

**ROBUST AND ACCURATE LOCALIZATION
ALGORITHMS FOR INDOOR POSITIONING
AND NAVIGATION**

HAN ZOU

HAN ZOU

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2016

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....
Date

.....
HAN ZOU

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Lihua Xie, for his professional guidance, continuous encouragement and unconditional support throughout my Ph.D. study. I truly appreciate the opportunities and the freedom he has offered for finding the research topics I am interested in. It was his profound insights, thoughtful discussions, endless patience and immense knowledge that enabled me to conquer the bottlenecks in my research.

I sincerely appreciate Prof. Costas J. Spanos and Prof. Alexandre Bayen from University of California at Berkeley, Prof. Qing-Shan Jia from Tsinghua University, Prof. Tseng King Jet for the precious exchange opportunities provided by them, their consistent encouragement and valuable suggestions. Furthermore, I would like to appreciate the research scholarship and other financial support from the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program.

Meanwhile, I would like to express my sincere gratitude to my colleagues, Prof. Hao Jiang, Prof. Baoqi Huang, Dr. Hengtao Wang, Zhirong Qiu, Xiaoxuan Lu, Ming Jin, Dr. Kevin Weekly, Yuxun Zhou, Zhenghua Chen, Ruoxi Jia and Kexin Guo for their meaningful collaborations and fruitful discussions. I also would like to appreciate my dear friends, Yaocheng Zhang, Yichao Zhao, and Yiquan Zhou, for their encouragement and support.

Last but not least, I owe my deepest appreciation to my parents, for their unconditional love, unwavering encouragement and continuous support. Without them never would I have been able to keep pursing the Ph.D. degree, and I dedicate this thesis to them.

Abstract

The explosive proliferation of mobile devices and the popularity of social networks have spurred extensive demands on Location Based Services (LBSs) in recent decades. Global Positioning System (GPS) is not capable of providing indoor LBS with sufficient localization accuracy due to the lack of line of sight (LoS) transmission channels between satellites and receivers. Hence, developing Indoor Positioning System (IPS) to provide reliable indoor LBS has been a hot research topic in recent years. IPS has been recognized as a crucial component in numerous applications such as asset tracking, logistics, tourism and security. Furthermore, IPS for occupancy detection can also play an important role in energy saving in buildings.

Various wireless communication technologies have been exploited for indoor positioning and navigation services in the past two decades. Since the existing IEEE 802.11 (WiFi) network infrastructures, such as WiFi routers, have been widely available in large numbers of commercial and residential buildings and nearly every commercial off-the-shelf (COTS) mobile device is WiFi enabled, WiFi based IPS has become the primary alternative to GPS for indoor positioning. Nevertheless, there are still several bottlenecks that restrain them from large-scale implementation.

In this thesis, we aim to propose systematic solutions to overcome the longstanding challenges of existing WiFi based IPSs, and develop novel algorithms and systems that outperform the existing ones in terms of accuracy, reliability, robustness and efficiency. Firstly, in order to overcome the device heterogeneity issue, we propose to standardize WiFi fingerprints by a statistical shape analysis method (i.e. Procrustes analysis), and define Signal Tendency Index (STI) to measure the similarity between

such standardized location fingerprints. Secondly, we address the issue of robustness of the WiFi fingerprinting-based IPS against environmental dynamics by proposing an online sequential extreme learning machine (OS-ELM) based localization algorithm. The fast learning speed of OS-ELM can reduce the time and manpower costs for offline site survey. Meanwhile, its online sequential learning ability enables the proposed localization algorithm to adapt to environmental dynamics in a timely manner. Moreover, we also propose a novel online mutual information (Online-MI) based access point (AP) selection strategy that is able to select the optimal subset of APs to reduce the computational burden and improve the indoor localization accuracy of the IPS. Furthermore, in order to remove the needs of tedious and laborious offline site survey process for WiFi based IPS, we design and develop WinIPS, a WiFi based non-intrusive IPS that enables automatic online radio map construction and adaptation for calibration-free indoor localization. It is able to capture data packets transmitted in the WiFi traffic and extract the received signal strength (RSS) and MAC addresses of both WiFi access points (APs) and mobile devices in a non-intrusive manner. Owing to this unique advantage, the online RSS measurements of APs are obtained and used as online reference points for radio map construction and adaptation in real-time. Our WinIPS received the 3rd Place Award in the IPSN 2014 Microsoft Indoor Localization Competition (the 3rd most accurate system in the Infrastructure-Free Category).

In addition, in order to seamlessly integrate the proposed IPS with GPS, we propose BlueDetect, an accurate, fast and energy-efficient scheme for indoor-outdoor (IO) detection and smooth LBS in all environments running on a mobile device based on the emerging low-power iBeacon technology. By leveraging the portable BLE beacons and Bluetooth module on mobile devices, BlueDetect provides precise IO detection results to turn on/off on-board sensors (such as WiFi and GPS) smartly, improve their performances and reduce the power consumption of mobile devices simultaneously. Furthermore, seamless LBS, such as positioning and navigation service, can be realized by BlueDetect, especially in semi-outdoor environments, which cannot be achieved easily by either GPS or IPS. By integrating BlueDetect with

on-board motion sensors on smartphone, including accelerometer, magnetometer and gyroscope, the system achieved indoor localization accuracy of 1.37m in the infrastructure-based category in the 2015 Microsoft indoor localization competition.

With the aid of the aforementioned algorithms and IPSs, numerous LBSs such as indoor navigation on wearable device (Google Glass), indoor geofencing for a smart lighting control system and seamless indoor outdoor navigation, have been successfully developed and implemented in a wide variety of indoor environments.

Symbols and Acronyms

Symbols

\mathbb{C}	set of complex numbers
\mathbb{R}	set of real numbers
\mathbb{R}^+	set of positive real numbers
\mathbb{R}^N	set of real N dimensional vectors
$\mathbb{R}^{M \times N}$	set of real $M \times N$ dimensional matrices
$\mathbf{A}, \mathbf{B}, \Phi, \dots$	matrices
$\mathbf{x}, \mathbf{y}, \dots$	vectors
x, y, \dots	scalars
$\mathbf{A}_{i,:}$	i th row of \mathbf{A}
$\mathbf{A}_{:,j}$	j th column of \mathbf{A}
A_{ij}	ij th entry of \mathbf{A}
\mathbf{A}^T	transpose of \mathbf{A}
\mathbf{A}^H	complex transpose of \mathbf{A}
x_i	i th entry of \mathbf{x}
$\ \cdot\ _2$	spectral norm for a matrix, or Euclidean norm for a vector
$ \cdot $	determinant for a matrix, or absolute value for a scalar, or cardinality for a set
$\text{supp } \{\mathbf{x}\}$	support of \mathbf{x} defined as $\{i : x_i \neq 0\}$
$\ \mathbf{x}\ _p$	ℓ_p norm ($p \geq 1$) or pseudo-norm ($0 < p < 1$) of \mathbf{x} defined as $(\sum_i x_i ^p)^{\frac{1}{p}}$
$\ \mathbf{x}\ _0$	ℓ_0 pseudo-norm of \mathbf{x} defined as the number of nonzero entries

$\ \mathbf{x}\ _\infty$	ℓ_∞ norm of \mathbf{x} defined as $\max\{ x_i \}$
$\langle \mathbf{x}, \mathbf{y} \rangle$	inner product of \mathbf{x} and \mathbf{y}
$\mathbf{x} \succeq \mathbf{y}$	$x_i \geq y_i$ for all i
$E\{\cdot\}$	expectation
$Var\{\cdot\}$	variance
L	number of hidden nodes

Acronyms

IPS	indoor positioning system
LBS	location based service
AP	access point
RSS	received signal strength
REP	repeatability
ML	machine learning
STD	stand deviation
MD	mobile device
RD	reference device
TD	testing device
RP	reference point
KNN	K nearest neighbor
WKNN	weighted K nearest neighbor
MMSE	minimum mean square error
HVAC	heating, ventilating, and air conditioning
SLFN	single-hidden layer feedforward neural network
ELM	extreme learning machine
WPL	weighted path loss
OnlineMI	online mutual information
STI	signal tendency index

OS-ELM	online sequential extreme learning machine
WinIPS	WiFi-based Non-intrusive Indoor Positioning System
PSFM-GPR	Gaussian Process Regression with Polynomial Surface Fitting Mean
WinSMS	WiFi-based Non-intrusive Sensing and Monitoring System
VRP	virtual reference point
IoT	Internet of Things

Contents

Acknowledgements	iii
Abstract	vii
Symbols and Acronyms	x
List of Contents	xvi
List of Figures	xx
List of Tables	xxii
List of Algorithms	xxiii
1 Introduction	1
1.1 Motivation and Objectives	1
1.2 Main Contributions	4
1.3 Organization of the Thesis	7

2 Literature Review	9
2.1 Overview of Sensing Technologies for Indoor Localization	9
2.1.1 Cameras and visible light	10
2.1.2 Magnetic field	10
2.1.3 Infrared (IR)	11
2.1.4 Ultrasound and acoustic signal	11
2.1.5 Ultra-wideband (UWB)	12
2.1.6 Inertial Measurement Unit (IMU)	12
2.1.7 Radio Frequency Identification (RFID)	13
2.1.8 Bluetooth Low Energy (BLE)	15
2.1.9 Wireless Local Area Networks (WiFi)	16
2.1.10 Summary of sensing technologies for indoor localization	17
2.2 Overview of Indoor Localization Algorithms	18
2.3 Memoryless Localization Algorithms	19
2.3.1 Signal propagation based localization algorithms	19
2.3.2 Fingerprinting based localization algorithms	20
2.3.3 Limitations of fingerprinting-based localization algorithms	23
2.4 Tracking	29
2.4.1 Pedestrian Dead Reckoning (PDR) based tracking	29
2.4.2 Learning based tracking	30
2.5 Integration of IPS and GPS	32
2.5.1 GPS-Based IO detection	32
2.5.2 Lightweight on-board sensor-based IO detection	33
2.6 Conclusion	34

3 Localization Algorithms for Indoor Positioning	35
3.1 Weighted Path Loss (WPL)	36
3.1.1 Indoor path loss model	36
3.1.2 Methodology of WPL	36
3.1.3 Experimental results and performance evaluation	37
3.2 Extreme Learning Machine (ELM) based Localization Algorithm . . .	44
3.2.1 Motivations	44
3.2.2 Preliminaries on ELM	45
3.2.3 Methodology of ELM	46
3.2.4 Experimental results and performance evaluation	47
3.3 Integrated WPL-ELM	51
3.3.1 Motivations	51
3.3.2 Methodology of integrated WPL-ELM	52
3.3.3 Experimental results and performance evaluation	53
3.4 Conclusion	55
4 Robustness of IPS Against Device Heterogeneity	57
4.1 Standardizing WiFi Fingerprints based on the Procrustes Analysis Method	59
4.1.1 Experimental analysis	60
4.1.2 Standardizing RSS fingerprints	61
4.1.3 Theoretical analysis	64
4.2 Proposed STI-WELM Algorithm	67
4.2.1 Preliminaries on WELM for indoor localization	67

4.2.2	STI-WELM	70
4.3	Experimental Study	72
4.3.1	Testbed	72
4.3.2	System setup and data collection procedure	73
4.3.3	Comparison between RSS, SSD and STI as location fingerprints	74
4.3.4	Experimental results and evaluation	75
4.3.5	Implementation of the proposed approach for real location-based service	87
4.4	Conclusion	88
5	Robustness of IPS Against Environmental Dynamics	89
5.1	An OS-ELM based Indoor Localization Algorithm	90
5.1.1	Preliminary on OS-ELM	90
5.1.2	OS-ELM based indoor localization algorithm	92
5.2	Simulation Results and Evaluation	95
5.3	Experimental Results and Performance Evaluation	98
5.3.1	System overview	98
5.3.2	Selection of parameters for the initial OS-ELM model	100
5.3.3	Comparison between OS-ELM and other methods	103
5.3.4	Performance evaluation of OS-ELM under specific environmental dynamics	106
5.4	Conclusion	111

6 Access Point Selection Strategy	113
6.1 OnlineMI based AP Selection Strategy and Localization Algorithm	115
6.1.1 OnlineMI AP selection strategy	115
6.1.2 OnlineMI-WKNN localization algorithm	118
6.2 Simulation Results and Evaluation	119
6.2.1 Simulation setup	119
6.2.2 Performance evaluation	120
6.3 Experimental Results and Performance Evaluation	124
6.3.1 Performance evaluation	124
6.4 Conclusion	129
7 Online Radio Map Construction	131
7.1 WiFi-based Non-Intrusive Indoor Positioning System (WinIPS)	134
7.1.1 System overview	134
7.1.2 WinSMS for RSS data acquisition	135
7.1.3 PSFM-GPR for online radio map construction and adaptation	138
7.1.4 STI-WKNN Localization algorithm for heterogeneous mobile device	145
7.2 Experimental Results and Discussions	148
7.2.1 Experimental setup	148
7.2.2 RSS estimation accuracy	150
7.2.3 Localization estimation Accuracy	155
7.3 Conclusion	160

8 Seamless Integration of IPS and GPS	163
8.1 System Design	165
8.1.1 System overview	165
8.1.2 Seamless transition between outdoors and semi-outdoors . . .	166
8.1.3 Seamless transition between semi-outdoors and indoors . . .	168
8.1.4 Semi-outdoors (iBeacon)	169
8.2 Evaluation	173
8.2.1 IO detection accuracy of BlueDetect	175
8.2.2 Localization accuracy of BlueDetect in semi-outdoor environments	177
8.2.3 Power consumption of BlueDetect	178
8.3 Conclusion	181
9 Conclusions and Future Works	183
9.1 Conclusions	183
9.2 Future works	187
Author's Publications	196
Bibliography	197

List of Figures

1.1	Indoor Positioning System [1]	2
2.1	Overview of indoor localization technologies in terms of accuracy and coverage [2]	18
2.2	Overview of memoryless localization algorithms for indoor LBS	19
3.1	Placement of RIFD reference tags, tracking tags, sensors and reader in Experiment I	38
3.2	Relationship between RSSI and distance	40
3.3	Cumulative percentile of error distance for different methods	42
3.4	Placement of RIFD reference tags, tracking tags, sensors and reader in Experiment II and Experiment III	43
3.5	Comparison of distance error distribution for different methods	43
3.6	Cumulative percentile of error distance for different methods	48
3.7	Comparison of distance error distribution for different methods	49
3.8	Flowchart of the integrated WPL-ELM approach	53
3.9	Cumulative percentile of error distance for different methods	56
4.1	WiFi RSS values measured by different mobile devices at the same location.	61

4.2	Layout of the testbed	73
4.3	Distribution histogram of the average standard deviations of different location fingerprints	75
4.4	Comparison of distance error distributions for different methods (KNN) .	80
4.5	Comparison of distance error distributions for different methods (ELM and WELM)	81
4.6	Comparison of mean localization error between different approaches under the influence of the number of APs.	86
5.1	Methodology of the online sequential extreme learning machine (OS-ELM) approach.	93
5.2	Positions of the WiFi APs, offline calibration points, online calibration points and online testing points in the simulated field.	96
5.3	Cumulative percentile of error distance.	98
5.4	Positions of the WiFi APs, offline calibration points, online calibration points and online testing points in the test-bed.	99
5.5	The test mobile device Samsung I929 and developed Android apps. .	100
5.6	Localization accuracy regarding different activation functions and different numbers of hidden nodes.	102
5.7	Deviation of distance error with different numbers of hidden nodes. .	103
5.8	Cumulative percentile of error distance for different methods.	106
5.9	Comparison of the distance error distribution for different methods. (a) KNN; (b) fuzzy KNN; (c) batch ELM; (d) OS-ELM.	107
6.1	Locations of the WiFi APs, offline calibration points and online testing points in the simulated field.	119

