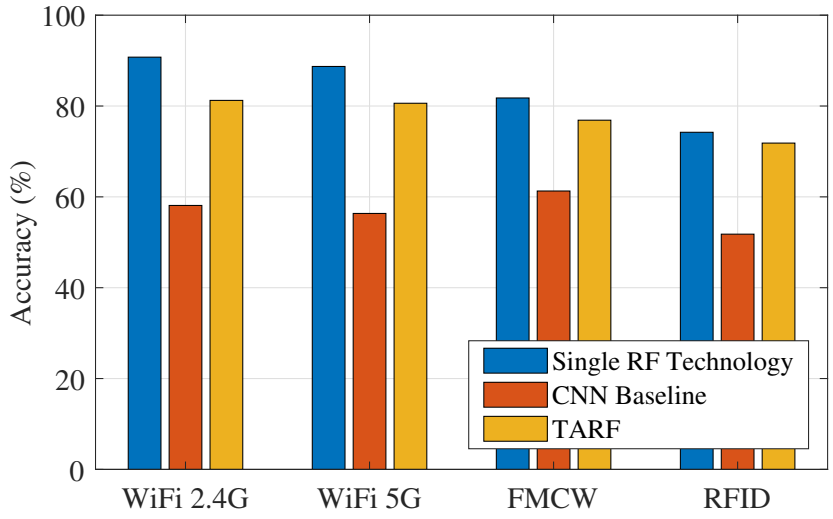


Figure 9.16: Accuracy performance in NLOS testing scenario.

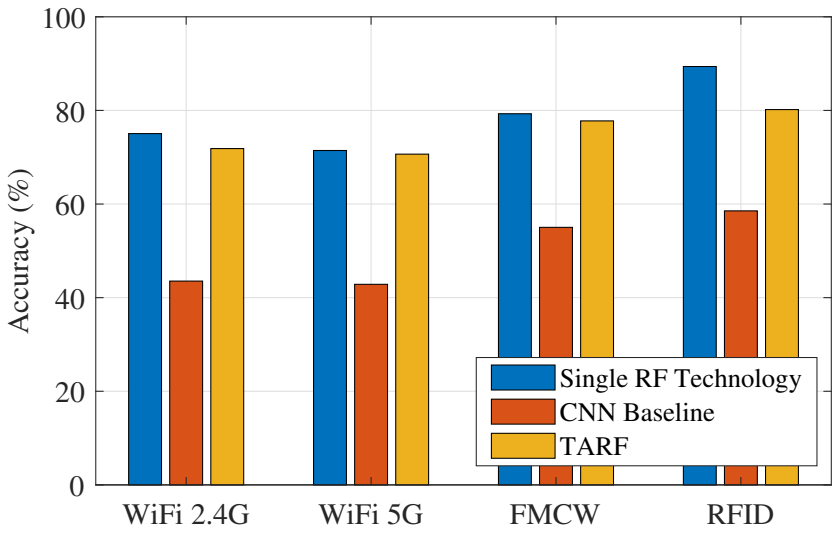


Figure 9.17: Accuracy performance in dynamic RF environment testing scenario.

with RF data from a single RF platform to generate the corresponding technology-specific system baseline, while the technology-generalized schemes are trained with data from all four RF technologies. Fig. 9.15 illustrates the activity recognition accuracy of different systems in LOS testing scenarios. The figure shows that all technology-specific schemes can achieve satisfactory activity recognition accuracy in the LOS testing scenario, which are all over 89.37%. The performance of technology-generalized systems, such as CNN baseline and TARF, is worse than the technology-specific system. However, as a generalized system, the accuracy of TARF is also much higher than the CNN baseline. Although TARF could not outperform the single

RF technology scheme, a trained technology-agnostic system could achieve satisfactory accuracy for any RF technology. For example, the lowest recognition accuracy 80.34% is achieved when tested with WiFi 5G, and the highest accuracy is 82.73% tested with RFID.

However, the LOS testing scenario is too ideal for the practical application, so we also evaluate the system in the NLOS scenario and dynamic, noisy environment. Fig. 9.16, and Fig. 9.17 illustrate the system accuracy when the test data is sampled from NLOS and dynamic RF environment, respectively. From the figures, we could observe that the interference of the NLOS environment varies between different RF technologies. Among these four RF technologies, WiFi-based schemes, such as 2.4GHz and 5GHz, could still achieve high accuracy, but the accuracy of FMCW and RFID-based systems descend to 81.77% and 71.22%, respectively. This is because the WiFi signal is wide broadcasting, so the NLOS component could also convey informative features of human activities. In contrast, due to the limited scanning range of the FMCW radar and RFID system, the LOS component contributes dominant information for human activity recognition. Especially for the RFID system, most tags are not normally interrogated because of the blockage of the obstacle.

When it comes to dynamic RF environments, the performance of these sigle RF systems is changed. As Fig. 9.17 shows, the accuracy of the two WiFi schemes is significantly attenuated by the noise generated by the moving person. The accuracy of the WiFi-specific schemes decreases to 75.05% and 71.44% for 2.4GHz and 5GHz, respectively. However, the accuracy of the RFID system remains relatively high accuracy as 89.38%. This is because the RFID tags attached to the subject clothes could convey reliable human movement features, which are more robust to the dynamic environment than WiFi-based schemes.

The results in Fig. 9.16, and Fig. 9.17 imply that single RF technology can not be adaptable to all kinds of testing environments. Therefore, an effective technology-agnostic system could incorporate all accessible RF technologies so that different RF technologies could complement the shortcomings of other RF technologies. Table 9.1 illustrate the accuracy of all single RF technology schemes, CNN baseline, and TARF system. The table shows that, when all accessible RF technologies are involved in the generalized system, TARF could achieve 81.24% and 80.18% activity recognition accuracy for NLOS and dynamic environment, which

Table 9.1: Accuracy comparison with different testing scenarios

Testing Environment	*WiFi 2.4GHz*	*WiFi 2.4GHz*	*FMCW*	*RFID*	*CNN Baseline*	*TARF*
LOS	91.86%	89.37%	91.22%	90.73%	63.41%	82.73%
NLOS	90.76%	88.71%	81.77%	74.22%	61.29%	*81.24%*
Dynamic Environment	75.05%	71.44%	79.29%	89.38%	62.54%	*80.18%*

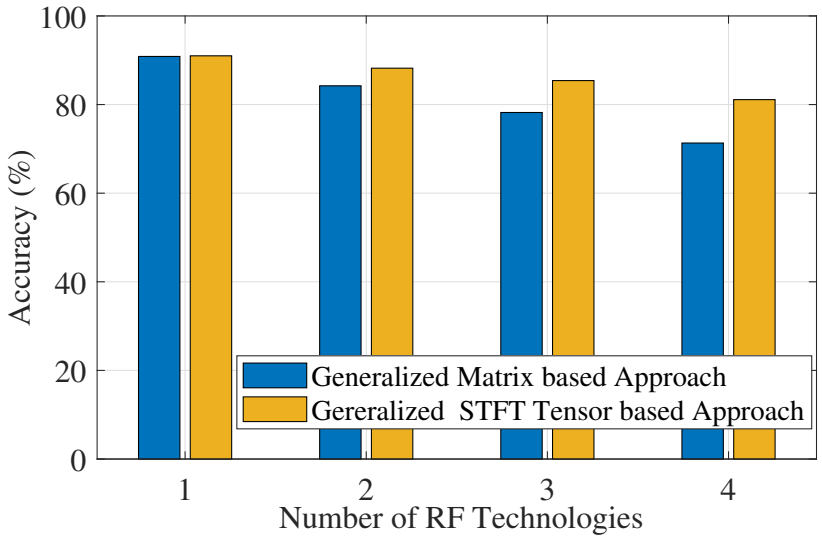


Figure 9.18: Performance comparison between the generalized matrix-based approach and the proposed STFT tensor-based approach.

is comparable to the accuracy in the ideal LOS testing scenario. The experiments result shown in the table demonstrate that, as a technology-agnostic system, TARF could be performed on various RF technologies and achieve robust HAR performance for different testing scenarios.

9.6.4 Impact of the Generalized Feature Tensor

We also conduct experiments to examine the benefit of utilizing the extended STFT feature tensor, and to establish the most appropriate tensor-related parameters. The accuracy performance of human activity recognition is presented in Fig. 9.18, where the blue bars are the results obtained by just utilizing the generalized matrix $\mathbf{S}_R$ as input to the deep neural network. It can be seen that using the generalized matrix can achieve a 90.86% recognition accuracy in a single-technology situation, but the accuracy drops dramatically to 71.32% when all the four technologies are used. The proposed STFT tensor-based technique results are represented by the green bars, which degrades from 91.00% to 81.11% instead, and is more resilient to

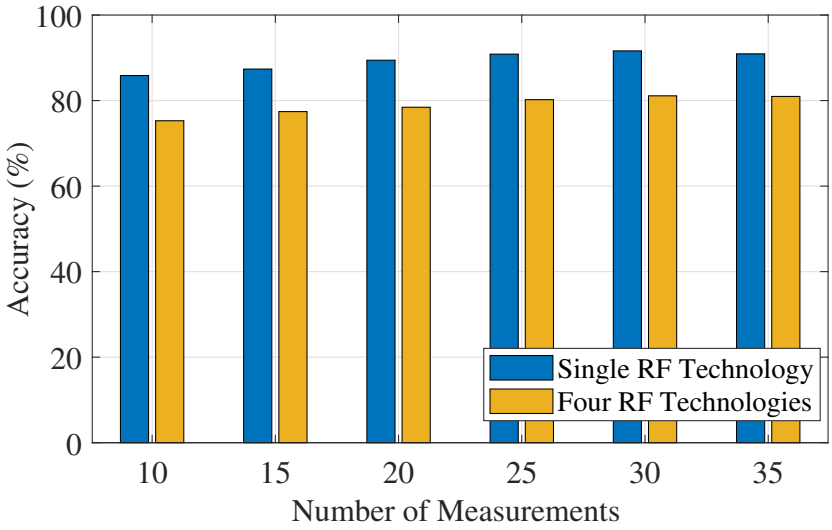


Figure 9.19: Activity recognition accuracy when different numbers of measurements are used.

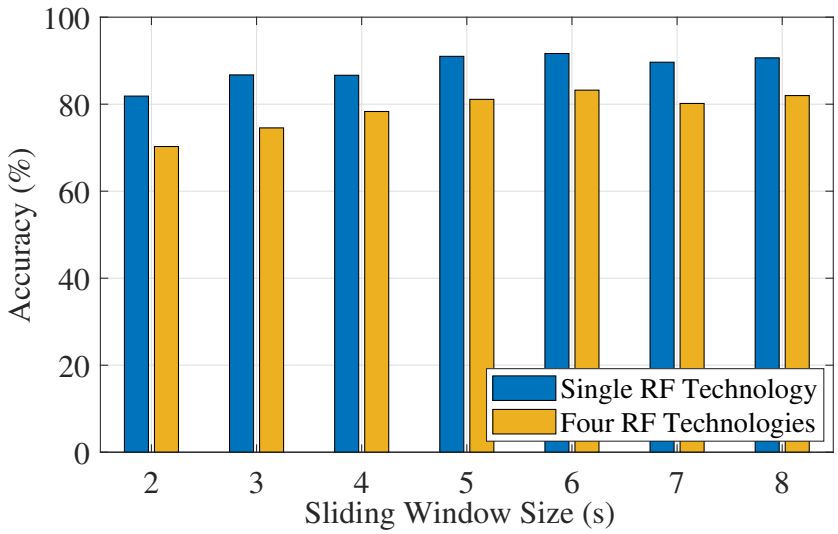


Figure 9.20: Activity recognition accuracy when different sliding window sizes are used.

the influence from various sensing technologies. The robustness demonstrated by these results validates that that the proposed STFT feature tensor can successfully extract the general characteristics of human behavior from diverse RF data collected by different RF technologies.

We also conduct experiments to explore appropriate parameter setting for the proposed TARF system. Fig. 9.19 and Fig. 9.20 show the impacts of the measurement number N_P and size of the sliding window, respectively. As Fig. 9.19 shows, the accuracy increases when more measurements are used for feature extraction. In the single-technology scenario, the accuracy is over 90% when $N_P \geq 25$, and the highest accuracy 91.62% is achieved when $N_P = 30$.

Similarly, In the four-technology scenario, the highest accuracy is achieved when 30 measurements are used for tensor generation. Fig. 9.20 shows the accuracy when different sliding window sizes are employed for STFT. The figure shows that in both scenarios, high accuracy is achieved when the sliding windows is 6 seconds. However, we notice that when the window size is 5 seconds, the accuracy in the two scenarios are 91.00% and 81.11%, respectively, which is sufficiently high for human activity recognition. Thus, to reduce the training complexity, we choose the smallest sliding window size which still achieves an acceptable system accuracy for STFT. As a result, we set N_P to 30 and the window size to 5 seconds. Although the STFT tensor requires 5 seconds of the RF data, the sliding window structure makes the system could perform activity recognition for each newly sampled RF data. Furthermore, since the classification of the trained DANN is processed very fast, the proposed TARF system is suitable for the realtime tracking of human activities.

9.7 Conclusion

In this paper, we proposed a generalized approach to human activity recognition, termed TARF, to mitigate the impact of technology-agnostic data acquisition. A novel signal preprocessing solution was proposed to combat the diversity caused by different RF sensing platforms.

The generalized tensor construction method was proposed to break the barriers of RF data collected from different RF technologies and extract the generalized features related to human activities. We then utilize a domain adversarial neural network to address the diversity issue of motion feature translation in different RF platforms. The experiments results demonstrate that the TARF system could be effectively implemented in various RF devices so that different RF technologies could complement each other. The technology-agnostic scheme could achieve robust HAR performance in different scenarios by incorporating all accessible RF technologies.

Chapter 10

Summary and Future Work

In my Ph.D. study, we propose several RF sensing techniques in Artificial Intelligence of Things, such as Indoor localization, Vital sign monitoring, human pose tracking, etc. For the RF based respiration monitoring systems, we propose effective RF vital sign monitoring systems in both static environments and more challenging dynamic environments. For the indoor localization system, we investigated the problem of localizing an RFID tag array. The proposed system was termed SparseTag, i.e., a sparse RFID tag array system for high accuracy backscatter indoor localization. The SparseTag system comprised four key components: (i) sparse array processing, (ii) difference co-array design, (iii) DOA estimation using a spatial smoothing method, and (iv) a DOA-based localization method. In addition, we prototype RFID-Pose, a vision-aided 3D human pose tracking system with RFID data. Based on the RFID-Pose, we extend the system by improving subject adaptability with a cross-skeleton learning methodology. We also combat Data Divergence in different environments with Meta-Learning algorithms.

10.1 Intelligent Disease Precaution System

With the human chest movement monitoring, the Phasebeat system can detect the apnea and other details of the respiration when the user is sleeping. These respiration details can reflect the health condition of the user, and certain apnea can represent the possibility of some special disease. The intelligent disease precaution system can record the sleeping data and evaluate the user health condition with learning algorithm.

10.2 Online Multiple Users Monitoring

With advanced signal processing technique, the system can also separate the breathing signals corresponding to different persons with blind signal separation. However, this kind of technique should need data from a long period of time, and long-time calculation is also necessary. With multiple antennas or other some other devices like RFID tags, the system can separate the different users breathing signals with low calculation time. We plan to extend the system with more devices, so that the system can monitor users respiration data in real-time.