6.2 Comparison of mean error between different AP selection methods for different numbers of APs.	121
6.3 Comparison of distance error distribution for different methods, where the number of APs is twelve.	123
6.4 Comparison of distance error distribution for different methods, where the number of APs is twelve.	123
6.5 Positions of the WiFi APs, offline calibration points and online testing points in the testbed.	125
6.6 Comparison of mean error between different AP selection methods for different numbers of APs.	126
6.7 Comparison of distance error distribution for different methods, where the number of APs is eight.	127
6.8 Comparison of distance error distribution for different methods, where the number of APs is twelve.	128
7.1 System architecture of WinIPS.	134
7.2 System architecture of WinSMS	136
7.3 Visualization of pairwise RSS matrix of 8 APs (dBm)	138
7.4 Scenario of online radio map construction in complex indoor environment.	139
7.5 Irregular RSS distribution of an AP in complex indoor environment. .	142
7.6 WiFi RSSs measured by different mobile devices at identical locations.	146
7.7 (a) Layout of the testbed at the beginning of the experiment (T_1). (b) Layout of the testbed six months later after renovation (T_2).	149
7.8 Comparison of RSS variation distribution between short-term and long-term time period.	151

7.9	RSS estimation accuracy using PSFM-GPR.	152
7.10	Comparison of estimated RSS error distribution of AP8 using different methods.	154
7.11	Plots of the RSS estimation error vs. the localization error, color coded by the number of APs in use.	157
7.12	Comparison of localization accuracy between different online RSS Prediction Methods.	157
7.13	Comparision of cumulative distribution of location error between WKN-N and STI-WKNN.	159
8.1	Representative scenes and corresponding localization technologies of three different environments.	166
8.2	SNR of GPS signals in different environments.	168
8.3	Relationship between RSS of a BLE beacon and distance.	174
8.4	The test walking route in the university campus and indoor-outdoor (IO) detection accuracy comparison.	175
8.5	Screenshots of BlueDetect in the three environment types.	176
8.6	Estimote BLE Beacon.	176
8.7	Comparison of the cumulative distribution of the location error between GPS and BlueDetect in semi-outdoor environments.	179
8.8	Screenshot of the power-monitoring app.	180
8.9	Power consumption of various sensors on a mobile device (Nexus 6) for different IO detection methods.	180

List of Tables

2.1	Overview of major sensing technologies for indoor localization.	10
3.1	Localization accuracy statistics	40
3.2	Comparison between WPL and other methods	42
3.3	Comparison between ELM and other methods	48
3.4	Comparison between WPL-ELM and ELM	55
4.1	Training inputs and training target for WiFi RSS fingerprint database	68
4.2	Mobile devices used for data collection	74
4.3	Detailed average localization errors (in meter) of STI-KNN without scaling and STI-KNN	77
4.4	Detailed average localization errors (in meter) under various situations (KNN)	79
4.5	Summary of average localization errors (in meter) (KNN)	82
4.6	Detailed average localization errors (in meter) under various situations (ELM)	83
4.7	Summary of average localization errors (in Meter) (ELM)	84
4.8	Average localization errors (m) under the influence of the No. of APs	86

5.1	Simulation results of OS-ELM.	98
5.2	Comparison between OS-ELM and other methods.	105
5.3	Impact of human presence and movements on localization accuracy. .	109
5.4	Impact of opening/closing doors on localization accuracy.	110
6.1	Performance comparison of different methods (No. of AP: 8)	122
6.2	Performance comparison of different methods (No. of AP: 12)	122
6.3	Performance comparison of different methods (No. of AP: 8)	126
6.4	Performance comparison of different methods (No. of AP: 12)	128
7.1	Online RSS observation among APs captured by WinSMS	137
7.2	Comparison of RSS estimation accuracy between different methods (dBm)	153
7.3	Comparison of localization accuracy (m) between online RSS prediction methods	155
7.4	Comparison of localization accuracy using offline site survey and PSFM-GPR (m)	156
7.5	Comparison of localization accuracy among different mobile devices (m)	159
8.1	Analysis of BLE beacon's RSS variations under LOS and NLOS conditions.	172
8.2	Analysis of BLE beacon's RSS variations with the influence of different holding orientations of the mobile device.	173

List of Algorithms

3.1	WPLELM	52
4.1	STI-WELM	70
7.1	Online radio map update algorithm	144
8.1	Bluedetect IO detection and localization algorithm (outdoor \rightleftharpoons semi-outdoor)	167
8.2	Bluedetect IO detection and localization algorithm (semi-outdoor \rightleftharpoons indoor)	169

Chapter 1

Introduction

1.1 Motivation and Objectives

The explosive proliferation of mobile devices and the popularity of social networks have spurred extensive demands on Location Based Services (LBSs) in recent decades. Although GPS is the primary choice for location estimation in outdoor environments, it is not capable of providing positioning services with sufficient localization accuracy in indoor environments due to the lack of line of sight (LoS) transmission channels between satellites and receivers [3]. Therefore, great efforts have been devoted to developing Indoor Positioning Systems (IPSs) so as to enable reliable and precise indoor positioning and navigation. As shown in Figure 1.1, IPS has been recognized as a crucial component in numerous applications such as asset tracking, logistics, tourism and security. According to a recent market research, the indoor location market is estimated to grow from \$597 million in 2014 to \$3961.8 million by 2019 at a Compound Annual Growth Rate (CAGR) of 46.0% from 2014 to 2019 [4].

In the past two decades, various wireless communication technologies, such as infrared (IR), ultrasound, Ultra-Wideband (UWB), Radio Frequency Identification (RFID), Bluetooth, IEEE 802.11 (WiFi) and so on, have been exploited for indoor



Figure 1.1: Indoor Positioning System [1]

positioning and navigation services [5,6]. Due to the rapid development of miscellaneous wireless communication technologies to accommodate LBSs, IPSs have been broadly developed and numerous algorithms have been studied recently [7,8].

Furthermore, IPS for human activity sensing can also play an important role in energy saving in buildings. Nowadays, building energy usage accounts for an increasing proportion of the total energy consumption. It's reported that the building electricity consumption accounted for 40% of the total electricity consumption in United States in 2005 and 33% in Singapore in 2010 respectively [9,10]. Specifically, most heating, ventilating, and air conditioning (HVAC) systems assume a maximum occupancy status of a room during normal working periods, and a minimum occupancy status at nights. This results in energy waste because the occupancy level of rooms is conditioned to the number of occupants who are actually present in the area [11]. The knowledge of occupancy distribution within a building provided by IPS offers the potential of significant energy savings [12]. There are basically three stages to link human activity sensing and a building management system. Firstly, by leveraging the information provided by IPS, we are able to capture human presence and behaviors in the indoor environment. After that, we can understand human activity pattern and develop human activity forecasting algorithms. Finally,

we can develop dynamic models of occupancy to map localized energy requirements and forecast building energy consumption.

Unlike other wireless technologies requiring the deployment of extra infrastructures, the existing IEEE 802.11 (WiFi) network infrastructures such as WiFi routers have been widely available in large numbers of commercial and residential buildings. More importantly, nearly every existing commercial mobile device is WiFi enabled. As such, WiFi based IPS has become the primary alternative to GPS for indoor positioning.

The existing WiFi based IPSs are able to provide 3-5m localization accuracy in general. Nevertheless, several drawbacks restrain them from large-scale implementation. First of all, most of the existing WiFi based IPSs involve an offline site survey process, which is extremely time-consuming and labor-intensive. Multiple received signal strength (RSS) samples need to be measured at numerous calibration points to ensure localization accuracy. Secondly, the offline calibrated database is vulnerable to environmental dynamics [13, 14], as the real-time RSS readings collected during online localization phase could deviate from those stored in the offline database due to the variation of temperature, humidity, occupancy distribution and multipath effects. Poor localization results will be introduced if the radio map is not updated adaptively. Furthermore, the existing WiFi based IPSs also suffer from device heterogeneity across commercial off-the-shelf (COTS) mobile devices [15]. Due to the proliferation of various types and brands of mobile devices, it is indispensable and urgent to develop a robust location fingerprinting technique to provide accurate, reliable and fast indoor positioning services for heterogeneous devices. Another challenge for existing WiFi based IPSs comes from the high computational burden due to the analysis of multiple RSS values from numerous access points (APs) simultaneously. Certain AP selection strategy is desired to reduce this computational burden while preserve or even improve the localization accuracy.

The objective of this thesis is to tackle the aforementioned issues and provide effective solutions to improve the performance of IPSs comprehensively. We demonstrate

that the major problems of existing WiFi based IPSs (e.g. device heterogeneity issue, calibration and maintenance efforts, the computational overhead) can be overcome by adopting our proposed algorithms accordingly. In addition, how to provide seamless LBS in all environments by integrating IPS with GPS is usually neglected and remains challenging. This will be also addressed in our work. We aim to develop accurate and reliable indoor LBS to locate and track mobile devices as well as the occupants.

1.2 Main Contributions

The main contributions of the thesis are listed as follows:

1. We propose three localization algorithms that can be easily implemented for all the radio frequency (RF) based IPS. Weighted Path Loss (WPL) is a centralized signal propagation model based localization algorithm which can provide an accurate location estimation of the mobile device without any extra infrastructure. Moreover, Extreme Learning Machine (ELM), a fast machine learning technique, is utilized as a fingerprinting based approach for indoor localization. Noticing that signal propagation model based approaches provide a fast location estimate but with limited accuracy, while fingerprinting based approaches provide a higher localization accuracy but with a higher computational complexity, we further propose an integrated WPL-ELM algorithm, which combines the fast estimation of WPL and the high localization accuracy of ELM.
2. After conducting a comprehensive theoretical and experimental analysis of the device heterogeneity issue for WiFi based IPS, we propose to standardize WiFi fingerprints based on a statistical shape analysis method (i.e. Procrustes analysis), and define Signal Tendency Index (STI) to measure the similarity between such standardized location fingerprints. Furthermore, considering the

fact that ELM provides a good generalization performance at an extremely fast learning speed, we integrate the weighted version of ELM, termed WELM, and STI to develop an efficient and robust method named as STI-WELM.

3. We propose an indoor localization algorithm based on an online sequential extreme learning machine (OS-ELM) to address the issue of non-robustness to environmental dynamics of the existing WiFi based IPSs. The fast learning speed of OS-ELM can reduce the time and manpower costs for the offline site survey. Meanwhile, its online sequential learning ability enables the proposed localization algorithm to adapt to environmental dynamics in a timely manner. Both simulation and experimental results verify the superiority of the proposed OS-ELM algorithm.
4. In order to reduce the computational burden of the WiFi based IPS, we propose a novel AP selection strategy, online mutual information (OnlineMI), that is able to select the optimal subset of APs, and in the meanwhile improve the localization accuracy. Unlike traditional AP selection methods that only consider the individual discriminate ability of all the APs, OnlineMI measures the collective discriminate ability of different groups of APs by the mutual information within the group. Furthermore, since its AP selection process is conducted online associated with the real-time location of the mobile device, OnlineMI can select the subset of APs that provides most online information for indoor positioning consistently and adaptively even under various environmental dynamics.
5. In order to remove the tedious and laborious offline site survey process, we design and develop WinIPS, a WiFi based non-intrusive IPS that enables automatic online radio map construction and adaptation for calibration-free indoor localization. For RSS data acquisition, we develop WinSMS, a novel intelligent wireless sensing system that captures data packets in the WiFi traffic and extract the RSS and MAC addresses of both WiFi APs and mobile devices in a non-intrusive manner. Owing to this unique advantage, we can

obtain the online RSS measurements of APs and use them as online reference points for radio map construction and adaptation in real-time. To establish a more fine-grained radio map, an upgraded Gaussian process regression (GPR) with polynomial surface fitting mean (PSFM-GPR) is introduced to precisely model the irregular RSS distribution over the complex indoor environment. This online generated fine-grained radio map is more adaptive and robust to environmental dynamics than the traditional offline calibrated radio map. The extensive experiments conducted in a real multi-functional office show that WinIPS outperforms existing solutions in terms of both RSS estimation accuracy and localization accuracy. Since it completely gets rid of the cumbersome offline site survey process while providing outstanding indoor LBS, these merits make it more suitable for practical large-scale implementation. The WinIPS received the *3rd* Place Award in the IPSN 2014 Microsoft Indoor Localization Competition (the *3rd* most accurate system in the Infrastructure-Free Category).

6. In an effort to seamlessly integrate the proposed IPS with GPS, we propose BlueDetect as an accurate, fast and energy-efficient scheme for indoor-outdoor (IO) detection and seamless LBS running on a mobile device based on the emerging low-power iBeacon technology. By leveraging the portable BLE beacons and Bluetooth module on mobile devices, BlueDetect provides precise IO detection results to smartly turn on/off on-board sensors (such as WiFi and GPS), improve their performance and reduce the power consumption of mobile devices simultaneously. Furthermore, seamless LBS such as positioning and navigation service can be realized by BlueDetect, especially in semi-outdoor environments. This cannot be achieved by GPS or IPS easily. We prototyped BlueDetect on multiple Android mobile devices and analyzed its performance comprehensively. It provides a higher IO detection accuracy, higher localization accuracy in semi-outdoor environments while consuming less battery than existing schemes. It is a feasible solution for IO detection and can be

extended for other services such as geo-fencing and floor identification in the near future.

Furthermore, some video demos of our systems developed in this thesis, including non-intrusive WiFi based indoor positioning and navigation system, indoor navigation on Google Glass, indoor geofencing for a smart lighting control system, and seamless indoor outdoor navigation, are available on our YouTube Channel [16].

1.3 Organization of the Thesis

The remainder of this thesis is organized as follows:

Chapter 2 presents a literature review on sensing technologies and algorithms for indoor localization. A variety of key technologies (e.g. vision, magnetic field, infrared, UWB, IMU, RFID, iBeacon and WiFi) are introduced and juxtaposed in terms of localization accuracy, cost, power consumption and the requirement of extra devices on the user-side. Existing signal propagation model based and fingerprinting based localization algorithms are also reviewed in this chapter. The main contributions of the thesis are detailed in Chapters 3-8.

Chapter 3 introduces our proposed three indoor localization algorithms, namely WPL, ELM and Integrated WPL-ELM, which can be easily implemented for all RF based IPS. Our work presented in this chapter has been published in *Unmanned Systems* [17], *IEEE CPSNA 2013* [18] and *IPIN 2013* [19].

Chapter 4 introduces our proposed STI location fingerprint, which aims to overcome the device heterogeneity issue for WiFi based IPS. Our work presented in this chapter has been published in *IEEE Transactions on Wireless Communication* [20] and *IEEE WCNC 2016* [21].

Chapter 5 describes the OS-ELM based localization algorithm we have proposed to make the WiFi based IPS more robust to various environmental dynamics. This

chapter is mainly based on our papers that have been published in *Sensors* [14] and *IEEE WF-IoT 2014* [22].

Chapter 6 presents the proposed OnlineMI AP selection strategy which aims to reduce the computational burden of the WiFi based IPS. Our work has been published in *IEEE CASE 2015* [23].

Chapter 7 is devoted to developing a WiFi based non-intrusive IPS, WinIPS, that enables automatic online radio map construction and adaptation for calibration-free indoor localization. This chapter is mainly based on our papers that have been published in *IEEE INFORCOM 2016* [24] and submitted to *IEEE Transactions on Industrial Informatics*.

Chapter 8 presents the proposed BlueDetect system, which is an accurate, fast response and energy-efficient scheme for Indoor-Outdoor (IO) detection and seamless LBS running on a mobile device based on the emerging low-power iBeacon technology. Our work presented in this chapter has been published in *Sensors* [25].

Chapter 9 concludes the thesis and highlights some potential future research directions.

Chapter 2

Literature Review

The explosive proliferation of mobile devices and the popularity of social networks have spurred extensive demands on LBSs in recent decades. Great efforts have been devoted to developing IPSs so as to enable reliable and precise indoor positioning and navigation in the past two decades [5–8]. This chapter is devoted to a comprehensive review of recent advances in indoor localization, from the angles of sensing technologies and localization algorithms. The major bottlenecks for existing IPS and localization algorithms are discussed as well. Furthermore, existing solutions for the integration of IPS and GPS are presented in this chapter.

2.1 Overview of Sensing Technologies for Indoor Localization

Various sensing technologies have been proposed and developed to provide indoor LBS. The main characteristics of these technologies (e.g. localization accuracy, cost, power consumption and the requirement of extra devices on the user-side) are summarized in Table 2.1. Multiple key technologies are highlighted in this section.

Table 2.1: Overview of major sensing technologies for indoor localization.

Technology	Localization Accuracy	Extra Device on User-Side	Power Consumption	Cost
Infrared	0.5-3 m	Yes	Low	Moderate
Acoustic signal	30-80 cm	No	Low	Moderate
RFID	1-3 m	Yes	Low	Moderate
UWB	10-50 cm	Yes	Low	High
PDR	1-5 m	No	High	Low
WiFi	2-5 m	No	High	Low
BLE (iBeacon)	1-5 m	No	Low	Low

2.1.1 Cameras and visible light

Nowadays, commercial cameras are able to provide high resolution of space, and accommodate information regarding objects in a scene, including texture, shape, color, size, and so on. Furthermore, the fast development of computer vision technology makes cameras widely adopted by robots for simultaneous localization and mapping (SLAM) tasks. Several vision-based IPSs have been proposed in recent years. For instance, Epsilon leverages visible light from smart LEDs coupled with custom light sensor receivers for localization [26]. It requires the LOS condition to achieve sub-meter localization accuracy. Luxapose utilizes commercial cameras and image processing techniques to realize indoor localization [27]. It is able to provide both location and orientation estimation of the object. However, the analysis process of the image introduces around 10 s delay for location estimation.

In summary, the requirements of specific environmental and lighting conditions, as well as the privacy issue are the bottlenecks for vision-based IPS.

2.1.2 Magnetic field

Several researchers have proposed to use magnetic field for indoor positioning and navigation [28–31]. These approaches can be classified into two categories: geomagnetic field based and artificial generated magnetic field based. Geomagnetic field is usually adopted to estimate the walking orientation of a pedestrian. However,

various magnetic inferences severely degrade the estimation accuracy. Meanwhile, some researchers actually leverage these anomalies as fingerprints (location signatures) to perform geomagnetic field fingerprinting based indoor localization [31, 32]. Artificially generated magnetic fields are also utilized for indoor positioning. Customized coils are required to be deployed in the environment. The location of user is estimated by trilateration algorithm according to the received strengths of multiple coils' magnetic fields [33]. A 3D indoor tracking system based on magnetic field is proposed in [30]. The requirement of dedicated infrastructure is the bottleneck of the magnetic field based IPSs.

2.1.3 Infrared (IR)

The range of infrared (IR) wavelength is from 700 nm to 1 mm, and the corresponding frequency range is from 430 THz to 300 GHz. Its wavelengths are longer than visible light, which makes IR-based IPSs less intrusive than vision-based IPSs. The Active Badge [34] is a pioneer work using infrared signals to realize indoor localization. There are basically three methods of using infrared signals for indoor localization: active infrared beacons, infrared imaging using natural radiation and artificial light sources. Artificial IR light source system is able to provide sub-mm localization accuracy. Active beacons and natural radiation are employed for providing coarse location estimation or computing the occupancy distribution of indoor environments.

The high deployment cost of extra infrared infrastructure is the main issue of IR-based IPSs in commercial applications.

2.1.4 Ultrasound and acoustic signal

Acoustic-based IPSs, such as Cricket [35], Bat [36] and Guoguo [37], are adopting ultrasound to realize indoor localization. Transmitters on the user side broadcast a short ultrasound pulse firstly. Then, multiple receivers at fixed locations pick up the

signal and estimate the location of the user. The work in [38] leverages high-pitched acoustic signals emitted from commercial mobile devices to realize indoor tracking.

To achieve higher localization accuracy, a dense deployment of dedicated devices is required for acoustic-based IPSs. Furthermore, the large sound noise in certain indoor environments and small coverage also restrict the large scale commercial applications of acoustic-based IPSs.

2.1.5 Ultra-wideband (UWB)

Ultra-Wideband (UWB) is a radio technology for high-bandwidth short-range communication with strong multipath resistance and penetrability through different building materials, which is beneficial for indoor ranging, localization and tracking. Unlike narrowband operation, UWB waves occupy a large frequency bandwidth (> 500 MHz). Its harmfulness to the human body is alleviated by the low power spectral density, which also facilitates restricting the interference of UWB signals with other narrowband receivers [2]. One major advantage of using UWB is that it is capable of penetrating construction materials, e.g., wood, concrete and glass. This is favorable for indoor localization as it capacitates ranging under non-line-of-sight (NLOS) transmission and enables inter-room distance measurements. Large bandwidth which translates into a high resolution in time and consequently in range is another superiority of using UWB for ranging [2]. Favored by the short duration (less than 1ns) of UWB pulses, the reflected signals from the original signals can be filtered, which largely improves its accuracy.

2.1.6 Inertial Measurement Unit (IMU)

The IMU is a device that is capable of measuring an entity's velocity, gravitational forces and orientation of moving by gyroscopes (angular velocity sensors) and accelerometers (acceleration sensors). Besides gyroscopes, most of mobile devices

also incorporate magnetic field sensors as a compass (magnetometers) in their IMU function.

IMU sensors are usually leveraged to infer precise motion status of user, including step detection, walking length estimation and walking direction estimation. Since the vertical acceleration signal contains periodic patterns when feet hit the ground during walking, it is usually employed for step detection [39]. Time-domain approaches such as peak detection [40], zero-crossing counting [41], auto-correlation [42], and thresholding [43] are proposed for step counting. Furthermore, frequency-domain approach [44] and feature clustering approach [45] are also leveraged to extract features from the inertial measurements for step detection. For walking length estimation, majority of studies aim to build regression models according to accelerations. For instance, step frequency [46] and the difference between the peak and trough values of the vertical accelerations [47] are utilized to infer the step length. Both gyroscope and magnetometer readings are usually utilized for walking direction estimation [39, 48]. In complex indoor environments, magnetometer measurements are easily disturbed by metal and electronic devices. The integration of gyroscope readings, i.e. angular velocities, drifts gradually due to the sensor noise of gyroscope. Therefore, the sensor fusion of both approaches is a robust method for walking direction estimation. This information provided by IMU sensors allows one to track an entity's position comprehensively and make pedestrian dead reckoning (PDR) feasible on mobile devices. However, the estimation accuracy falls dramatically as different placement of devices by holders can result in different measurements [49]. Thus, various RF based IPSs including ultrasound [50], RFID [51], iBeacon [39] and WiFi [52] are integrated with PDR and inertial approaches to make the entire system more accurate and robust.

2.1.7 Radio Frequency Identification (RFID)

Radio Frequency Identification (RFID) is a means of storing and retrieving data through electromagnetic transmission to a RF compatible integrated circuit. RFID-based IPSs usually consist of RFID readers with antenna which communicate with

nearby active or passive RFID tags. RFID technology has several advantages, including anti-interference, unique identification for objects and no requirement of LOS condition. It has been widely used in asset tracking, industrial automation and medical care. The application of RFID technology in IPS has become a hot research topic in recent years. RFID based IPSs such as SpotON [53] and LANDMARC [54] can successfully uniquely identify, localize and track equipment and persons.

A typical RFID based IPS consists of three basic components: RFID readers, RFID tags and their interconnecting communication network. Both RFID readers and tags use a predefined RF frequency and protocol to transmit and receive data. The RFID reader is able to read the data emitted from RFID tags. RFID tags can be classified into two categories: passive and active tags.

Passive RFID tags operate without batteries and are mainly used to replace the traditional barcode technology. A variety of RFID based IPSs by adopting passive RFID tags have been proposed in the past two decades [55, 56]. Although they are lighter and less expensive than active tags, the range of the passive RFID tags is limited to approximately 1 m to 2 m which largely restricts the coverage area of their system [57].

Active RFID tags are small transceivers equipped with button-cell batteries. They can actively transmit their ID and additional information to RFID readers. In contrast to passive RFID tags, a typical active RFID tag enables long transmission range of 30m or more with the help of an onboard radio and a small antenna, rendering it quite suitable for identifying and tracking high-unit-value products or persons in complex indoor environments. RFID based IPSs which use active RFID tags such as SpotON [53] and LANDMARC [54] have been proposed in recent decades. SpotON is a fine-grained IPS based on RFID signal strength. SpotON tags are custom devices that operate standalone or potentially as a plug in card enabling larger devices to take advantage of location-sensing technology. It can provide 3D location information of the tag. LANDMARC is one of the earliest and most famous IPSs by using active RFID tags and readers. In order to increase accuracy

without placing more readers, extra fixed location reference tags are introduced in LANDMARC to facilitate location calibration. An RSS radio map is constructed by collecting the RSS from each tag to readers. The system receives the RSS data from both reference tags and tracking tags in real time. After comparing the RSSs of the tracking tags with those of reference tags, the weighted k-nearest neighbour algorithm is adopted to estimate the locations of the tracking tags. It is reported that the localization accuracy of LANDMARC is around 1.5 m to 2 m with 50 percent probability. An enhanced LANDMARC approach has been proposed in [58], aiming to make the calculated coordinate of the tracking tags closer to the real time measurements without extra readers and reference tags.

Nevertheless, the deployment of extra infrastructure is still costly compared with those modalities based on existing infrastructure. Besides, on the user's side, they are required to carry the tags, which is not convenient and may bring maintenance issues.

2.1.8 Bluetooth Low Energy (BLE)

Bluetooth Low Energy (BLE) aims to transmit data between devices in short range with lower power consumption, lower cost and higher security compared to classic Bluetooth protocol [59, 60]. In [60], K nearest neighbor (KNN) algorithm and the BLE RSS are used to estimate the location of device. iBeacon, an advanced Bluetooth protocol is proposed by Apple [61], which is built upon BLE. It makes use of BLE proximity sensing to broadcast their unique identifiers to nearby portable mobile devices and trigger a location-based action on these devices. Since the iBeacon protocol uses very short duration messages and does not need a paired connection with mobile devices (broadcast only), it is much more power efficient than classical Bluetooth protocols and less power hungry on the user-side than GPS and WiFi [59]. With such a merit, a BLE beacon can run on a coin cell battery for months or even for years. According to a recent study on the battery life of 16 major iBeacon hardware devices [62], by setting the advertising interval as 645 ms, an iBeacon with

a CR2450 620-mAh coin cell battery is able to provide 11.2 months of life, which increases to two years as the advertising interval is increased to 900 ms. Nowadays, the iBeacon protocol is becoming a built-in standard for mobile devices, and a high density deployment of BLE beacons in buildings for multiple purposes will be expected in the near future.

2.1.9 Wireless Local Area Networks (WiFi)

Wireless Local Area Networks (IEEE 802.11 standard, normally called Wi-Fi) have been exploited and used to estimate the location of a mobile target in indoor environments for a few years. Since the existing IEEE 802.11 network infrastructures such as WiFi routers are widely available in large numbers of commercial and residential buildings and nearly every mobile device now is equipped with a WiFi receiver [63, 64], it is low-cost and practical to develop a WiFi based IPS to provide LBS in indoor environment.

Fingerprinting-based localization algorithm is the most widely adopted algorithm for WiFi based IPS due to its ability to capture the signal variance in complex indoor environment more accurately than other algorithms [65–69]. It consists of two phases: offline calibration phase and online localization phase. In the offline calibration phase, a site survey is performed to record the RSS from multiple APs at many calibration points and their physical coordinates to form a fingerprint. Then the data form a RSS fingerprint database (a.k.a. radio map). During the online localization phase, the location of mobile device is estimated by matching the observed RSS readings against the offline RSS fingerprint database. The pioneering work of employing fingerprinting-based approaches for WiFi based IPSs is RADAR [65] from Microsoft Research. It is able to achieve 2-3 m with a probability of 50%. Different from RADAR, which collects RSSI values at each reference point deterministically, Horus [66] stores the RSSI distribution of each AP in the fingerprint database and makes use of probabilistic model to perform indoor positioning.

Both RADAR [65] and Horus [66] are based on the raw RSS value, which is an unstable and noisy measurement. To improve the localization accuracy, both temporal and spatial observations are proposed to integrate with the RSS readings [70, 71]. Temporal patterns present the WiFi signal patterns along the path trajectory during walking. This temporal information can be leveraged to correct the localization error of fingerprinting-based localization algorithms. For instance, Walkie-Markie [72] utilizes the WiFi signal sequence in corridors for fingerprint matching. It stores all the RSS patterns in different corridors and compares the increase and decrease of real-time RSS readings with the patterns to infer the location of user. The peak (a predefined high RSS value) in a given signal sequence is also considered for fingerprinting-based localization [73]. Spatial patterns indicate the geographical signal patterns that can be leveraged to constrain the location of user. The WiFi RSS values may display certain unique signatures at some specific indoor locations, such as corridor, open space and office area. Thus, these signatures can be used as WiFi landmark to classify the corresponding regions. MapCraft [52] and UnLoc [74] leverage the unique readings of WiFi APs to reduce the localization error. Furthermore, the order of signal strength from different APs is also proposed for fingerprint matching. HALLWAY [75] employs the RSS orders classifying different rooms. The experimental results have demonstrated that this approach is able to mitigate the influence of signal fluctuation as well as device heterogeneity.

2.1.10 Summary of sensing technologies for indoor localization

The achievable system coverage in indoor environments is normally between $10m^2$ and $1000m^2$ and the required localization accuracy is usually around $1m$ to $5m$. As shown in Figure 2.1, the most suitable technologies for IPSs are RFID, BLE and WiFi.

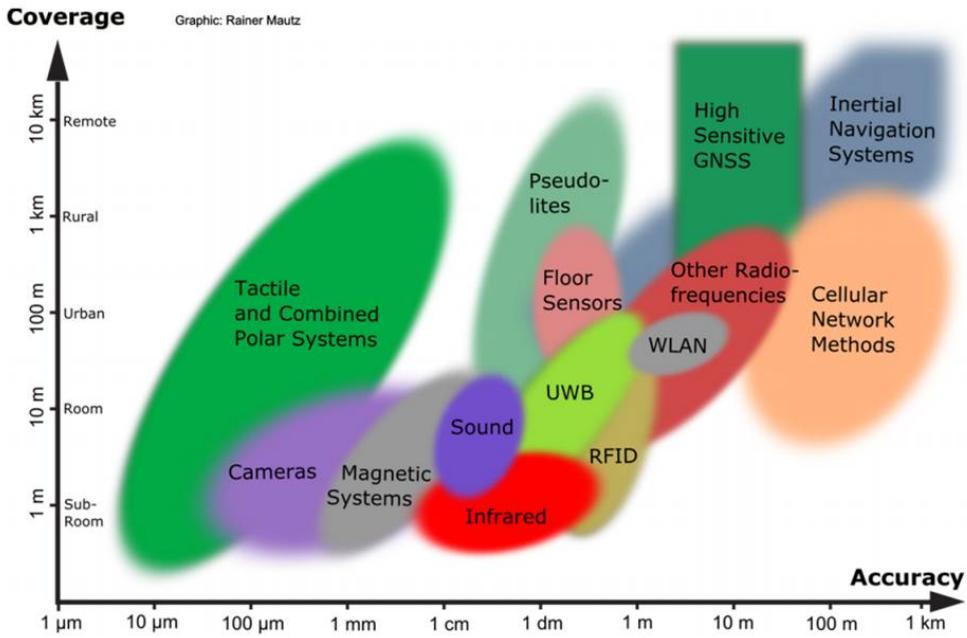


Figure 2.1: Overview of indoor localization technologies in terms of accuracy and coverage [2]

2.2 Overview of Indoor Localization Algorithms

With the booming development of leveraging various wireless communication technologies to provide indoor positioning and navigation, indoor localization algorithms have been extensively studied and a number of approaches have been proposed over the past two decades [6–8]. In general, indoor localization algorithms can be classified into two classes, memoryless localization and tracking. Memoryless localization algorithms employ real-time measurements only to estimate the location of a mobile device. Tracking algorithms leverage both real-time measurements and historical data, which are linked by a dynamic motion model, to infer the location.