10.3 Vital Signs Monitoring in Noisy Environment

Traditionally, RF signal based health monitoring system is very sensitive to the environment. It is very hard to do the vital signs monitoring when in driving environment, especially when the user is driving the vehicle. Large movement will generate large noise in the received signal, which is hard to be separated because the vital signal is usually small signal. Besides, vehicle vibration could also overwhelm the small vital sign signal. In the future, we will propose to make the system can resist the noise coming from user's large body movement. Since the breathing rate is also an indicator of human drowsiness, respiration monitoring in driving environment could be a effective way to detect the driving fatigue.

10.4 Data Augmentation for the vision-aided training supervision

In this dissertation, we have proposed the vision-aided approach to achieve the RFID based human pose estimation system. However, the data collection for the vision-aided supervised training is a extremely high cost process. This is because, the synchronized RF data and Kinect data sampling consumes a lot of time. In addition, the since we intend to improve high subject adaptability and environment adaptability for the system. The data should be collected in different environment and with different subject, which is considerable high cost on time. Thus, a data augmentation approach should be proposed to achieve an inexpensive data collection for the model training. With an effective training data augmentation approach, the network could be trained with a limited real sampled data and corresponding generated fake training data.

Appendices

Conference Publications

1. **C. Yang**, L. Wang, X. Wang, and S. Mao, "Meta-Pose: Environment-adaptive human skeleton tracking with RFID," in *Proc. IEEE GLOBECOM 2021*, Madrid, Spain, Dec. 2021.

2. X. Wang, M. Patil, **C. Yang**, S. Mao, and P.A. Patel, "Deep Convolutional Gaussian Processes for mmWave outdoor localization," in *Proc. IEEE ICASSP 2021*, Toronto, Canada, June 2021, pp.8323–8327.

3. J. Purohit, X. Wang, S. Mao, X. Sun, and **C. Yang**, "Fingerprinting-based indoor and outdoor localization with LoRa and deep learning," in Proc. IEEE GLOBECOM 2020, Taipei, Taiwan, Dec. 2020.

4. **C. Yang**, X. Wang, and S. Mao, "Demo Abstract: Vision-aided 3D human pose estimation with RFID," in *Proc. The 16th IEEE International Conference on Mobility, Sensing and Networking (MSN 2020)*, Tokyo, Japan, Dec. 2020, pp.628–629.

5. **C. Yang**, X. Wang, and S. Mao, "Subject-adaptive skeleton tracking with RFID," in *Proc. The 16th IEEE International Conference on Mobility, Sensing and Networking (MSN 2020)*, Tokyo, Japan, Dec. 2020, pp.599–606.

6. **C. Yang**, X. Wang, and S. Mao, "RFID-based driving fatigue detection," in *Proc. IEEE GLOBECOM 2019*, Waiko-loa, HI, Dec. 2019.

7. **C. Yang**, X. Wang, and S. Mao, "SparseTag: High-precision backscatter indoor localization with sparse RFID tag arrays," in *Proc. IEEE SECON 2019*, Boston, MA, June 2019, pp.1–9.

8. **C. Yang**, X. Wang, and S. Mao, "AutoTag: Recurrent variational autoencoder for unsupervised apnea detection with RFID tags," in *Proc. IEEE GLOBECOM 2018*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp.1–7.

9. X. Wang, **C. Yang**, and S. Mao, "ResBeat: Resilient breathing beats monitoring with online bimodal CSI data," in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017.

10. X. Wang, **C. Yang**, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp.1230–1239.

11. **C. Yang**, X. Wang, and S. Mao, "RFID-based vital sign monitoring," invited paper in *Proc. The 2021 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI'21)*, Virtual Conference, July, 2021.

12. J. Ma, **C. Yang**, S. Mao, J. Zhang, S. CG Periaswamy, and J. Patton, "Human Trajectory Completion with Transformers," in *Proc. IEEE ICC 2022*, accepted.

Journal Publications

1. **C. Yang**, X. Wang, and S. Mao, "Environment Adaptive RFID based 3D Human Pose Tracking with a Meta-learning Approach," *IEEE Journal of Radio Frequency Identification*, early access.

2. **C. Yang**, X. Wang, and S. Mao, "RFID Tag Localization with a Sparse Tag Array," *IEEE Internet of Things Journal*, early access.

3. **C. Yang**, X. Wang, and S. Mao, "RFID based 3D human pose tracking: A subject generalization approach," *Elsevier/KeAi Digital Communications and Networks*, Sept. 2021. DOI: XXX.

4. X. Wang, R. Huang, **C. Yang**, and S. Mao, "Smartphone sonar based contact-free respiration rate monitoring," *ACM Transactions on Computing for Healthcare*, vol.2, no.2, Article 15, Mar. 2021. DOI: 10.1145/3436822.

5. **C. Yang**, X. Wang, and S. Mao, "RFID-Pose: Vision-aided 3D human pose estimation with RFID," *IEEE Transactions on Reliability*, vol.70, no.3, pp.1218–1231, Sept. 2021. DOI: 10.1109/TR.2020.3030952.

6. **C. Yang**, X. Wang, and S. Mao, "Unsupervised drowsy driving detection with RFID," *IEEE Transactions on Vehicular Technology*, vol.69, no.8, pp. 8151–8163, Aug. 2020. DOI: 10.1109/TVT.2020.2995835.

7. X. Wang, **C. Yang**, and S. Mao, "Resilient respiration rate monitoring with realtime bimodal CSI data," *IEEE Sensors Journal*, vol.20, no.17, pp.10187–10198, Sept. 2020. DOI: 10.1109/JSEN.2020.2989780.

8. **C. Yang**, X. Wang, and S. Mao, "Respiration monitoring with RFID in driving environments," *IEEE Journal on Selected Areas in Communications*, vol.39, no.2, pp.500–512, Feb. 2021. DOI: 10.1109/JSAC.2020.3020606.

9. X. Wang, **C. Yang**, and S. Mao, "On CSI-based vital sign monitoring using commodity WiFi," *ACM Transactions on Computing for Healthcare*, vol.1, no.3, pp.12:1–12:27, Apr. 2020. DOI: 10.1145/3377165.

10. X. Wang, **C. Yang**, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi," *ACM Transactions on Intelligent Systems and Technology*, vol.9, no.1, Article 8, pp.8:1–8:27, Sept. 2017. DOI: 10.1145/3078855.

11. **C. Yang**, S. Mao, and X. Wang, "An overview of 3GPP positioning standards," *ACM GetMobile*, vol.26, no.1, pp.9–13, Mar. 2022.

12. **C. Yang**, X. Wang, and S. Mao, "TARF: Technology-agnostic RF Sensing for Human Activity Recognition" *IEEE Journal of Biomedical and Health Informatics*, accepted.

References

[1] C. Yang, S. Mao, and X. Wang, "An overview of 3GPP positioning standards," *ACM GetMobile*, vol.26, no.1, pp.9–13, Mar. 2022.

[2] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao,, "Contour stella image and deep learning for signal recognition in the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol.7, no.1, pp.34–46, Mar. 2021.

[3] K. Wang, W. Zhou, and S. Mao, "On joint BBU/RRH resource allocation in heterogeneous Cloud-RANs," *IEEE Internet of Things Journal*, vol.4, no.3, pp.749–759, June 2017.

[4] M. Feng, S. Mao, and T. Jiang, "BOOST: Base station on-off switching strategy for energy efficient massive MIMO HetNets," in *Proc. IEEE INFOCOM 2016*, San Francisco, CA, Apr. 2016, pp.1395–1403.

[5] H. Zhou, S. Mao, and P. Agrawal, "Optical power allocation for adaptive transmissions in wavelength-division multiplexing free space optical networks," *Elsevier Digital Communications and Networks Journal*, vol.1, no.3, pp.171–180, Aug. 2015.

[6] M. Feng, S. Mao, and T. Jiang, "Joint duplex mode selection, channel allocation, and power control for full-duplex cognitive femtocell networks," *Elsevier Digital Communications and Networks Journal*, vol.1, no.1, pp.30–44, Feb. 2015.

[7] X. Wang, S. Mao, and M.X. Gong, "A survey of LTE Wi-Fi coexistence in unlicensed bands," *ACM GetMobile: Mobile Computing and Communications Review*, vol.20, no.3, pp.17–23, July 2016.

[8] C. Yang, X. Wang, and S. Mao, "AutoTag: Recurrent vibrational autoencoder for unsupervised apnea detection with RFID tags," in *Proc. IEEE GLOBECOM 2018*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–7.

[9] O. Boric-Lubeke and V. Lubecke, "Wireless house calls: Using communications technology for health care and monitoring," *IEEE Microwave Mag.*, vol. 3, no. 3, pp. 43–48, Apr. 2002.

[10] X. Wang, X. Wang, and S. Mao, "RF sensing for Internet of Things: A general deep learning framework," *IEEE Communications*, vol. 56, no. 9, pp. 62–69, Sept. 2018.

[11] X. Wang, R. Huang, and S. Mao, "Sonarbeat: Sonar phase for breathing beat monitoring with smartphones," in *Proc. ICCCN 2017*, Vancouver, Canada, July/Aug. 2017, pp. 1–8.

[12] C. Hunt and F. Hauck, "Sudden infant death syndrome," *Can. Med. Assoc. J*, vol. 174, no. 13, pp. 1309–1310, Apr. 2006.

[13] M. L. R. Mogue and B. Rantala, "Capnometers," *Journal of Clinical Monitoring*, vol. 4, no. 2, pp. 115–121, Apr. 1988.

[14] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. ACM CHI'15*, Seoul, Korea, April 2015, pp. 837–846.

[15] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and fine-grained breathing volume monitoring from afar using wireless signals," in *Proc. IEEE INFOCOM'16*, San Francisco, CA, Apr. 2016, pp. 1–9.

[16] H. Abdelnasser, K. A. Harras, and M. Youssef, "Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator," in *Proc. IEEE MobiHoc'15*, Hangzhou, China, June 2015, pp. 277–286.

[17] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proc. ACM Mobihoc'15*, Hangzhou, China, June 2015, pp. 267–276.

[18] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp. 1–10.

[19] ——, "Tensorbeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi," *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 1, pp. 8:1–8:27, Sept. 2017.

[20] ——, "ResBeat: Resilient breathing beats monitoring with online bimodal CSI data," in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017, pp. 1–6.

[21] J. Wang and D. Katabi, "Dude, where's my card? RFID positioning that works with multipath and non-line of sight," in *ACM SIGCOMM Comput. Commun.Rev.*, vol. 43, no. 4, Oct. 2013, pp. 51–62.

[22] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proc. ACM MobiCom'14*, Maui, HI, Sept. 2014, pp. 237–248.

[23] T. Wei and X. Zhang, "Gyro in the air: Tracking 3D orientation of batteryless internet-of-things," *ACM GetMobile*, vol. 21, no. 1, pp. 35–38, Mar. 2017.

[24] Y. Ma, N. Selby, and F. Adib, "Drone relays for battery-free networks," in *Proc. ACM SIGCOMM 2017*, Los Angeles, CA, Aug. 2017, pp. 335–347.

[25] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S. C. Periaswamy, J. Patton, and X. Wang, "RFHUI: An intuitive and easy-to-operate human-uav interaction system for controlling a UAV in a 3D space," in *Proc. EAI MobiQuitous 2018*, New York City, NY, Nov. 2018, pp. 69–76.

[26] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S. Periaswamy, J. Patton, and X. Wang, "RFHUI: An RFID based human-unmanned aerial vehicle interaction system in an indoor environment," *Elsevier Digital Communica-tions and Networks Journal*, to appear.

[27] Y. Hou, Y. Wang, and Y. Zheng, "Tagbreathe: Monitor breathing with commodity RFID systems," in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp. 404–413.

[28] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," in *Proc. NIPS 2015*, Montreal, Canada, Dec. 2015, pp. 2980–2988.

[29] I. Habibie *et al.*, "A recurrent variational autoencoder for human motion synthesis," in *Proc. British Machine Vision Conference 2017*, London, UK, Sept. 2017, pp. 1–12.

[30] J. Salmi and A. F. Molisch, "Propagation parameter estimation, modeling and measurements for ultrawideband mimo radar," *IEEE Trans. Microw. Theory Technol.*, vol. 59, no. 11, pp. 4257–4267, Nov. 2011.

[31] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," in *Proc. IEEE PerCom 2003*, Dallas, TX, Mar. 2003, pp. 407–415.

[32] H. Ding, J. Han, C. Qian, F. Xiao, G. Wang, N. Yang, W. Xi, and J. Xiao, "Trio: Utilizing tag interference for refined localization of passive RFID," in *Proc. IEEE INFOCOM 2018*, Honolulu, HI, Apr. 2018, pp. 828–836.

[33] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," in *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, Oct. 2014, pp. 235–246.

[34] L. Shangguan and K. Jamieson, "The design and implementation of a mobile RFID tag sorting robot," in *Proc. ACM MobiSys'16*, Singapore, June 2016, pp. 31–42.

[35] T. Liu, L. Yang, Q. Lin, Y. Guo, and Y. Liu, "Anchor-free backscatter positioning for RFID tags with high accuracy," in *Proc. IEEE INFOCOM'14*, Toronto, Canada, Apr./May 2014, pp. 379–387.

[36] X. Wang, J. Zhang, Z. Yu, E. Mao, S. Periaswamy, and J. Patton, "RF Thermometer: A temperature estimation method with commercial UHF RFID tags," in *Proc. IEEE ICC 2019*, Shanghai, China, May 2019, pp. 1–6.

[37] C. Wang, J. Liu, Y. Chen, H. Liu, L. Xie, W. Wang, B. He, and S. Lu, "Multi-touch in the air: Device-free finger tracking and gesture recognition via COTS RFID," in *Proc. IEEE INFOCOM 2018*, Honolulu, HI, Apr. 2018, pp. 1691–1699.

[38] Impinj Support Portal, "Low level user data support," 2013, Impinj Speedway Revolution Reader Application, [online] Available: https://support.impinj.com.

[39] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, May 2014.

[40] O. Fabius and J. R. van Amersfoort, "Variational recurrent auto-encoders," *arXiv preprint arXiv:1412.6581*, June 2015.

[41] L. Almazaydeh, K. Elleithy, M. Faezipour, and A. Abushakra, "Apnea detection based on respiratory signal classification," *Elsevier Procedia Computer Science*, vol. 21, pp. 310–316, 2013.

[42] W. Yang, E. Shen, X. Wang, S. Mao, Y. Gong, and P. Hu, "Wi-Wheat+: Contact-free wheat moisture sensing with commodity WiFi based on entropy," *Elsevier/KeAi Digital Communications and Networks*, to appear.

[43] National Road Safety Foundation, "Drowsy driving," 2018 (accessed April 24, 2018). [Online]. Available: https://www.ghsa.org/issues/drowsy-driving

[44] J. Wang, J. Liu, and N. Kato, "Networking and communications in autonomous driving: A survey," *IEEE Commun. Sur. & Tut.*, vol. 21, no. 2, pp. 1243–1274, June 2019.

[45] J. Hu, L. Xu, X. He, and W. Meng, "Abnormal driving detection based on normalized driving behavior," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6645–6652, Aug. 2017.

[46] A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel, "Recent trends in driver safety monitoring systems: State of the art and challenges," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4550–4563, June 2017.