In the rest of this section, we will provide an overview of these two categories of indoor localization algorithms as well as their limitations respectively.

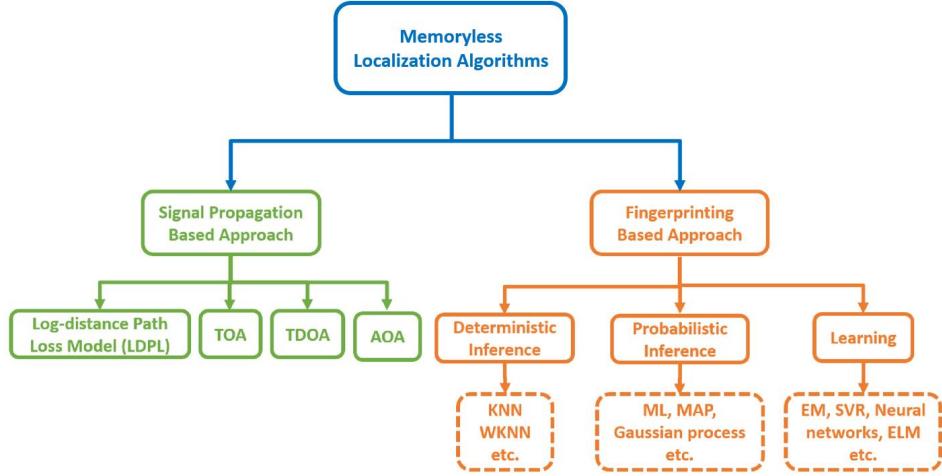


Figure 2.2: Overview of memoryless localization algorithms for indoor LBS

2.3 Memoryless Localization Algorithms

As shown in Figure 2.2, memoryless localization algorithms can be classified into two categories: signal propagation based approaches and fingerprinting based approaches.

2.3.1 Signal propagation based localization algorithms

Signal propagation based localization algorithms calculate the location of mobile devices based on geometrical models. The most widely used model is the Log-Distance Path Loss model (LDPL), which establishes the relationship between the measured RSS and the Radio Frequency (RF) propagation distance [76, 77]. Other geometric models have been utilized to characterize the relationship between signal transmitters and receivers as well. Time of arrival (TOA) algorithm leverages the ranging measurements from multiple transmitters to estimate the location of a mobile device. The speed of light multiplied by the duration of travel is the estimated distance between the transmitter and the mobile device [78]. PinPoint [79] is based on Time of Arrival (ToA). Angle of Arrival (AOA) has been employed for indoor localization [80]. The location of the mobile device is estimated according to the angle of arrival of an RF signal. Cricket [35] adopts Time Difference of Arrival

(TDoA) method to perform indoor positioning. The least squares method can be applied to derive a location estimate.

Several signal propagation based approaches have been investigated in [81]. In general, the average localization accuracy of these IPSs is around 5 meters. The advantage of the signal propagation based approach is that it requires less calibration efforts than the fingerprinting based approach. However, these theoretical models are unable to capture the RSS variations timely due to environmental dynamics. Numerous facts (e.g. multi-path effects, reflection, diffraction and lack of LOS conditions) severely deteriorate the performance of signal propagation based algorithms in complex indoor environments.

2.3.2 Fingerprinting based localization algorithms

Fingerprinting based approaches adopt fingerprint matching as the basic scheme. It consists of two phases: offline calibration phase and online localization phase. In the offline calibration phase, a site survey is performed to record the RSS from multiple APs at many calibration points and their physical coordinates to form a fingerprint. Then the data form a RSS fingerprint database (a.k.a. radio map). During the online localization phase, the location of a mobile device is estimated by matching the observed RSS readings against the offline RSS fingerprint database.

A majority of these approaches leverages RF signals for RSS fingerprinting. To name a few, RADAR [65] and Horus [66] are based on WiFi signal, while LANDMARC [54] utilizes RFID signal; FM radio [82], geomagnetism [83] and GSM signals [84] are also adopted as fingerprints to build up the fingerprint database.

Miscellaneous matching techniques have been incorporated into fingerprinting-based approaches. They can be classified into two categories: deterministic inference, probabilistic inference and learning.

2.3.2.1 Deterministic inference based localization

Deterministic inference matching techniques estimate the location of mobile devices by leveraging predefined voting rules such as KNN and weighted K nearest neighbor (WKNN) algorithms. To be more specific, suppose M most similar reference points have been selected for location estimation. Then, the estimated location of a target $\hat{\mathbf{x}}$ is calculated based on WKNN as follows:

$$\hat{\mathbf{x}} = \sum_{m=1}^M w_m \mathbf{c}_m \quad (2.1)$$

where w_m is the weight for the m th reference point according to certain similarity metrics, while \mathbf{c}_m is the physical location of the m th reference point. KNN is a special case of WKNN when w_m are set to be the same for all reference points. The weighting scheme depends on the distance function that measures the similarity between the real-time RSS vector and fingerprints stored in the database [85]. The most commonly used distance function is Euclidean distance. Other similarity equations such as Manhattan distance, Minkowski distance, Chebyshev distance, Intersection distance, Fidelity similarity and JensenShannon distance are also utilized to measure the similarity between RSS vectors [86]. Comprehensive performance evaluations of different distance and similarity metrics for WiFi fingerprinting based localization can be found in [85, 86].

Deterministic inference matching techniques can be implemented easily. The localization accuracy could be improved when denser reference points are generated. However, additional infrastructure cost will be introduced.

2.3.2.2 Probabilistic inference based localization

These algorithms utilize probabilistic models to estimate the locations of mobile devices. During the offline phase, the signal strength distribution of each AP at each reference location is stored as radio map in the database. During the online phase, with the online measured RSS s , the objective is to find the location \hat{x} that

maximizes the probability $P(x|s)$. According to the Bayes' theorem, the estimated location \hat{x} is inferred as follows:

$$\hat{x}(t) \propto \arg \max_x P(x|s) = \arg \max_x P(s|x) \quad (2.2)$$

Numerous Bayesian inference tools have been adopted to estimate \hat{x} based on equation (2.2), such as Gaussian process [87], maximum a posteriori estimation [88] and maximum likelihood estimation [89]. The problem for probabilistic inference approaches is that they require high computational overhead which is not feasible on mobile devices.

2.3.2.3 Learning based localization

Learning based algorithms follow a data-driven manner, which learn and establish the relationship between input data (RSS values) and target data (locations) without a specific model, to estimate the locations of mobile devices. Machine learning methods, including neural networks [90], Back-propagation (BP) [6], support vector machine for regression (SVR) [91], compressed sensing [92], factor graphs [93], kernel estimation [94, 95], have been utilized to learn the parameters in the training dataset. Most of these methods perform non-linear feature mapping to map the raw input data from input space to feature space to enhance the representations of data. In this way, certain nonlinear relationship is able to be captured instead of using a linear model. It is noteworthy that one of the machine learning algorithms called extreme learning machine (ELM) has attracted a lot of attention in recent years due to its fast learning speed and easy implementation [96, 97]. In *Chapter 3*, our proposed RFID-based IPS, which adopts ELM as the localization algorithm, is able to deliver a better performance in terms of both the efficiency and localization accuracy [17–19]. In addition, online sequential extreme learning machine (OS-ELM), which is able to adapt to various environmental dynamics by its online sequential learning ability, can provide higher localization accuracy more consistently than traditional approaches [14, 22]. The OS-ELM based localization algorithm is presented in *Chapter 5*.

2.3.3 Limitations of fingerprinting-based localization algorithms

Although several published results have shown that the fingerprinting-based localization algorithms outperform signal propagation based approaches [8, 98], there are still several drawbacks that restrain them from large-scale implementation.

First of all, existing fingerprinting-based localization algorithms utilize as many available APs as possible in indoor environments to improve the localization accuracy. In this case, the localization algorithm needs to handle high computational burden due to the usage of large numbers of APs. One major challenge is how to reduce this computational burden while preserving or even improving the localization accuracy. Secondly, it is acknowledged that the fingerprinting approach results in high localization accuracy provided that the testing device is the same as the reference device, but the localization accuracy can be severely degenerated for heterogeneous devices [15]. Due to the proliferation of various types and brands of mobile devices, it is indispensable and urgent to develop a robust location fingerprinting technique so as to provide accurate, reliable and fast indoor positioning services for heterogeneous devices. Moreover, the offline calibrated database is vulnerable to environmental dynamics [13, 14], as the real-time RSS readings collected during online localization phase could deviate from those stored in the offline radio map due to the variation of temperature, humidity, occupancy distribution and multipath effects. Serious localization errors will be introduced if the radio map is not updated adaptively. In addition, the offline site survey process is extremely time-consuming and labor-intensive. Multiple RSS samples need to be measured at numerous calibration points to ensure the localization accuracy. The following sections describe the aforementioned limitations for fingerprinting-based approaches elaborately.

2.3.3.1 Data mining for improving localization accuracy

Certain research works have been conducted to extract the most relevant information from the high dimension signal space caused by massive numbers of APs for the purpose of improving localization accuracy. It is straightforward to leverage as many available APs as possible since they are densely deployed in indoor environments. However, the primary objective of this deployment plan is for the wireless network coverage rather than for the localization purpose. Thus, a portion of the RSS data from APs is redundant. It is necessary to identify the most vital information to enhance the localization performance from the large and noisy signal space. Existing approaches can be classified into two categories: feature extraction and AP selection.

The basic idea of feature extraction is to leverage statistical methods, such as linear discriminant analysis (LDA) and principal component analysis (PCA), to extract the most valuable feature components for further localization [99, 100]. These methods are suitable for linear features. However, the larger variation of WiFi RSS limits the performance enhancement of feature extraction based methods.

Another category is AP selection methods [95, 101, 102]. Due to the channel overlap of APs and large noise of WiFi RSS, a portion of available APs is redundant for localization purpose. The objective of AP selection is to reduce the computational burden and meanwhile preserve or even improve the localization accuracy of the entire IPS, by using only a proper subset of available APs. The basic idea is that N_f APs are selected from a total of N_t available APs according to some importance criterion. Therefore, the computational cost for localization process is reduced since $(N_t - N_f)$ dimensions are discarded. APs with the top N_f discriminative values are involved to construct lower dimensional features for further calculation.

To this end, certain importance criterion has been designed. The existing works mainly focus on developing an individual importance index for each AP, and accordingly selecting those with the highest scores. A common approach is MaxMean proposed in [101], which employs the absolute mean RSS value as the importance index and chooses those with the strongest mean RSS values.

Apart from the RSS value, individual information gain is also used as another importance index to assess the discriminate ability of each AP [102]. Assume that there are a total of N_t APs available in an indoor environment. Each AP ($AP_i, 1 \leq i \leq N_t$) can be treated as a location information source or feature. For fingerprinting-based approaches, suppose WiFi RSSs are collected at p calibration points during the offline phase. Therefore, N_t mean RSS values can be obtained at each calibration point ($C_j, 1 \leq j \leq p$). The InfoGain criterion of each AP is calculated as follows:

$$InfoGain(AP_i) = H(C) - H(C | AP_i)$$

where $H(C)$ is the entropy of the reference points where the RSS from AP_i is unknown; $H(C | AP_i)$ denotes the conditional entropy of the reference points given the RSS of AP_i . Since all the RSS values from N_t APs are collected at reference points, $H(C)$ and $H(C | AP_i)$ are calculated accordingly as follows:

$$\begin{aligned} H(C) &= - \sum_{j=1}^p Pr(C_j) \log Pr(C_j) \\ H(C | AP_i) &= - \sum_v \sum_{j=1}^p [Pr(C_j, AP_i = v) \times \\ &\quad \log Pr(C_j | AP_i = v)] \end{aligned}$$

where $Pr(C_j)$ is the prior probability of reference point C_j . Since there are p reference points in total, we assume $Pr(C_j) = 1/p$ for each reference point. v indicates one possible RSS value from AP_i . $Pr(C_j | AP_i = v)$ is the conditional probability which can be calculated as follows:

$$Pr(C_j | AP_i = v) = \frac{Pr(AP_i = v | C_j) Pr(C_j)}{Pr(AP_i = v)}$$

Since the location of C_j is known, $Pr(AP_i = v | C_j)$ can be calculated directly based on the offline WiFi RSS database.

After the calculation, all APs are sorted in a descending order based on their InfoGain values. Then, top N_f APs with the highest InfoGain values as their discrimi-

nate ability are selected for online localization. Other ($N_t - N_f$) APs are discarded for the reduction of computational burden.

We propose a novel online mutual information (OnlineMI) AP selection strategy that measures the collective discriminate ability among APs based on their mutual information to reduce the computational burden in *Chapter 6*.

2.3.3.2 Device heterogeneity

The device heterogeneity issue occurs when the clients' mobile devices (testing devices) are different from the reference device (device utilized for the offline site survey). Due to the heterogeneous factors of mobile devices, including distinct WiFi chipsets, WiFi antennas, hardware drivers, encapsulation materials, and even operating systems [103, 104], RSS detected by heterogeneous devices at the same location usually has different mean values, and will be translated into different physical locations by the traditional WiFi RSS fingerprinting technique, with the result that localization accuracy is severely degraded [15, 94]. To handle the device heterogeneity issue encountered by the WiFi fingerprinting-based IPS, different schemes were proposed [15, 94, 103, 105–111].

One effective but time-consuming solution is to manually adjust RSS values for distinct testing devices by a linear transformation method [94, 106, 107]. Various transformation functions, such as Kullback-Leibler divergence [94], time-space sampling [106], Gaussian fit sensor model [107], have been leveraged. The main drawback of this approach is that it requires the types of the heterogeneous mobile devices be known in advance such that an offline regression procedure can be conducted to derive pairwise linear relationships. This imposes strict limitations on widespread applications involving a mass of new and unknown mobile devices. Moreover, as pointed out in [94], the linear transformation could not satisfactorily resolve the device heterogeneity issue since a simple linear relationship cannot effectively characterize the difference across mobile devices.

Some calibration-free methods were proposed in [103, 109] to avert the tedious manual RSS calibration procedure for each testing device. Collaborative mapping was employed to estimate a linear mapping function by training online measured RSS values [109]. Unsupervised learning methods such as online regression and expectation-maximization have been leveraged to learn the mapping function [103]. Nevertheless, these methods rely on time-consuming online processing to guarantee localization accuracy. Another way to address the device heterogeneity issue is to define and use alternative location fingerprint instead of absolute RSS values. For instance, signal strength difference (SSD), which leverages the difference of RSS values as a location fingerprint, was proposed in [15, 110]. The main drawback of SSD is the effect of shadowing variation and reduced number of RSS fingerprint vectors. On the other hand, hyperbolic location fingerprinting (HLF) employs the RSS ratio between a pair of APs as a location fingerprint [108, 111]. In [15, 105], the experimental results demonstrated that SSD is better than HLF for heterogeneous devices as a location fingerprint.