[47] Y. Xun, J. Liu, N. Kato, Y. Fang, and Y. Zhang, "Automobile driver fingerprinting: A new machine learning based authentication scheme," *IEEE Trans. Ind. Inform.*, vol. 16, no. 2, pp. 1417–1426, Feb. 2020.

[48] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Elsevier Expert Systems with Applications*, vol. 36, no. 2, pp. 2352–2359, Mar. 2009.

[49] W.-B. Horng, C.-Y. Chen, Y. Chang, and C.-H. Fan, "Driver fatigue detection based on eye tracking and dynamic template matching," in *Proc. IEEE ICNSC'04*, Taipei, Taiwan, Mar. 2004, pp. 7–12.

[50] W. Jia and H. Peng, "Wifind: Driver fatigue detection with fine-grained Wi-Fi signal features," in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017, pp. 1–6.

[51] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "D3-Guard: Acoustic-based drowsy driving detection using smartphones," in *Proc. IEEE INFOCOM'19*, Paris, France, Apr./May 2019, pp. 1–9.

[52] X. Wang, J. Zhang, Z. Yu, S. Mao, S. Periaswamy, and J. Patton, "On remote temperature sensing using commercial UHF RFID tags," *IEEE Internet of Things J.*, vol. 6, no. 6, pp. 10 715–10 727, Dec. 2019.

[53] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S. C. Periaswamy, J. Patton, and X. Wang, "RFHUI: An intuitive and easy-to-operate human-uav interaction system for controlling a UAV in a 3D space," in *Proc. EAI MobiQuitous'18*, New York City, NY, Nov. 2018, pp. 69–76.

[54] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S. Periaswamy, J. Patton, and X. Wang, "RFHUI: An RFID based human-unmanned aerial vehicle interaction system in an indoor environment," *KeAi Digital Communications and Networks Journal*, vol. 6, no. 1, pp. 14–22, Feb. 2020.

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

[56] L. Yao, Q. Sheng, W. Ruan, T. Gu, X. Li, N. Falkner, and Z. Yang, "RF-care: Device-free posture recognition for elderly people using a passive RFID tag array," in *Proc. EAI MobiQuitous'15*, Coimbra, Portugal, July 2015, pp. 120–129.

[57] C. Yang, X. Wang, , and S. Mao, "SparseTag: High-precision backscatter indoor localization with sparse RFID tag arrays," in *Proc. IEEE SECON'19*, Boston, MA, June 2019, pp. 1–9.

[58] C. Yang, X. Wang, and S. Mao, "Unsupervised detection of apnea using commodity RFID tags with a recurrent variational autoencoder," *IEEE Access J.*, vol. 7, no. 1, pp. 67 526–67 538, June 2019.

[59] J. Zhou, J. Shi, and X. Qu, "Landmark placement for wireless localization in rectangular-shaped industrial facilities," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3081–3090, July 2010.

[60] T. Jing, et al., "An efficient scheme for tag information update in RFID systems on roads," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2435–2444, Apr. 2016.

[61] H. Qin, Y. Peng, and W. Zhang, "Vehicles on RFID: Error-cognitive vehicle localization in GPS-less environments," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9943–9957, Nov. 2017.

[62] C. Yang, X. Wang, and S. Mao, "RFID-based driving fatigue detection," in *Proc. IEEE GLOBECOM 2019*, Waiko-loa, HI, Dec. 2019, pp. 1–6.

[63] Y.-F. Zhang, X.-Y. Gao, J.-Y. Zhu, W.-L. Zheng, and B.-L. Lu, "A novel approach to driving fatigue detection using forehead EOG," in *Proc. 2015 Int. IEEE/EMBS Conf. Neural Engineering (NER)*, Montpellier, France, Apr. 2015, pp. 707–710.

[64] L. Li, M. Xie, and H. Dong, "A method of driving fatigue detection based on eye location," in *Proc. IEEE 3rd Int. Conf. Commun. Software Netw.*, Xi'an, China, May 2011, pp. 480–484.

[65] B. Warwick, N. Symons, X. Chen, and K. Xiong, "Detecting driver drowsiness using wireless wearables," in *Proc. IEEE MASS'15*, Dallas,TX, Oct. 2015, pp. 585–588.

[66] Z. Yang, M. Bocca, V. Jain, and P. Mohapatra, "Contactless Breathing Rate Monitoring in Vehicle Using UWB Radar." in *Proc. RealWSN Workshop*, Shenzhen, China, Nov. 2018, pp. 13–18.

[67] J. He, S. Roberson, B. Fields, J. Peng, S. Cielocha, and J. Coltea, "Fatigue detection using smartphones," *Journal of Ergonomics*, vol. 3, no. 3, pp. 1–7, Jan. 2013.

[68] C. Zhao, Z. Li, T. Liu, H. Ding, J. Han, W. Xi, and R. Gui, "RF-Mehndi: A Fingertip Profiled RF Identifier," in *Proc. IEEE INFOCOM'19*, Paris, France, June 2019, pp. 1513–1521.

[69] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, "TagScan: Simultaneous target imaging and material identification with commodity RFID devices," in *Proc. ACM MobiCom'17*, Snowbird, Utah, Oct. 2017, pp. 288–300.

[70] P. Li, Z. An, L. Yang, and P. Yang, "Towards physical-layer vibration sensing with rfids," in *Proc. IEEE INFOCOM'19*, Paris, France, June 2019, pp. 892–900.

[71] J. Guo, T. Wang, Y. He, M. Jin, C. Jiang, and Y. Liu, "Twinleak: Rfid-based liquid leakage detection in industrial environments," in *Proc. IEEE INFOCOM'19*, Paris, France, Apr. 2019, pp. 883–891.

[72] S. Pradhan, E. Chai, K. Sundaresan, L. Qiu, M. A. Khojastepour, and S. Rangarajan, "Rio: A pervasive rfid-based touch gesture interface," in *Proc. ACM MobiCom'17*, Snowbird, Utah, Oct. 2017, pp. 261–274.

[73] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRFID: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2016.

[74] J. Liu, X. Chen, S. Chen, X. Liu, Y. Wang, and L. Chen, "TagSheet: Sleeping Posture Recognition with an unobtrusive Passive Tag Matrix," in *Proc. IEEE INFOCOM'19*, Paris, France, Apr. 2019, pp. 874–882.

[75] National Highway Traffic Safety Administration, "U.S. DOT Announces 2017 Roadway Fatalities Down," 2018 (accessed April 24, 2018). [Online]. Available: https://www.nhtsa.gov/press-releases/us-dot-announces-2017-roadway-fatalities-down

[76] Driver Knowledge, "Driving Facts," 2019 (accessed April 25, 2016). [Online]. Available: https://www.driverknowledge.com/car-accident-statistics/

[77] J. Solaz, J. Laparra-Hernández, D. Bande, N. Rodríguez, S. Veleff, J. Gerpe, and E. Medina, "Drowsiness detection based on the analysis of breathing rate obtained from real-time image recognition," *Transportation Research Procedia*, vol. 14, pp. 3867–3876, Apr. 2016.

[78] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, "BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals," in *Proc. ACM MobiSys 2019*, Seoul, Republic of Korea, June 2019, pp. 54–66.

[79] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010, [online] Available: https://arxiv.org/abs/1009.5055.

[80] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. ACM CHI 2015*, Seoul, Republic of Korea, Apr. 2015, pp. 837–846.

[81] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and fine-grained breathing volume monitoring from afar using wireless signals," in *Proc. IEEE INFOCOM 2016*, San Francisco, CA, Apr. 2016, pp. 1–9.

[82] H. Abdelnasser, K. A. Harras, and M. Youssef, "Ubibreathe: A ubiquitous non-invasive WiFi-based breathing estimator," in *Proc. IEEE MobiHoc'15*, Hangzhou, China, June 2015, pp. 277–286.