We propose a novel location fingerprint in *Chapter 4* to address robustness issue with respect to the device heterogeneity.

2.3.3.3 Performance robustness against environmental dynamics

WiFi RSS is known to be susceptible to various environmental changes, including instant interference, such as opening and closing of doors, moving metal objects, as well as continuous interference, such as variation of temperature, humidity and occupancy distribution [13, 14]. Another source is the multipath effects, which include the reflection, diffusion and diffraction in indoor environments. As a consequence, the real-time RSS samples collected during the online localization phase could severely deviate from those stored in the offline radio map, which leads to serious location errors. We propose an OS-ELM based indoor localization algorithm in *Chapter 5*, which is able to adapt to various environmental dynamics by its online sequential learning ability and provides more accurate and robust indoor LBS consistently than the traditional approaches [14, 22].

2.3.3.4 Time-consuming and labor-intensive calibration

The offline site survey process is time consuming, labor exhaustive and expensive. In order to achieve sufficient localization accuracy, the WiFi RSS fingerprints from multiple APs need to be measured at a huge number of calibration points, which is impractical for large indoor environments (such as shopping mall, stadium and airport).

Several schemes have been proposed to reduce the manual efforts for offline site survey and update the radio map online, including fixed reference anchor methods [54, 112, 113], calibration-free methods [77, 114], learning-based methods [115, 116] and Crowdsourcing methods [72, 117].

Specifically, LANDMARC [54] and LEASE [112], developed an adaptive offset of the RSS variations by employing reference anchors deployed at known fixed locations with real-time RSS observations. Nonetheless, these approaches require a very dense deployment of reference anchors to construct the radio map accurately. In [113], self-made WiFi anchors are introduced to obtain real-time RSS observations and Geography Weighted Regression (GWR) is adopted for online radio map construction to reduce the workload for offline site survey. It is noted that all these methods still require extra hardware to be deployed and are infeasible for large-scale implementation. A calibration-free method is presented in [77] which uses an indoor radio propagation model for online radio map construction to remove the offline site survey process. Nevertheless, the simple log-distance path loss model cannot describe the complex RSS distribution precisely. In [114], the idea of employing RSS data among APs to establish a radio map and using the GPR with Log-Distance path loss model for RSS modeling is introduced. However, they fail to modify the AP firmware to realize the idea due to technical difficulty and put a wireless monitor beside each AP instead. As a result, extra devices are still needed.

Several learning-based approaches are also introduced to reduce the number of reference anchors to be deployed [115, 116, 118]. LEMT [115] performed radio map adaptation by training the functional relationship between each location and its

neighboring locations based on nonlinear regression analysis and model tree method, since neighboring locations have highly correlated RSS characteristics generally. The drawback of LEMT is that the process of building huge numbers of trees in each RSS sniffing period is time-consuming, which makes it difficult for real-time application. Other learning techniques such as multiview learning [118] and manifold alignment [116] are also utilized to transfer RSS information across different times and devices. Nevertheless, they still need to collect certain numbers of offline RSS fingerprints as label data for learning purposes.

Crowdsourcing methods, which employ the full sensing capabilities of mobile devices, are introduced to reduce the efforts for radio map construction as well [72, 117]. Zee [117] utilized inertial measurement unit (IMU), comprised of accelerometers, gyroscopes and magnetometers, and RSS reading from the mobile devices to build up a radio map. Walkie-Markie [72] used landmarks, such as turns, escalators and elevators, to enhance crowdsourcing performance. Nevertheless, extra user intervention is needed for these approaches and continuous IMU monitoring will consume a lot of mobile device's battery, which is an impractical solution.

In order to overcome the aforementioned issues, we propose, WinIPS, a WiFi based non-intrusive indoor positioning system that enables automatic online radio map construction and adaptation for calibration-free indoor localization in *Chapter 7*.

2.4 Tracking

Tracking algorithms leverage both real-time measurements and historical data, which are linked by a dynamic motion model, to infer the location of mobile device. It can be classified into two categories: Pedestrian Dead Reckoning (PDR) based tracking approaches and learning based approaches.

2.4.1 Pedestrian Dead Reckoning (PDR) based tracking

Pedestrian Dead Reckoning (PDR) is the most widely used tracking algorithm for indoor positioning and navigation [49, 74, 119]. It uses the on-board IMU sensor

to estimate the current position of the mobile device according to the previous position, step length and walking direction of the user. The step length and the number of steps a user has walked can be estimated according to the accelerometer readings. The direction of users' movement can be derived from the gyroscope or the compass measurements. Therefore, a simple linear motion model can be constructed as follows:

$$\mathbf{X}_t = \mathbf{X}_{t-1} + length_t \begin{bmatrix} \sin(\theta_t) \\ \cos(\theta_t) \end{bmatrix} \quad (2.3)$$

where \mathbf{X} is the estimated location of mobile device at time step t , $length_t$ is the step length and θ_t is the walking direction at time step t . PDR is able to provide accurate location estimation in a short range with a precise initial location estimation [49]. Nonetheless, it will drift along with the walking distance due to vibrations and inherent noise in the continuous coarse measurements. In order to eliminate these accumulative errors, several recent works such as fusion with landmarks [74] and WiFi fingerprinting algorithms [119] have been proposed to improve the performance of PDR. Ranging measurement provided by portable ultrasound range sensors is also integrated with PDR using particle filter to mitigate the drift error with walking distance caused by the integration of gyroscope data in [50]. A Kalman filter is employed for the fusion of the PDR approach and active RFID IPS [51]. Furthermore, a novel PDR localization algorithm with iBeacon corrections using extended Kalman filter is proposed in [39]. All of these sensor fusion algorithms are based on directed graphical models. Meanwhile, fusion methods based on undirected graphical model such as a conditional random field (CRF) are also proposed to improve the localization accuracy of the entire IPS [52].

2.4.2 Learning based tracking

Learning based tracking algorithms construct and train a motion model using historical data and track the location of user with real-time measurements. The historical data can be treated as the prior knowledge in a probabilistic model with a set of

parameters Θ known a priori. The objective is to estimate Θ by training the dataset and refine it with real-time measurements. The maximum likelihood estimate $\hat{\Theta}_{ML}$ or maximum a posteriori estimate $\hat{\Theta}_{MAP}$ derived from Expectation Maximization (EM) approach [120] has been widely utilized in learning based tracking approaches. It is an iterative algorithm that alternates between calculating the expectation of the log-likelihood function given the parameters (*E step*), and optimizing the parameters that maximize the expected log-likelihood (*M step*). Let the expected complete data log likelihood in the *i*th iteration be given as follows:

$$Q(\Theta, \Theta_{i-1}) = E_{x(t)|Z(t), \Lambda, \Theta_{i-1}} \{ \log p(x(t), Z(t), \Lambda | \Theta) \} \quad (2.4)$$

where Λ represents the prior knowledge on $x(t)$, $Z(t)$ denotes the measurements at timestamp t . The objective of *E step* is to calculate $Q(\Theta, \Theta_{i-1})$, while *M step* will optimize $Q(\Theta, \Theta_{i-1})$ with respect to Θ :

$$\Theta_i = \arg \max_{\Theta} Q(\Theta, \Theta_{i-1}) \quad (2.5)$$

EM method is able to estimate the parameters in a predefined motion model effectively. However, the complex movement of occupants in indoor environments cannot be described by a simple motion model comprehensively. The performance of EM learning based tracking approaches will be severely degraded if there is an incorrect assumption in the model. As an alternative, model-free learning based tracking algorithms are able to estimate the parameters without explicit assumptions. For instance, Conditional Random Field (CRF) is employed in [119] to maximize the conditional probability of states given observations without explicitly specifying the distributions. It allows us to capture correlations among observations over time, and to express the extent to which observations support not only states but also state transitions. The high computation cost is the bottleneck for learning based tracking algorithms.

2.5 Integration of IPS and GPS

As presented in the previous sections, extensive research and studies have been conducted in IPS, and some feasible solutions have been proposed. However, the majority of existing research adopts the common assumption that the operating environment is either indoor or outdoor and is known, which does not necessarily hold in reality. Some areas adjacent to buildings (covered corridor, connections between buildings) or semi-open buildings (parking garage) have partial characteristics of both indoor and outdoor environments. In this case, a sole dependence on GPS or IPS is unable to deliver a precise indoor-outdoor (IO) detection. Meanwhile, an accurate and effective IO detection scheme provides basic but critical information for upper-layer applications to serve individual users, e.g., mobile applications leveraging reliable IO status to give better services and alleviate battery consumption. Furthermore, an accurate and effective IO detection scheme could enable a seamless transition between outdoor and indoor LBSSs.

In this section, we present an overview of existing IO detection schemes; one leverages the availability of GPS signals as the indicator to determine the status of indoors or outdoors; another employs various sensors equipped on mobile devices to infer the IO status.

2.5.1 GPS-Based IO detection

Generally, GPS provides reliable and outstanding localization and navigation services for users in outdoor environments when more than four visible satellites are available. On the contrary, since the line-of-sight paths between the mobile device and satellites are usually blocked in indoor environments, the performance of GPS is jeopardized. According to these facts, the localization accuracy of GPS or the availability of GPS signals is leveraged as an indicator to determine the status of indoors or outdoors [76, 121–123].

The GPS-based IO detection suffers from several drawbacks. First of all, its classification accuracy is not reliable, especially in some indoor environments with large windows. The reason is that the number of visible satellites in this area is greater than that in other indoor environments; therefore, the status of users will be misclassified into an outdoor environment if the availability of GPS signals is the sole indicator for IO detection. Secondly, the bootstrap process of the GPS module takes around one minute in general. Its performance is largely deteriorated during this slow process. Furthermore, GPS-based IO detection is not an energy-efficient method. The GPS module is the most power-hungry sensor in a mobile device, which is not worth being turned on as an IO hint all the time, from the energy saving perspective. In particular, GPS will consume much power in indoor environments since it needs to conduct GPS satellite scanning continuously until a sufficient number of satellites have been connected to perform localization.

In summary, GPS-based IO detection is vulnerable, ineffective, energy consuming and impractical.

2.5.2 Lightweight on-board sensor-based IO detection

Another category is to leverage various sensors equipped on the mobile device to determine whether the environment status is indoor, outdoor or an intermediate zone [124, 125]. IODetector [124] mainly utilizes three on-board sensors, light, cell and magnetic field sensors, to detect the IO status. According to their observations, high light intensity indicates outdoor environments, and low light intensity represents indoor environments in general; the fluctuation of magnetic field intensity is higher in indoor than outdoor environments due to severe disturbances from steel structures and the chaotic electromagnetic fields generated by electric devices in indoor environments; the cell RSS of the user’s mobile device drops rapidly from outdoor to indoor environments because of the attenuation of walls and ceilings. These three detectors provide their individual estimates and corresponding confidence in those estimates. Then, IODetector aggregates these results, and the final

estimation is the status with the best overall confidence in estimation. A first-order hidden Markov model is also employed to realize a stateful IODetector in [124], although only a slight enhancement was achieved compared to the stateless one. The main drawback of IODetector is that it leverages hard thresholds for each sensor feature to determine indoor or outdoor environments. The accuracy of the estimation across different devices and environments is negatively impacted.

Besides the three sensors employed in IODetector [124], other sensor features, including sound intensity from the microphone, temperature from the battery thermometer and the proximity sensor are utilized for IO detection in [125]. Furthermore, co-training as a semi-supervised learning method was adopted to improve the IO estimation accuracy across different devices and environments.

2.6 Conclusion

In this chapter, we have surveyed some major sensing technologies and localization algorithms for indoor LBS. Indoor localization technologies based on camera, magnetic field, IR, ultrasound, UWB, IMU, RFID, BLE and WiFi have been reviewed. The indoor localization algorithms can be classified into two classes, memoryless localization approaches and tracking approaches. The limitations of these approaches were elaborated. For memoryless localization algorithms, signal propagation based approaches are unable to capture the RSS variations timely due to environmental dynamics in complex indoor environments. Fingerprinting-based localization algorithms suffer from several problems (e.g. high computational complexity, device heterogeneity issue, time-consuming and labor-intensive calibration process and vulnerability to environmental dynamics) restrain them from large-scale implementation. On the other hand, high computation cost is the major bottleneck for tracking algorithms. In addition, existing IO detection schemes and solutions for the integration of IPS and GPS were also presented in this chapter.

Chapter 3

Localization Algorithms for Indoor Positioning

In this chapter, we propose three localization algorithms, namely Weighted Path Loss (WPL) [18], Extreme Learning Machine (ELM) [17] and integrated WPL-ELM [19], that can be directly utilized for all RF based IPS. As presented in Section 3.1, WPL is a path loss model-based approach which does not require offline site survey procedure and provides a reasonably accurate location estimation of the target effectively. In Section 3.2, we propose an ELM based localization algorithm which can provide higher localization accuracy than other existing fingerprinting based approaches. The integrated WPL-ELM approach which combines the fast estimation of WPL and the high localization accuracy of ELM is presented in Section 3.3. This integrated approach provides higher localization efficiency and accuracy than existing approaches, e.g. the LANDMARC approach [54] and the support vector machine for regression (SVR) approach [91]. Extensive experiments have been conducted on our RFID based IPS [17] to evaluate the performance of proposed algorithms comprehensively. These three localization algorithms have been also implemented on our other RF based IPSs including WiFi based IPS [20] and iBeacon based IPS [25].

3.1 Weighted Path Loss (WPL)

3.1.1 Indoor path loss model

The most commonly used path loss model for indoor environments is the ITU Indoor Propagation Model [126]. It provides a relation between the total path loss PL (dBm) and distance d (m) as:

$$PL = 20 \log(f) + 10\alpha \log(d) + c(k, f) + X_\sigma \quad (3.1)$$

where f (MHz) is the radio frequency, c is an empirical floor loss penetration factor, k is the number of floors between transmitter and receiver, α is the path loss exponent, and X_σ is a noise which is typically assumed to be Gaussian with zero mean and standard deviation σ . The signal propagation conditions depend on different indoor environments due to multipath fading and shadow fading. Therefore, the path loss exponent α which ranges from 2 to 4, depending on the layout of indoor environment, should be determined empirically. After summing with the constant term, the indoor path loss model can be further expressed as:

$$PL(d) = PL_0 + 10\alpha \log(d) + X_\sigma \quad (3.2)$$

with PL_0 being the reference path loss coefficient.

3.1.2 Methodology of WPL

Suppose that we have A sensors and B targets. Each sensor can pick up the signal strengths of all B targets. In order to calculate the estimated location of each target, we define the signal strength of the j th target received at the i th sensor as s_{ij} , where $i \in [1, 2, \dots, A]$, $j \in [1, 2, \dots, B]$. The real position of the i th sensor is defined as (x_i, y_i) . Based on the path loss model defined in the previous section, the signal

strength s_{ij} can be expressed as:

$$s_{ij} = PL(d_{ij}) = PL_0 + 10\alpha \log(d_{ij}) + X_\sigma \quad (3.3)$$

Therefore, based on (3.3), the distance between the j th target and the i th sensor can be calculated by:

$$d_{ij} = 10^{\frac{s_{ij} - PL_0 - X_\sigma}{10\alpha}} \quad (3.4)$$

The distances between these A sensors and the j th target can be expressed as a d vector. The weighting factor of the i th sensor with respect to the j th target is defined as:

$$w_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{i=1}^A \frac{1}{d_{ij}}} \quad (3.5)$$

The unknown location coordinate (u_j, v_j) of the j th target is obtained by

$$(u_j, v_j) = \sum_{i=1}^A w_{ij}(x_i, y_i) \quad (3.6)$$

3.1.3 Experimental results and performance evaluation

In order to evaluate the performance of WPL, extensive experiments have been conducted on our RFID based IPS [17]. The test-bed is set up in the Internet of Things Laboratory in the School of Electrical and Electronic Engineering, Nanyang Technological University. The area of the test-bed is around 110 m² (6.4 m × 17.1 m). As shown in Figure 3.1, there are 19 RFID sensors installed in the room. The positions of 9 tracking tags (targets) and the sensors are also shown in Figure 3.1.