[83] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital sign and sleep monitoring using millimeter wave," *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 2, pp. 1–32, June 2017.

[84] S. Shi, Y. Xie, M. Li, A. X. Liu, and J. Zhao, "Synthesizing Wider WiFi Bandwidth for Respiration Rate Monitoring in Dynamic Environments," in *Proc. IEEE INFOCOM 2019*, Paris, France, Apr. 2019, pp. 874–882.

[85] C. Chen, Y. Han, Y. Chen, H.-Q. Lai, F. Zhang, B. Wang, and K. R. Liu, "TR-BREATH: Time-reversal breathing rate estimation and detection," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 489–501, Apr. 2017.

[86] X. Wang, C. Yang, and S. Mao, "ResBeat: Resilient breathing beats monitoring with online bimodal CSI data," in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017, pp. 1–6.

[87] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "FullBreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–19, Sept. 2018.

[88] R. Zhao, D. Wang, Q. Zhang, H. Chen, and A. Huang, "CRH: A Contactless Respiration and Heartbeat Monitoring System with COTS RFID Tags," in *Proc. IEEE SECON 2018*, Hong Kong, China, June 2018, pp. 1–9.

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

[90] J. Zhang, X. Wang, Z. Yu, Y. Lyu, S. Mao, S. Periaswamy, J. Patton, and X. Wang, "Robust RFID based 6-DoF localization for unmanned aerial vehicles," *IEEE Access Journal*, vol. 7, no. 1, pp. 77 348–77 361, June 2019.

[91] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultra-wideband localization for deployed RFID tags," in *Proc. ACM MobiCom 2017*, Snowbird, Utah, Oct. 2017, pp. 248–260.

[92] C. Duan, L. Yang, H. Jia, Q. Lin, Y. Liu, and L. Xie, "Robust spinning sensing with dual-RFID-tags in noisy settings," in *Proc. IEEE INFOCOM 2018*, Honolulu, HI, Apr. 2018.

[93] L. Yang, Y. Li, Q. Lin, X.-Y. Li, and Y. Liu, "Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals," in *Proc. ACM MobiCom 2016*, New York, NY, Oct. 2016, pp. 16–28.

[94] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, Jan. 2012.

[95] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, Feb. 2015.

[96] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[97] X. Wang, J. Zhang, Z. Yu, E. Mao, S. Periaswamy, and J. Patton, "RF Thermometer: A temperature estimation method with commercial UHF RFID tags," in *Proc. IEEE ICC 2019*, Shanghai, China, May 2019, pp. 1–6.

[98] Y. Zhang, M. G. Amin, and S. Kaushik, "Localization and tracking of passive rfid tags based on direction estimation," *International Journal of Antennas and Propagation*, vol. 2007, 2007.

[99] H. Jin, Z. Yang, S. Kumar, and J. I. Hong, "Towards wearable everyday body-frame tracking using passive RFIDs," *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technol.*, vol. 1, no. 4, p. No. 145, Dec. 2018.

[100] Y. Bu, L. Xie, J. Liu, B. He, Y. Gong, and S. Lu, "3-Dimensional reconstruction on tagged packages via RFID systems," in *Proc. IEEE SECON'17*, San Diego, CA, June 2017, pp. 1–9.

[101] Y. Zhang, L. Xie, Y. Bu, Y. Wang, J. Wu, and S. Lu, "3-Dimensional localization via RFID tag array," in *Proc. IEEE MASS'17*, Orlando, FL, Oct. 2017, pp. 353–361.

[102] F. Guidi, N. Decarli, S. Bartoletti, A. Conti, and D. Dardari, "Detection of multiple tags based on impulsive backscattered signals," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3918–3930, 2014.

[103] M. Bolic, M. Rostamian, and P. M. Djuric, "Proximity detection with RFID: A step toward the internet of things," *IEEE Pervasive Computing*, vol. 14, no. 2, pp. 70–76, Apr. 2015.

[104] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: do user location and body orientation matter?" in *Proc. ACM Ubicomp 2016*, Heidelberg, Germany, Sept. 2016, pp. 25–36.

[105] H. Ding, J. Han, C. Qian, F. Xiao, G. Wang, N. Yang, W. Xi, and J. Xiao, "Trio: Utilizing tag interference for refined localization of passive RFID," in *Proc. IEEE INFOCOM'18*, Honolulu, HI, Apr. 2018.

[106] F. Xiao, Z. Wang, N. Ye, R. Wang, and X.-Y. Li, "One more tag enables fine-grained RFID localization and tracking," *IEEE/ACM Trans. Netw.*, vol. 26, no. 1, pp. 161–174, Jan. 2018.

[107] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[108] D. M. Dobkin, *The rf in RFID: UHF RFID in practice*. Newnes, 2012.

[109] F. Lu, X. Chen, and T. Y. Terry, "Performance analysis of stacked RFID tags," in *Proc. 2009 IEEE International Conference on RFID*, Orlando, FL, Apr. 2009, pp. 330–337.

[110] Y. Tanaka, Y. Umeda, O. Takyu, M. Nakayama, and K. Kodama, "Change of read range for uhf passive rfid tags in close proximity," in *RFID, 2009 IEEE International Conference on*. IEEE, 2009, pp. 338–345.

[111] P. Pal and P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[112] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[113] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[114] C.-L. Liu and P. Vaidyanathan, "Super nested arrays: Linear sparse arrays with reduced mutual coupling–Part I: Fundamentals," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3997–4012, Aug. 2016.

[115] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. ACM MobiSys 2013*, Taipei, Taiwan, June 2013, pp. 249–262.

[116] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 13)*, Lombard, IL, Apr. 2013, pp. 71–84.

[117] W. Gong and J. Liu, "Robust indoor wireless localization using sparse recovery," in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp. 847–856.

[118] P. Bahl and V. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proc. IEEE INFOCOM 2000*, Tel Aviv, Israel, Apr. 2000, pp. 775–784.

[119] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for Indoor Localization with a Deep Learning Approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.

[120] M. Youssef and A. Agrawala, "The Horus TWLAN location determination system," in *Proc. ACM MobiSys 2005*, Seattle, WA, June 2005, pp. 205–218.

[121] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE CVPR 2017*, Honolulu, HI, July 2017, pp. 7291–7299.

[122] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," in *Proc. IEEE CVPR 2010*, San Francisco, CA, June 2010, pp. 623–630.

[123] Tom's Guide, "Millions of wireless security cameras are at risk of being hacked: What to do," 2020 (accessed June 20, 2020). [Online]. Available: https://www.tomsguide.com/news/hackable-security-cameras

[124] P. A. Laplante and J. F. DeFranco, "Software engineering of safety-critical systems: Themes from practitioners," *IEEE Trans. Rel.*, vol. 66, no. 3, pp. 825–836, Sept. 2017.

[125] S. Siboni, V. Sachidananda, Y. Meidan, M. Bohadana, Y. Mathov, S. Bhairav, A. Shabtai, and Y. Elovici, "Security testbed for Iternet-of-Things devices," *IEEE Trans. Rel.*, vol. 68, no. 1, pp. 23–44, Mar. 2019.

[126] M. Noor-A-Rahim, M. Khyam, G. M. N. Ali, Z. Liu, D. Pesch, and P. H. Chong, "Reliable state estimation of an unmanned aerial vehicle over a distributed wireless IoT network," *IEEE Trans. Rel.*, vol. 68, no. 3, pp. 1061–1069, Sept. 2019.

[127] S. Wang and X. Yao, "Using class imbalance learning for software defect prediction," *IEEE Trans. Rel.*, vol. 62, no. 2, pp. 434–443, June 2013.