Before the performance evaluation, we define the reasonable range of received signal strength the sensor can pick up from the tags firstly. We put one tag directly next to one sensor in order to estimate the maximum signal strength a sensor can receive from a tag. After collecting 3600 RSSI samples in one hour, we find that the average signal strength received by the sensor is around -42dBm with the standard deviation of 2.8dBm. We also put one tag at the right lower corner and one sensor at the left

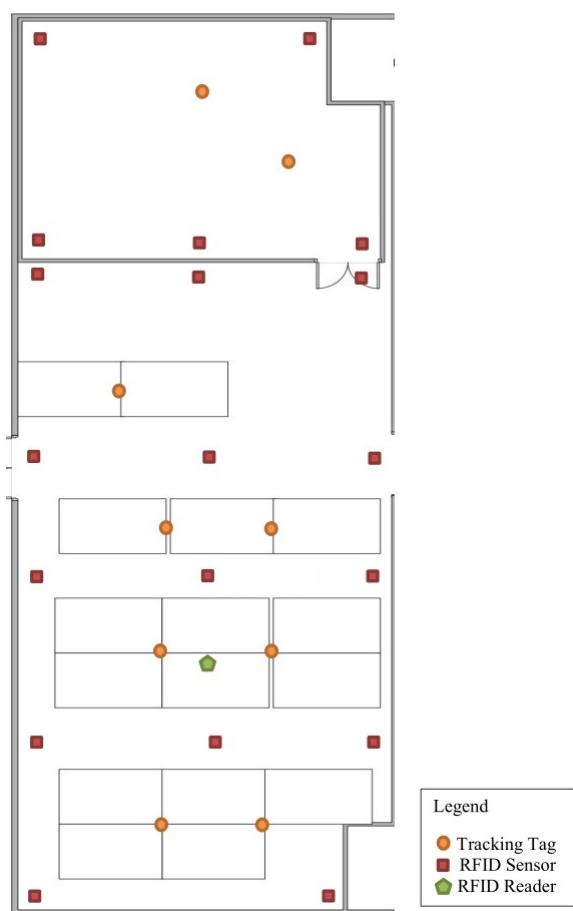


Figure 3.1: Placement of RIFD reference tags, tracking tags, sensors and reader in Experiment I

upper corner of the room (as shown in Figure 3.1) in order to estimate the minimum signal strength a sensor can pick up from a tag. The average signal strength received by the sensor is around -98dBm with the standard deviation of 5.6dBm. Therefore, we define the reasonable range of received signal strength to be from -40 to -100dBm for our system.

In order to evaluate the performance of the proposed localization algorithms, the distance error is used to measure the localization accuracy of the system. We define the location estimation error e to be the distance between the real location coordinates (x_0, y_0) and the estimated location coordinates (x, y) ,

$$e = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (3.7)$$

3.1.3.1 Estimation of the path loss exponent α in WPL

The WPL approach largely depends on the path loss exponent α . Therefore, an experiment is conducted to measure the RSSI values of different distances from a sensor in order to find out the relationship between RSSI and distance. As shown in Figure 3.1, 7 reference tags located on the left side and the sensor at the left upper corner of the test-bed are selected in this experiment since there are relatively clearer line-of-sight between the sensor and these tags. We measure the signal strength at 1.50m, 3.45m, 5.06m, 7.64m, 10.64m, 13.54m and 17.09m. At each location, 3000 RSSI samples are collected in 1 day. Figure 3.2 shows the average signal strength of the collected RSSI data at various locations. Based on the collected data and the path loss model as introduced in Equation 3.3, we use a curve fitting method to construct the relationship between RSSI and distance as:

$$PL(d_i) = -52.40 - 10 \times 3.58 \times \log(d_i) \quad (3.8)$$

i.e., the path loss exponent α is taken as 3.58 and the reference path loss coefficient PL_0 as -52.40dBm. We assume that α and PL_0 remain unchanged during the entire testing period.

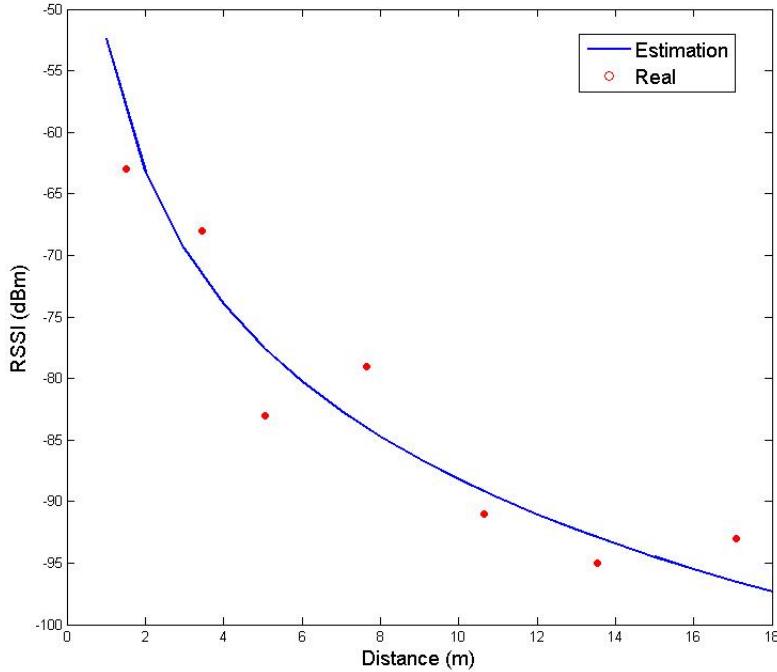


Figure 3.2: Relationship between RSSI and distance

Table 3.1: Localization accuracy statistics

Approach	Average
	Localization Accuracy (m)
LANDMARC	2.642
Enhanced LANDMARC	1.990
WPL	1.651

3.1.3.2 Evaluation on localization accuracy of WPL

We keep collecting data of the signal strength of the 9 tracking tags from the 19 sensors for 7 days in Experiment I in order to evaluate the localization accuracy of the WPL approach. Since WPL is classified as a model-based approach, the performance of LANDMARC [54] and enhanced LANDMARC [58] are chosen for comparison. Furthermore, the LANDMARC and the enhanced LANDMARC use the weighted k-nearest neighbour algorithm to estimate the location of the tracking tags, and we choose k as the maximum number of reference tags in order to optimize the localization accuracy of these methods.

Based on the RSSI samples of all the tracking tags we collected during Experiment I, the localization performance comparison between the LANDMARC, enhanced

LANDMARC and WPL are presented in Table 3.1 and Figure 3.3. As shown in Table 3.1, the average localization accuracy by using LANDMARC, enhanced LANDMARC and WPL is respectively 2.642m, 1.990m and 1.651m. WPL enhances the precision of localization accuracy by 38% over LANDMARC and 17% over enhanced LANDMARC. To conclude, out of three approaches, WPL provides the highest localization accuracy.

The rationale of the outstanding performance of WPL is that it determines the weight of each sensor based on the distance estimated by a dedicated path loss model in that specific indoor environment instead of using the noisy RSS measurements. Before performing the WPL approach, two key parameters: the path loss exponent α and the path loss coefficient PL_0 are carefully estimated based on offline testing data, as elaborated in Section 3.1.3.1. Numbers of RSSI samples at multiple distances from the transmitter are collected through one day to fully characterize the relationship between the physical space and the signal space in that region. On the other hand, both LANDMARC and enhanced LANDMARC assign the weight based on the Euclidean distance between the instantaneous RSS vectors of each sensor and the target. These raw RSS readings are coarse and unstable due to various interference in complex indoor environment as elaborated in Chapter 2. Thus, the weight of WPL, which is determined based on reciprocal of the estimated distance, is more reliable for ranking the importance of each sensor than the raw RSS readings adopted LANDMARC and enhanced LANDMARC, resulting in better localization performance.

3.1.3.3 Evaluation on robustness of WPL

In order to evaluate the robustness of WPL when the number of sensors is reduced, we turn off 5 sensors in the test-bed during Experiment II. As shown in Figure 3.4, we keep collecting data of the signal strength of the 9 tracking tags from the 14 sensors for 7 days in this experiment. Then we compare the performance of LANDMARC, enhanced LANDMARC and WPL using 14 sensors with Experiment I database, where 19 sensors are used.

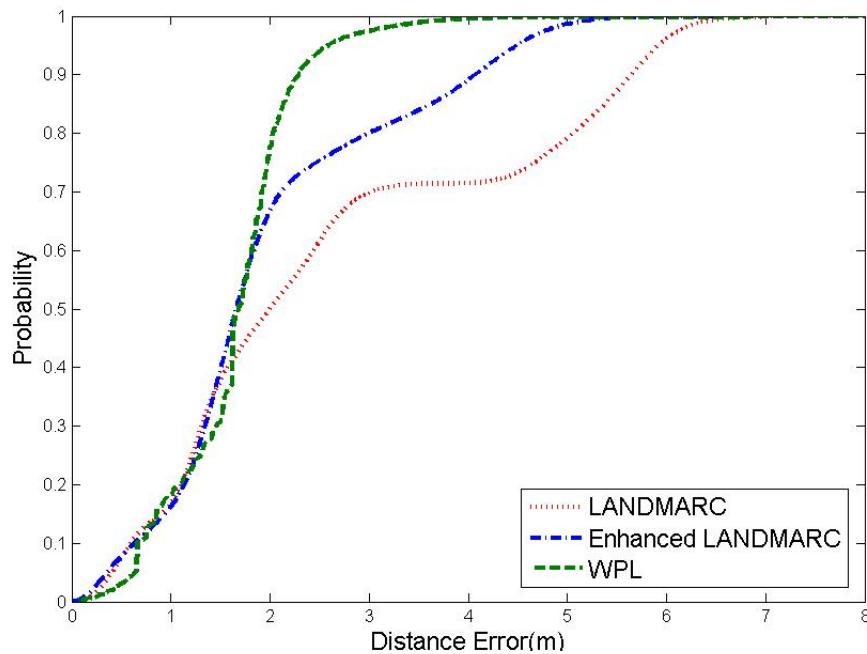


Figure 3.3: Cumulative percentile of error distance for different methods

Table 3.2: Comparison between WPL and other methods

No. of Sensor		19		14	
Approach		Average	Variance	Average	Variance
LANDMARC		2.642	0.108	2.973	0.191
Enhanced LANDMARC		1.990	0.192	2.214	0.420
WPL		1.651	0.290	1.782	0.337

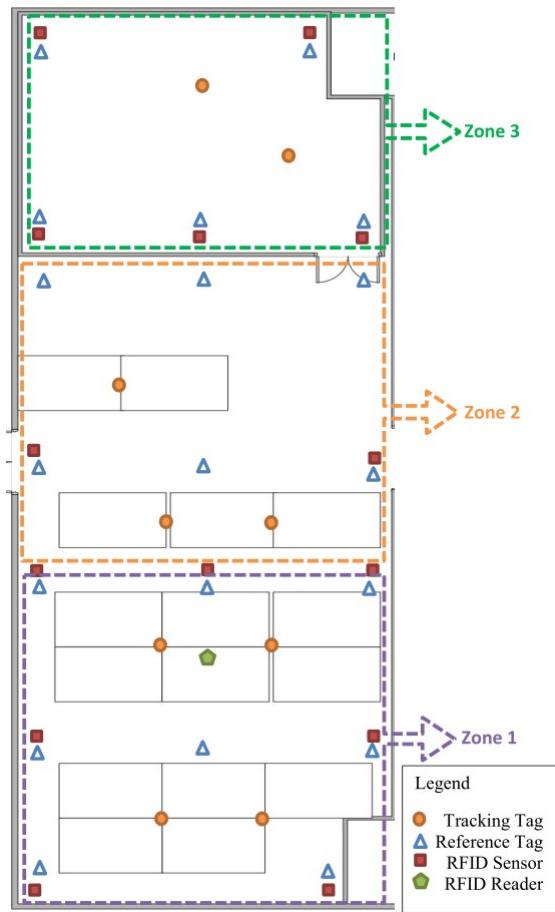


Figure 3.4: Placement of RFID reference tags, tracking tags, sensors and reader in Experiment II and Experiment III

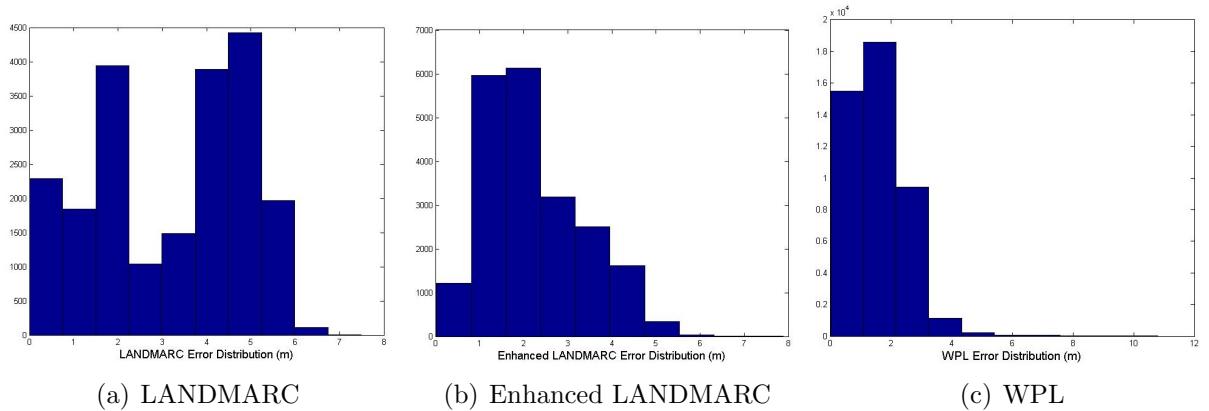


Figure 3.5: Comparison of distance error distribution for different methods

The mean and variance of the location estimation error of the three approaches with different numbers of sensors are presented in Table 3.2. Figure 3.5 demonstrates the distance error distribution of the three approaches. As shown in Table 3.2, the localization performances of all the three approaches become worse with lower density of sensors. However, it can be observed that the localization accuracy of WPL still remains the best among the three approaches when 5 sensors are removed from the test-bed. Under this circumstance, WPL still enhances the precision of localization accuracy by 40% over LANDMARC and 19% over enhanced LANDMARC. Compared with the results when 19 sensors are used, the localization accuracy of LANDMARC, enhanced LANDMARC and WPL decreased by 13%, 11% and 8%, respectively. The performance decay of WPL is the smallest. As shown in Figure 3.5, the distance error distribution of LANDMARC in Figure 3.5(a) and enhanced LANDMARC in Figure 3.5(b) are much more scattered, while that of WPL is mainly limited within 2.7 m as shown in Figure 3.5(c).

Thus we can conclude that WPL remains more reliable and robust when the number of sensors is reduced.