[128] X. Yang, K. Tang, and X. Yao, "A learning-to-rank approach to software defect prediction," *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 234–246, Mar. 2014.

[129] M. Liu, L. Miao, and D. Zhang, "Two-stage cost-sensitive learning for software defect prediction," *IEEE Trans. Rel.*, vol. 63, no. 2, pp. 676–686, June 2014.

[130] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proc. IEEE ICCV 2019*, Seoul, Republic of Korea, Oct. 2019, pp. 5452–5461.

[131] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, and L. Su, "Towards 3D human pose construction using WiFi," in *Proc. ACM MobiCom'20*, London, UK, Sept. 2020, pp. 1–14.

[132] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proc. IEEE CVPR 2018*, Salt Lake City, UT, June 2018, pp. 7356–7365.

[133] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs," *IEEE Sensors J.*, vol. 20, no. Sept., pp. 10 032–10 044, 17 2020.

[134] C. Wang, J. Liu, Y. Chen, L. Xie, H. B. Liu, and S. Lu, "RF-Kinect: A wearable RFID-based approach towards 3D body movement tracking," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–28, Mar. 2018.

[135] H. Jin, Z. Yang, S. Kumar, and J. I. Hong, "Towards wearable everyday body-frame tracking using passive RFIDs," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–23, Dec. 2018.

[136] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Elsevier Computer Vision and Image Understanding*, vol. 192, no. 3, p. 102897, Mar. 2020.

[137] J. Zhang, S. Periaswamy, S. Mao, and J. Patton, "Standards for passive UHF RFID," *ACM GetMobile*, vol. 23, no. 3, pp. 10–15, Sept. 2019.

[138] C. Yang, X. Wang, and S. Mao, "Unsupervised drowsy driving detection with RFID," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8151–8163, Aug. 2020.

[139] C. Yang, X. Wang, and S. Mao, "Respiration monitoring with RFID in driving environments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, Feb. 2021.

[140] R. Mitra, N. B. Gundavarapu, A. Sharma, and A. Jain, "Multiview-consistent semi-supervised learning for 3D human pose estimation," in *Proc. IEEE CVPR 2020*, Seattle, WA, June 2020, pp. 6907–6916.

[141] X. Fan, K. Zheng, Y. Lin, and S. Wang, "Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation," in *Proc. IEEE CVPR 2015*, Boston, MA, June 2015, pp. 1347–1355.

[142] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[143] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Springer Int. J. Computer Vision*, vol. 87, no. 1/2, pp. 1–24, July 2010.

[144] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "RF-based 3D skeletons," in *Proc. ACM SIGCOM 2018*, Budapest, Hungary, Aug. 2018, pp. 267–281.

[145] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2012.

[146] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Oct. 2013, arXiv preprint arXiv:1009.5055. [Online]. Available: https://arxiv.org/abs/1009.5055

[147] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proc. IEEE CVPR 2018*, Salt Lake City, UT, June 2018, pp. 8639–8648.

[148] C. Yang, X. Wang, S. Mao, Subject-adaptive skeleton tracking with RFID, in: Proc. IEEE MSN 2020, Tokyo, Japan, 2020, pp. 1–8.

[149] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, in: Proc. IEEE CVPR 2017, Honolulu, HI, 2017, pp. 7291–7299.

[150] C. Yang, X. Wang, S. Mao, Demo Abstract: Vision-aided 3D human pose estimation with RFID, in: Proc. IEEE MSN 2020, Tokyo, Japan, 2020, pp. 1–2.

[151] M. Andriluka, S. Roth, B. Schiele, Monocular 3D pose estimation and tracking by detection, in: Proc. IEEE CVPR 2010, San Francisco, CA, 2010, pp. 623–630.

[152] Tom's Guide, Millions of wireless security cameras are at risk of being hacked: What to do, https://www.tomsguide.com/news/hackable-security-cameras (2020 (accessed Aug. 28, 2020)).

[153] F. Wang, S. Zhou, S. Panev, J. Han, D. Huang, Person-in-WiFi: Fine-grained person perception using WiFi, in: Proc. IEEE ICCV 2019, Seoul, Republic of Korea, 2019, pp. 5452–5461.

[154] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, D. Katabi, Through-wall human pose estimation using radio signals, in: Proc. IEEE CVPR 2018, Salt Lake City, UT, 2018, pp. 7356–7365.

[155] A. Sengupta, F. Jin, R. Zhang, S. Cao, mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs, IEEE Sensors J. 20 (17) (2020) 10032–10044.

[156] J. Zhang, S. Periaswamy, S. Mao, J. Patton, Standards for passive UHF RFID, ACM GetMobile 23 (3) (2019) 10–15.

[157] C. Yang, X. Wang, S. Mao, RFID-Pose: Vision-aided 3D human pose estimation with RFID, IEEE Transactions on Reliability. In press. DOI: 10.1109/TR.2020.3030952.

[158] M. A. A. da Cruz, J. J. P. C. Rodrigues, P. Lorenz, V. Korotaev, V. H. C. de Albuquerque, In.iot—a new middleware for internet of things, IEEE Internet of Things Journal (2020) 1–10

[159] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (10) (2009) 1345–1359.

[160] R. Raina, A. Battle, H. Lee, B. Packer, A. Y. Ng, Self-taught learning: Transfer learning from unlabeled data, in: Proc. ACM ICML 2007, Corvalis, OR, 2007, pp. 759–766.

[161] W. Jiang, et al., Towards environment independent device free human activity recognition, in: Proc. ACM MobiCom 2018, New Delhi, India, 2018, pp. 289–304.

[162] Y. Chen, Y. Tian, M. He, Monocular human pose estimation: A survey of deep learning-based methods, Elsevier Computer Vision and Image Understanding 192 (3) (2020) 102897.

[163] R. Mitra, N. B. Gundavarapu, A. Sharma, A. Jain, Multiview-consistent semi-supervised learning for 3D human pose estimation, in: Proc. IEEE CVPR 2020, Seattle, WA, 2020, pp. 6907–6916.

[164] X. Fan, K. Zheng, Y. Lin, S. Wang, Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation, in: Proc. IEEE CVPR 2015, Boston, MA, 2015, pp. 1347–1355.

[165] Z. Zhang, Microsoft Kinect sensor and its effect, IEEE Multimedia 19 (2) (2012) 4–10.

[166] L. Sigal, A. O. Balan, M. J. Black, Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, Springer Int. J. Computer Vision 87 (1/2) (2010) 1–24.

[167] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, A. Torralba, RF-based 3D skeletons, in: Proc. ACM SIGCOM 2018, Budapest, Hungary, 2018, pp. 267–281.

[168] J. Liu, P. Musialski, P. Wonka, J. Ye, Tensor completion for estimating missing values in visual data, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2012) 208–220.

[169] R. Villegas, J. Yang, D. Ceylan, H. Lee, Neural kinematic networks for unsupervised motion retargetting, in: Proc. IEEE CVPR 2018, Salt Lake City, UT, 2018, pp. 8639–8648.

[170] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: ICCV 2017, Venice, Italy, 2017, pp. 2223–2232.

[171] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, A. Zisserman, Temporal cycle-consistency learning, in: Proc. IEEE CVPR 2019, Long Beach, CA, 2019, pp. 1801–1810.

[172] C. Yang, L. Wang, X. Wang, and S. Mao, "Meta-Pose: Environment-adaptive human skele-ton tracking with RFID," in *Proc. IEEE GLOBECOM 2021*, Madrid, Spain, Dec. 2021, pp. 1–6.

[173] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE CVPR 2017*, Honolulu, HI, July 2017, pp. 7291–7299.

[174] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," in *Proc. IEEE CVPR 2010*, San Francisco, CA, June 2010, pp. 623–630.