3.2 Extreme Learning Machine (ELM) based Localization Algorithm

3.2.1 Motivations

A large body of existing IPSs apply the fingerprinting-based approaches, which usually involve two phases: an offline training phase and an online localization phase. During the offline training phase, a site survey dedicated to collecting the RSS fingerprints from different APs at some known locations is performed in the indoor environment and consequently, a RSS fingerprint database is built up. During the online localization phase, when a user sends a location query containing his or her current RSS fingerprint, the location of the user will be estimated by matching the

measured fingerprint with the fingerprints stored in the database, and the location associated with the matching fingerprint will be returned as his or her location estimate.

Machine learning and neural network techniques have been leveraged in fingerprinting-based approaches. However, those techniques face challenging issues such as slow learning speed, intensive human intervene, and poor computational scalability. Extreme Learning Machine (ELM) as emergent technology has attracted attention recently [127]. It has been proved to provide good generalization performance at an extremely fast learning speed [96].

WiFi based IPS [128] by adopting the ELM approach have been provided to give a better performance in terms of both the efficiency and the localization accuracy.

3.2.2 Preliminaries on ELM

ELM is a kind of machine learning algorithm based on a Single-hidden Layer Feed-forward neural Network (SLFN) architecture. The output with L hidden nodes of a SLFN can be represented as:

$$\mathbf{y}_N(\mathbf{x}) = \sum_{i=1}^L \beta_i g_i(\mathbf{x}) = \sum_{i=1}^L \beta_i G(\mathbf{a}_i, b_i, \mathbf{x}) \quad (3.9)$$

where a_i , b_i are the weights and bias connecting the input nodes and the i th hidden node, β_i are the output weights connecting the i th hidden node and the output nodes, and $G(a_i, b_i, x)$ is the activation function which gives the output of the i th hidden node with respect to the input vector x .

It is shown in [129] that a SLFN with at most N hidden nodes and with almost any nonlinear activation function can exactly learn N distinct observations. Given N arbitrary distinct training samples $(\mathbf{x}_j, \mathbf{t}_j), j = 1, 2, \dots, N$, by substituting \mathbf{x} with \mathbf{x}_j in (3) we obtain

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (3.10)$$

where

$$\mathbf{H} = \begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_1) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_1) \\ \vdots & \dots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_N) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_N) \end{bmatrix}_{N \times L}, \quad (3.11)$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m}. \quad (3.12)$$

In the above, \mathbf{H} is the hidden layer output matrix of ELM; the i th column of \mathbf{H} is the i th hidden node's output vector with respect to inputs x_1, x_2, \dots, x_N , and the j th row of \mathbf{H} is the output vector of the hidden layer with respect to the input vector of x_j .

Unlike the traditional training algorithms for neural networks which need to adjust the input weights and hidden layer biases, [96] has proved that the parameters of SLFN can be randomly assigned provided that the activation function is infinitely differentiable. Therefore, the hidden layer output matrix \mathbf{H} remains unchanged once these parameters are randomly initialized. To train a SLFN is simply equivalent to finding an optimal solution $\boldsymbol{\beta}_{LS}$ of (4) as:

$$\|\boldsymbol{\beta}_{LS} - \mathbf{T}\| = \min_{\boldsymbol{\beta}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|. \quad (3.13)$$

The optimal solution of the above equation can be found as $\boldsymbol{\beta}_{LS} = \mathbf{H}^\dagger \mathbf{T}$, where \mathbf{H}^\dagger is the Moor-Penrose generalized inverse of \mathbf{H} .

3.2.3 Methodology of ELM

The ELM approach considers the localization problem as a regression problem. It consists of an offline phase and an online phase. During the offline phase, suppose that P reference points will be selected and Q RSSI samples will be collected for each point. Each RSSI sample is denoted as $((X_{pq}, Y_{pq}), RSS_{pq})$, $p \in [1, 2, \dots, P]$, $q \in [1, 2, \dots, Q]$. The RSSI readings are the training inputs and the corresponding

location vectors (X_{pq}, Y_{pq}) are the training targets of ELM. The hard-limit transfer function is usually adopted as the activation function. The training process of ELM can be conducted in the following three main steps:

Step 1: Randomly assign values to hidden node parameters.

Step 2: Calculate the hidden layer output matrix \mathbf{H} .

Step 3: Calculate the output weight $\boldsymbol{\beta}$ by $\boldsymbol{\beta} = \mathbf{H}^\dagger \mathbf{T}$, where \mathbf{H}^\dagger is the Moor-Penrose generalized inverse of \mathbf{H} .

During the online phase, the only thing we need to do is to feed the RSS vector into the ELM model. The output given by ELM is the estimated location of the target.

3.2.4 Experimental results and performance evaluation

Since ELM is a fingerprinting-based localization algorithm, we build up a historical RSSI fingerprints database for ELM offline training during Experiment II in the first place. As shown in Figure 3.4, we keep collecting data of the signal strength of the 19 reference tags and the 9 testing tags from the 14 RFID sensors for 5 days during the offline phase. 318500 RSSI samples for each tags are obtained in this experiment.

For each of 19 reference tags, 5000 RSSI samples are randomly chosen as training fingerprints from Experiment II database. Here we choose 5000 RSSI samples of each reference tag for the ELM offline training process, considering the limitation of the number of input samples in ELM. When the number of input samples is too large, unnecessary hidden nodes parameters will be introduced and cause ELM to be unstable and overfitted easily. In our system, we found that 5000 input variables (RSSI samples in our case) is appropriate for ELM training.

Besides the number of input variables, another parameter that could affect the localization accuracy of ELM is the number of hidden nodes in the ELM hidden layer. The localization accuracy of ELM can be improved with the increase of the

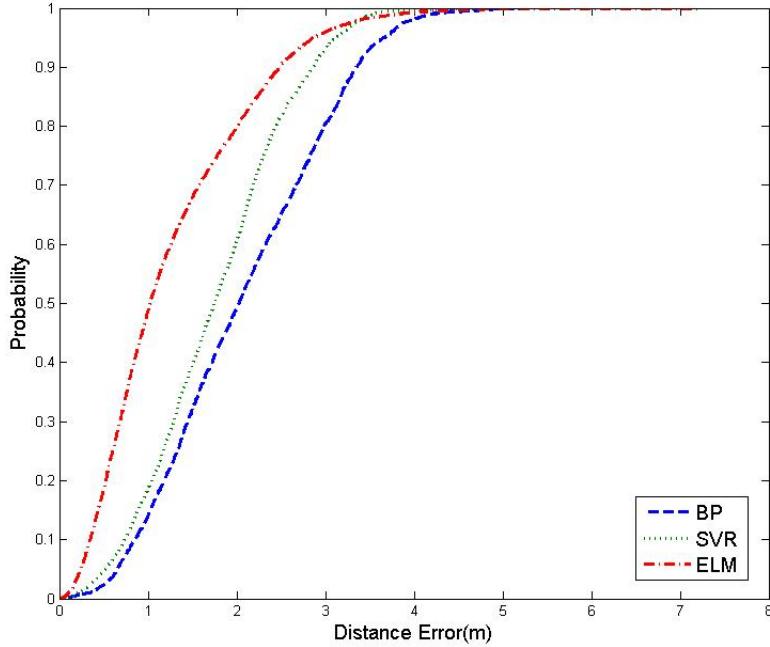


Figure 3.6: Cumulative percentile of error distance for different methods

Table 3.3: Comparison between ELM and other methods

Approach	Training Time (s)	Testing Time (s)	Accuracy (m)
BP	97200	0.007	2.084
SVR	1402.887	0.013	1.769
ELM	248.026	1.825	1.198

number of hidden nodes in the ELM hidden layer. However, both training time and testing time also increase. For instance, the ELM approach with 2500 hidden nodes enhances the precision of localization accuracy by 33% over WPL but the testing time is as long as 1.937s, which is too long for real-time localization. Thus, there is a tradeoff between the localization accuracy and the testing time when applying the ELM approach. Based on our evaluation, we choose 2000 hidden nodes in the ELM hidden layer in our system.

Another two machine learning algorithms, Back-propagation (BP) algorithm and support vector machine for regression (SVR) algorithm, will be chosen in comparison with ELM. After building up the ELM model, we evaluate the performance of these three approaches based on the RSSI samples of the 9 tracking tags from

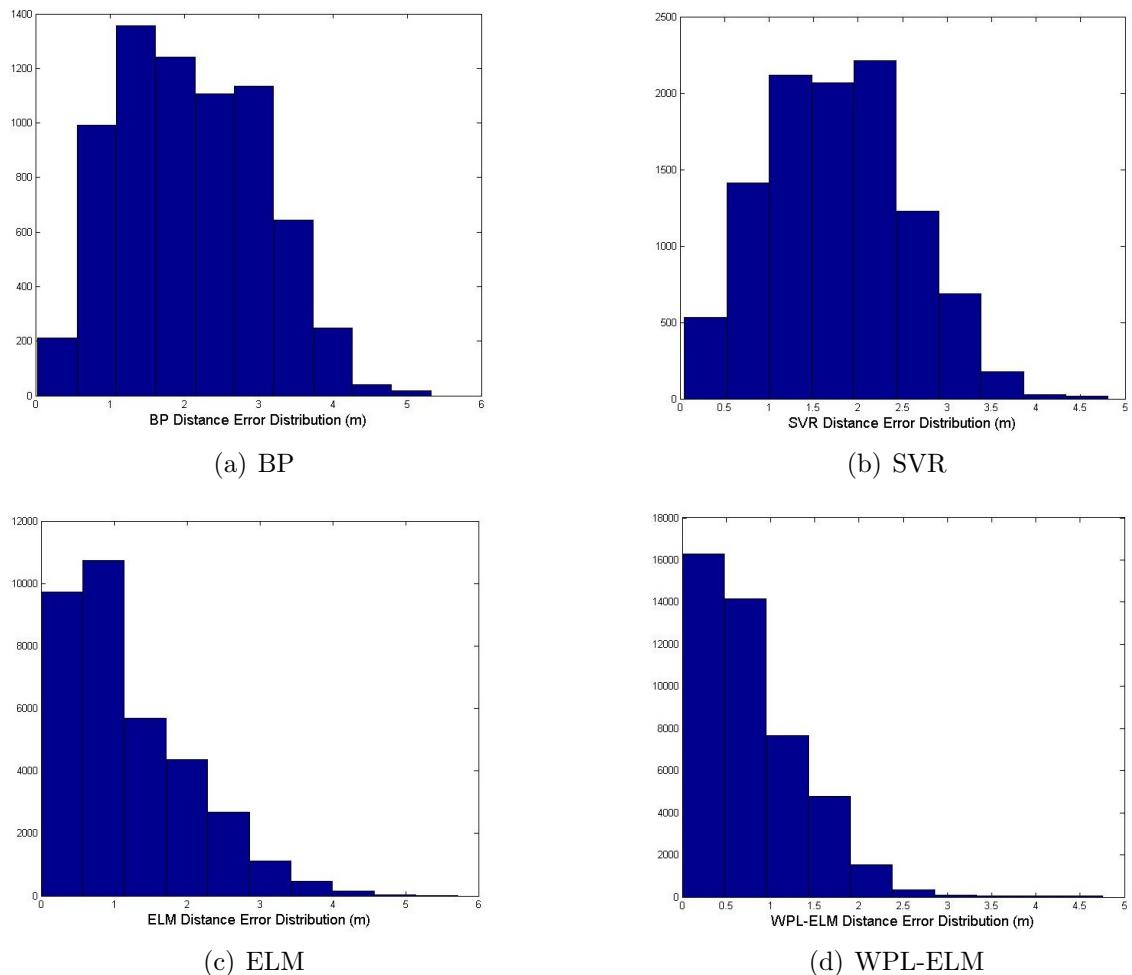


Figure 3.7: Comparison of distance error distribution for different methods

Experiment II database. The cumulative percentile of error distance for the three approaches is shown in Figure 3.6. Table 3.3 demonstrates the performance comparison between the three approaches in terms of the training time, testing time and average localization accuracy of the samples.

As observed from Table 3.3, ELM has an enormous advantage in training time and learning speed. It learns up to 391.94 times and 5.66 times faster than BP and SVR respectively. On the other hand, BP and SVR require shorter testing time than ELM. The testing time of 1.825s is really a drawback of ELM because it will introduce certain delay in real-time localization of the tracking tags. It can be seen in Table 3.3, the average localization accuracy of all 9 tracking tags by using BP, SVR and ELM is 2.084m, 1.769m and 1.198m. ELM enhances the precision of localization accuracy by 43% over BP and 32% over SVR respectively. Figure 3.7 demonstrates the distance error distribution of the three different approaches. The distance errors of ELM as shown in Figure 3.7(c) are mostly within 3m. In contrast, the distance error distribution of BP in Figure 3.7(a) and SVR in Figure 3.7(b) are much more scattered. ELM has tremendous advantages in offline training time and online localization accuracy due to its faster learning speed and better generalization performance compared with BP and SVR. The most critical issue of BP is the time-consuming learning process since it is a gradient-based learning algorithm. Moreover, when its learning rate is large, BP becomes unstable and may stop at a local minima during learning process that will definitely impact its localization performance [96]. The major difference between SVR and ELM is the computational complexity. For SVR, the approximation bias still requires in the optimization constraints to absorb the system errors because it may not have universal approximation capability [130]. On the other hand, all the parameters of the ELM mapping are randomly generated and the bias may not be required in its output nodes. Since ELM has less decision variables to be determined, it has computational superiority compared with SVR. This is the main reason why the training process of ELM is faster than SVR. Researcher in [130] also indicates that the feasible solution space of SVR is a subset of ELM feasible solution space if same kernels are leveraged,

which may cause the lower localization accuracy of SVR.

It is worth noting that the requirement of large number of RSSI samples and high density of reference tags for offline training is a limitation for RFID fingerprinting-based localization algorithms. According to the experimental results, the localization accuracy decreases 26% when 75% of the reference tags are used for training. The mean localization error reaches to 2.137 m (43.9% worse) when only half of the reference tags are leveraged. One feasible solution to reduce the high density deployment of reference tags is to upgrade the software of RFID sensors to enable them to collect the RSS emitted from other sensors. In this manner, all the sensors are becoming online reference points therefore no reference tags are required. On the other hand, we found that the influence of number of RSSI samples collected at each reference tag on localization accuracy is much smaller. The localization accuracy decreases only 5.6% when the numbers of collected samples reduced by half.

3.3 Integrated WPL-ELM

3.3.1 Motivations

Since the model-based approaches can provide a location estimation of a target in a short time and fingerprinting-based approaches can provide a higher localization accuracy in general, we propose another localization algorithm: WPL-ELM in [19], which integrates the fast estimation of WPL and the high localization accuracy of ELM. During the offline phase, the indoor environment is divided into small zones firstly and an ELM model is developed for each zone. During the online phase, the WPL approach is used to determine the zone of the target primarily, then the ELM model of that zone is deployed to provide a location estimate of the target.

3.3.2 Methodology of integrated WPL-ELM

Model-based approaches can provide a location estimate of a target in a short time since they do not require any site survey during the offline phase [7]. On the other hand, fingerprinting-based approaches can provide a higher localization accuracy than the model-based approaches, with an extra offline calibration [6, 7]. Since WPL and ELM can be classified as a signal propagation model based approach and a fingerprinting-based approach respectively, it is brilliant if we can make use of both the fast estimation of WPL and the high localization accuracy of ELM together for indoor localization. Following this thought, another localization algorithm: WPL-ELM integrating the advantages of both WPL and ELM is proposed. The methodology of the integrated WPL-ELM approach is presented in Algorithm 3.1 and Figure 3.8 demonstrates its flowchart.