[175] Tom's Guide, "Millions of wireless security cameras are at risk of being hacked: What to do," 2020 (accessed Aug. 28, 2020). [Online]. Available: https://www.tomsguide.com/news/hackable-security-cameras

[176] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proc. IEEE CVPR 2018*, Salt Lake City, UT, June 2018, pp. 7356–7365.

[177] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10 032–10 044, Sept. 2020.

[178] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proc. IEEE ICCV 2019*, Seoul, Republic of Korea, Oct. 2019, pp. 5452–5461.

[179] J. Zhang, S. Periaswamy, S. Mao, and J. Patton, "Standards for passive UHF RFID," *ACM GetMobile*, vol. 23, no. 3, pp. 10–15, Sept. 2019.

[180] ——, "Subject-adaptive skeleton tracking with RFID," in *Proc. IEEE MSN 2020*, Tokyo, Japan, Dec. 2020.

[181] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE ICCV 2017*, Venice, Italy, Oct. 2017, pp. 2223–2232.

[182] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. ACM MobiCom 2018*, New Delhi, India, Sept. 2018, pp. 289–304.

[183] L. Wang, S. Mao, B. Wilamowski, and R. Nelms, "Pre-trained models for non-intrusive appliance load monitoring," *IEEE Transactions on Green Communications and Networking*, in press. DOI: 10.1109/TGCN.2021.3087702.

[184] J. Vanschoren, "Meta-learning: A survey," arXiv preprint arXiv:1810.03548, Oct. 2018. [Online]. Available: https://arxiv.org/abs/1810.03548

[185] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. ICML 2017*, Sydney, Australia, Aug. 2017, pp. 1126–1135.

[186] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Elsevier Computer Vision and Image Understanding*, vol. 192, no. 3, p. 102897, Mar. 2020.

[187] R. Mitra, N. B. Gundavarapu, A. Sharma, and A. Jain, "Multiview-consistent semi-supervised learning for 3D human pose estimation," in *Proc. IEEE CVPR 2020*, Seattle, WA, June 2020, pp. 6907–6916.

[188] X. Fan, K. Zheng, Y. Lin, and S. Wang, "Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation," in *Proc. IEEE CVPR 2015*, Boston, MA, June 2015, pp. 1347–1355.

[189] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[190] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Springer Int. J. Computer Vision*, vol. 87, no. 1/2, pp. 1–24, July 2010.

[191] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "RF-based 3D skeletons," in *Proc. ACM SIGCOM 2018*, Budapest, Hungary, Aug. 2018, pp. 267–281.

[192] F. Wang, J. Liu, and W. Gong, "Multi-adversarial in-car activity recognition using RFIDs," *IEEE Trans. Mobile Comput.*, in press. DOI: 10.1109/TMC.2020.2977902.

[193] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proc. IEEE CVPR 2018*, Salt Lake City, UT, June 2018, pp. 8639–8648.

[194] L. Ukkonen and L. Sydanheimo, "Threshold power-based radiation pattern measurement of passive UHF RFID tags," *PIERS Online*, vol. 6, no. 6, pp. 523–526, 2010.

[195] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," arXiv preprint arXiv:1803.02999, Oct. 2018. [Online]. Available: https://arxiv.org/abs/1803.02999

[196] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," in *Proc. 15th Learning and Technol. Conf.*, Jeddah, Saudi Arabia, Feb. 2018, pp. 29–34.

[197] R. Liu, A. A. Ramli, H. Zhang, E. Datta, E. Henricson, and X. Liu, "An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence," *arXiv preprint arXiv:2103.15990*, Aug. 2021. [Online]. Available: https://arxiv.org/abs/2103.15990

[198] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, June 2019.

[199] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, Los Cabos, Mexico, Oct. 2019, pp. 51–56.

[200] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1629–1645, Aug. 2019.

[201] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *MDPI Remote Sensing*, vol. 11, no. 9, p. 1068, Apr. 2019.

[202] F. Adib and D. Katabi, "See through walls with WiFi!" in *Proc. ACM SIGCOMM 2013*, Hong Kong, China, Aug. 2013, pp. 75–86.

[203] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. USENIX NSDI'14*, Seattle, WA, Apr. 2014, pp. 317–329.

[204] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "MU-ID: Multi-user identification through gaits using millimeter wave radios," in *Proc. IEEE INFOCOM'20*, Toronto, Canada, Aug. 2020, pp. 2589–2598.

[205] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in *Proc. ACM MobiCom 2020*, London, UK, Sept. 2020, pp. 1–14.

[206] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su, "mmMesh: towards 3D real-time dynamic human mesh construction using millimeter-wave," in *Proc. ACM MobiSys 2021*, Virtual Conf., June 2021, pp. 269–282.

[207] Z. Chen, T. Zheng, C. Cai, and J. Luo, "MoVi-Fi: motion-robust vital signs waveform recovery via deep interpreted RF sensing," in *Proc. ACM MobiCom 2021*, New Orleans, LA, Oct. 2021, pp. 392–405.

[208] S. Ding, Z. Chen, T. Zheng, and J. Luo, "RF-net: A unified meta-learning framework for RF-enabled one-shot human activity recognition," in *Proc. ACM SenSys'20*, Virtual Conf., Nov. 2020, pp. 517–530.

[209] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, Aug. 2010.

[210] S. M. Hernandez and E. Bulut, "Performing WiFi sensing with off-the-shelf smartphones," in *Proc. IEEE PerCom'20 Workshops*, Austin, TX, Mar. 2020, pp. 1–3.

[211] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *Proc. ACM WiNTECH'20*, Los Cabos, Mexico, Oct. 2019, pp. 21–28.

[212] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *Proc. ACM Mobicom'15*, Paris, France, Sept. 2015, pp. 53–64.

[213] X. Jiao, M. Mehari, W. Liu, M. Aslam, and I. Moerman, "Openwifi CSI fuzzer for authorized sensing and covert channels," *arXiv preprint arXiv:2105.07428*, May 2021. [Online]. Available: https://arxiv.org/abs/2105.07428

[214] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, June 2017.

[215] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal CSI-based human activity recognition using GANs," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, Dec. 2021.

[216] C. Yang, L. Wang, X. Wang, and S. Mao, "Meta-Pose: Environment-adaptive human skele-ton tracking with RFID," in *Proc. IEEE GLOBECOM 2021*, Madrid, Spain, Dec. 2021, pp. 1–6.

[217] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. ACM MobiCom 2018*, New Delhi, India, Sept. 2018, pp. 289–304.

[218] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *Proc. ICML'17*, Sydney, Australia, Aug. 2017, pp. 4100–4109.

[219] H. Xue, W. Jiang, C. Miao, F. Ma, S. Wang, Y. Yuan, S. Yao, A. Zhang, and L. Su, "DeepMV: Multi-view deep learning for device-free human activity recognition," *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–26, Mar. 2020.

[220] F. Wang, J. Liu, and W. Gong, "WiCAR: WiFi-based in-car activity recognition with multi-adversarial domain adaptation," in *Proc. IWQoS'19*, Phoenix, AZ, June 2019, pp. 1–10.

[221] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," whitepaper, Texas Instruments, July 2017. [Online]. Available: https://www.ti.com/lit/wp/spyy005a/spyy005a.pdf

[222] V. Winkler, "Range doppler detection for automotive fmcw radars," in *Proc. 2007 European Radar Conf.*, Munich, Germany, Oct. 2007, pp. 166–169.

[223] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. ACM Mobicom'15*, Paris, France, Sept. 2015, pp. 65–76.

[224] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, Apr. 2016.

[225] F. M. Noori, M. Riegler, M. Z. Uddin, and J. Torresen, "Human activity recognition from multiple sensors data using multi-fusion representations and CNNs," *ACM Trans. Multimedia Comput., Commun., and Appl.*, vol. 16, no. 2, pp. 1–19, May 2020.

Associates, 1994.