Algorithm 3.1: WPLELM

Offline calibration phase

1. Divide the indoor environment into multiple small zones according to the distribution of the sensors
2. Construct an ELM localization model in each zone

Online localization phase

1. The zone of the target is estimated by using the WPL approach as introduced in section 3.1 primarily
 2. After the zone is confirmed, the location of the target is determined by the ELM localization model of that zone
-

During the offline calibration phase, big indoor space is divided into multiple small zones according to the distribution of the sensors. Then, an ELM localization model for each zone is developed.

During the online localization phase, the WPL approach is used to determine the zone of the target primarily in the first step. After we know the zone the target is located at, the ELM localization model of that zone is deployed in the second step to provide a location estimate of the target.

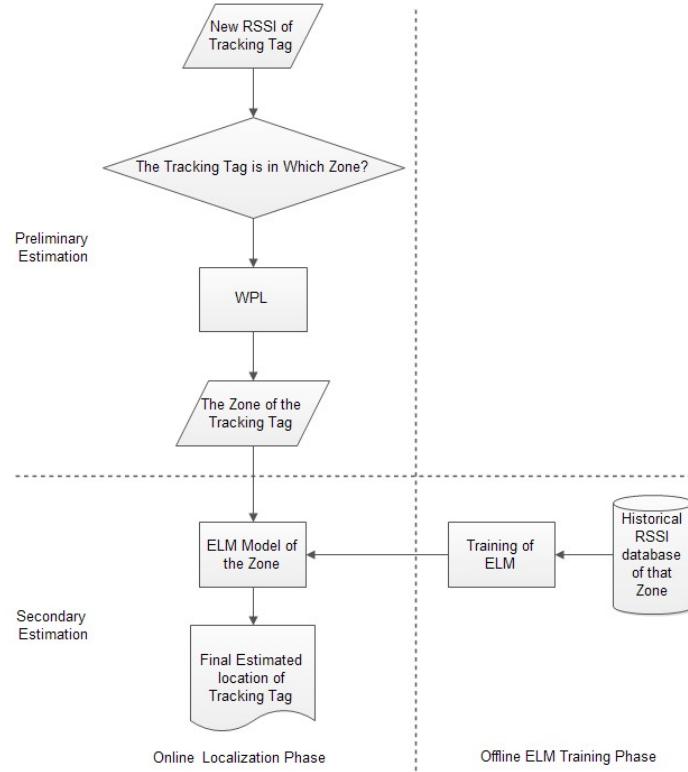


Figure 3.8: Flowchart of the integrated WPL-ELM approach

3.3.3 Experimental results and performance evaluation

We conduct Experiment III on our RFID based IPS [17] to evaluate the performance of the integrated WPL-ELM approach. During the offline phase, As shown in Figure 3.4, the entire room is divided into three small zones firstly. Zone 1 contains 7 sensors, 8 reference tags and 4 tracking tags. Zone 2 contains 8 sensors, 9 reference tags and 3 tracking tags. There are 5 sensors, 5 reference tags and 2 tracking tags in Zone 3. We keep collecting data of the signal strength of the 19 reference tags and the 9 testing tags from the 14 RFID sensors for 5 days during the offline phase. 318500 RSSI samples for each tag are obtained in this experiment. After that, for each of 19 reference tags, 5000 RSSI samples are randomly chosen as training fingerprints from Experiment III database. These RSSI samples with their corresponding physical location coordinates are put into the ELM training process and are adopted to build up the ELM model in each zone for real-time localization.

Until all the ELM models for each zone are established, we evaluate the performance

of WPL-ELM based on the RSSI samples of 9 tracking tags from Experiment III database. Since WPL is adopted as the preliminary estimation of the tracking tag, we first analyze the reliability of WPL in classifying the tracking tags into the correct zone. By evaluating all the 318500 RSSI samples for each tracking tag in Experiment II, WPL can determine the zone of the tracking tag with a 97.8% accuracy. With fast estimation and 1.782m localization accuracy of the tracking tag, WPL is fully capable of providing the correct zone of the tracking tag, or equally an estimate of its location.

After we get the preliminary location estimation of the tracking tag (the zone the tracking tag is located at), ELM is adopted to provide a location estimate of the target by using the ELM model of that zone which is developed during the offline phase. The performance comparison between ELM and WPL-ELM is shown in Figure 3.9. The distance errors of WPL-ELM as shown in Figure 3.7 are mostly within 2.3m which is the best among the four approaches.

Table 3.4 demonstrates the performance comparison between ELM and WPL-ELM in terms of the training time, average testing time and average localization accuracy of the samples in each zone followed by the overall performance. As observed in Table 3.4, the overall average localization accuracy of WPL-ELM is 0.799m, which enhances the precision of localization accuracy by 62% over BP, 55% over SVR and 33% over ELM respectively. In addition, the more noteworthy point is that WPL-ELM largely reduces both training time during the offline phase and testing time during the online phase as compared with ELM. The overall training time of WPL-ELM is 147.089s, which is 41% less time than ELM. The overall testing time of WPL-ELM is 0.432s, 4.22 times faster than ELM. Therefore, WPL-ELM can overcome the drawback of ELM, namely the tedious testing time during the online phase.

In summary, WPL-ELM is a hybrid localization algorithm that inherits the merits of the high efficiency of propagation model-based approach and the high accuracy of fingerprinting-based approaches, and is able to address the challenges and achieve

Table 3.4: Comparison between WPL-ELM and ELM

Approach	Training Time (s)	Testing Time (s)	Accuracy (m)
ELM	248.026	1.825	1.198
WPL-ELM			
Zone 1	59.338	0.524	0.763
Zone 2	62.057	0.428	0.901
Zone 3	25.694	0.252	0.719
Overall	147.089	0.432	0.799
Improvement	41%	76%	33%

performance than each of them. It performs localization in two steps. In the first step, WPL is utilized to determine the zone of target. This step acts like a filter to narrow down the searching range for the location estimation using ELM. Since we train a dedicated ELM localization model for each zone using the fingerprints within it, the corresponding model is more fine-grained and able to fully describe the relationship between the physical space and the signal space in that region. On the other hand, the original ELM trains the model using all the fingerprints in the entire area with the same number of hidden nodes (to make a fair comparison). Its model is relatively coarse and may not characterize the links between the physical coordinates and RSS vectors in some regions. With the fine-grained localization model, WPL-ELM is able to outperform original ELM in terms of localization accuracy.

3.4 Conclusion

In this chapter, we have proposed three localization algorithms: WPL, ELM and integrated WPL-ELM. WPL is a path loss model-based approach which does not require any reference tags and provides a reasonably accurate location estimate of the target effectively. ELM is a machine learning fingerprinting-based localization algorithm which can provide higher localization accuracy than other existing fingerprinting-based approaches. The integrated WPL-ELM approach combines the fast estimation of WPL and the high localization accuracy of ELM. Our experimental

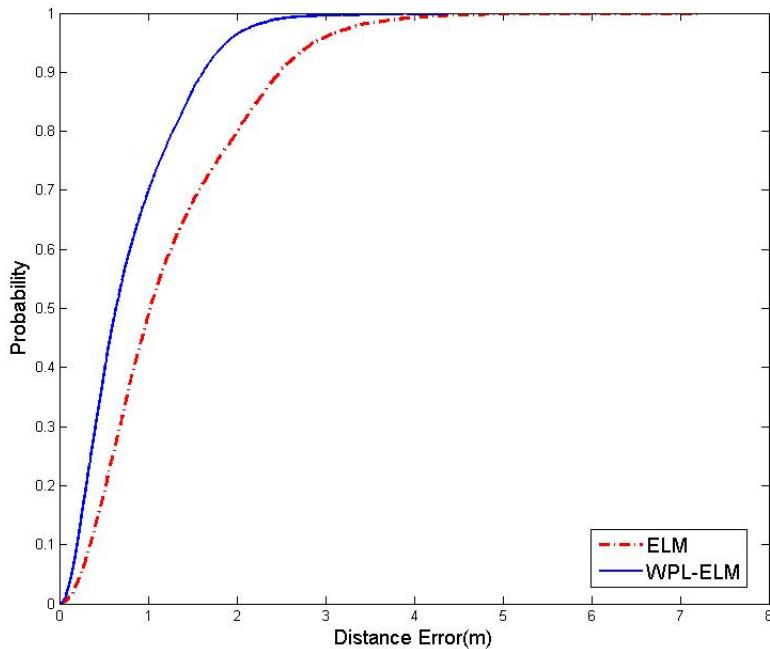


Figure 3.9: Cumulative percentile of error distance for different methods

results demonstrate that the WPL approach enhances the precision of localization accuracy by 38% over LANDMARC and 17% over enhanced LANDMARC. In addition, WPL is more robust when the number of RFID sensors is reduced than existing approaches. The ELM approach has tremendous advantages in offline training time and online localization accuracy compared with other approaches. It improves the precision of indoor localization by 43% over BP and 32% over SVR respectively.

Considering the fast estimation by WPL and the high localization accuracy by ELM, another localization algorithm, integrated WPL-ELM, which integrates the advantages of both approaches, was further proposed. Based on our experimental results, the training time and testing time of WPL-ELM are 1.69 times and 4.22 times faster than ELM. Furthermore, it improves the precision of indoor localization by 62% over the BP approach, 55% over the SVR approach and 33% over the ELM approach respectively. In conclusion, WPL-ELM can provide a higher localization accuracy of the target in a more efficient way than existing approaches.

Chapter 4

Robustness of IPS Against Device Heterogeneity

In Chapter 3, three localization algorithms have been proposed. As stated in Section 2.1.9, WiFi has been recognized as a promising wireless technique to provide LBS in large scale indoor environments due to the availability of existing WiFi network infrastructures in buildings. However, there are still several drawbacks that restrain WiFi based IPS from large-scale implementation. In this chapter, we aim to tackle the device heterogeneity issue for WiFi based IPS.

Though RSS is related to the distance of a transmitter-receiver pair, it is hard to characterize the relationship by using explicit formulas. Hence, the WiFi fingerprinting approach [65–68] is proposed by leveraging RSS as location fingerprints. As illustrated in Section 2.3.2, the fingerprinting-based approach consists of two phases: an offline training phase and an online localization phase. During the offline training phase, WiFi RSSs from various WiFi APs received by a mobile device at known locations are recorded. During the online localization phase, the RSS fingerprint measured by each mobile device is compared with the RSSs stored in the database, and its location is estimated by examining the difference between them.

It is acknowledged that the fingerprinting approach results in high localization accuracy provided that the testing device is the same as the reference device and under

the same environment, but its localization performance can be severely degenerated for heterogeneous devices [15]. Due to the proliferation of various types and brands of mobile devices, it is indispensable and urgent to develop a robust location fingerprinting technique so as to provide accurate, reliable and fast indoor positioning services for heterogeneous devices. In addition, indoor environments often change over time, and consequently, the fingerprint database built in the offline phase can deviate from the truth during the online phase, and thus, the robustness of the fingerprinting-based IPS is inevitably affected.

In the literature, certain methods have been developed for the treatment of device heterogeneity [15, 94, 103, 105–110] as well as for mitigating the influences of indoor environmental changes. For example, the crowdsourcing based IPSs (e.g. [117, 131–133]) are able to partially adapt to indoor environmental changes for the sake of fingerprints crowdsourced at different times, but such fingerprints suffer from low accuracy.

In order to address robustness issue with respect to the device heterogeneity, in this chapter, we propose to standardize WiFi fingerprints based on a statistical shape analysis method (i.e. Procrustes analysis) [134], and define Signal Tendency Index (STI) to measure the similarity between such standardized location fingerprints. A theoretical analysis indicates that STI, which is convenient to be derived in an online fashion, displays outstanding tolerance of device heterogeneity and indoor environmental changes. More importantly, STI can be straightforwardly integrated with existing WiFi localization schemes, such as propagation model based schemes [17, 18], KNN based schemes [54, 65, 66], ELM based schemes [14, 135] and so on, to improve their robustness. Furthermore, considering the fact that ELM [96, 97] provides good generalization performance at an extremely fast learning speed, we integrate the weighted version of ELM, termed WELM [136], and STI to develop an efficient and robust IPS, termed STI-WELM. To be specific, STI-WELM employs STI to standardize RSS values measured by online testing devices and collected by the reference device during the offline site survey. By leveraging our proposed weighting scheme,

which considers the relative importance of each RSS sample according to its corresponding STI value, a weight matrix for STI-WELM offline training is constructed, which establishes a STI-WELM model with high robustness. Extensive experiments have been conducted, and the results show that the proposed STI-WELM scheme provides more reliable and precise localization accuracy than other approaches.

The rest of the chapter is organized as follows. Section 4.1 introduces STI and then provides a theoretical and experimental analysis to demonstrate its capability and usefulness in handling device heterogeneity and environmental dynamics. Section 4.2 presents the proposed STI-WELM algorithm. In Section 4.3, our experimental testbed and data collection procedure are elaborated at first, and then experimental results and performance evaluation of the proposed scheme are reported. We conclude this chapter in Section 4.4.

4.1 Standardizing WiFi Fingerprints based on the Procrustes Analysis Method

In this section, we shall introduce a technique to standardize WiFi fingerprints to improve the robustness of the fingerprinting-based IPS.

During the offline site survey phase, only one mobile device (MD) is required as a reference device (RD), and the RSS fingerprints from all the APs at each reference point (RP) are collected and stored in the fingerprint database. Suppose that there are m RPs and n APs in total, and at each RP, p RSS fingerprints are collected by the RD from n APs. The mean RSS vector at the i -th RP (denoted by RP_i) is defined as $RDS_i \in \mathbb{R}^n$, in which the j -th element is the mean RSS value collected at RP_i from the j -th AP (denoted AP_j) during a period of time. In the case when the location of RP_i is out of the detectable range of AP_j , we let the corresponding mean RSS be the minimum detectable RSS value, i.e. -100 dBm.

During the online phase, the RSS values measured by a testing device (TD) from all the APs are denoted by a vector $TDS = [P_1, \dots, P_j, \dots, P_n]$, in which P_j is the

mean RSS value collected over a period of time from AP_j . Likewise, the minimum detectable RSS value of any TD is $-100dBm$.

4.1.1 Experimental analysis

In order to better understand device heterogeneity and grasp the key features of RSS values from heterogeneous devices, we conduct an experiment using five different mobile devices, including two mobile phones (iPhone 5S and Nokia E71), two tablets (iPad Air and Samsung GT-P1000 Galaxy Tab) and one laptop (Fujitsu LifeBook T4220). In the IoT testbed as introduced in Section 4.3.1, 60 RSS samples are measured within one minute for each of the five mobile devices at the same location with respect to 8 WiFi APs installed at different locations. As can be seen in Figure 4.1, each curve connects the average RSS values between one device and 8 APs. The RSS values associated with different mobile devices are significantly different, which verifies the effect of device heterogeneity. It is also conceivable that, if one device (say Nokia E71) is employed as a reference device in the offline site survey to create the WiFi fingerprint database and another device (say iPad Air) is considered to be positioned in the online phase, then the fingerprint matching result (say using Euclidean distance) will return the true location or any nearby location at an extremely low probability due to the obvious gap between any pair of the curves. Hence, the indoor localization accuracy will be remarkably degraded.

It is notable that, although the differences exist between any pair of curves from different devices, the shapes of the curves display certain similarities, as shown in Figure 4.1; in other words, one curve can be roughly recovered from another one via translation and scale operations. This observation motivates us that, instead of matching RSS fingerprints directly, a better performance may be obtainable by comparing the shapes of the curves associated with different devices. Intuitively, since shape comparison is immune to rotation, translation and scale, the negative effect of device heterogeneity can be mitigated.

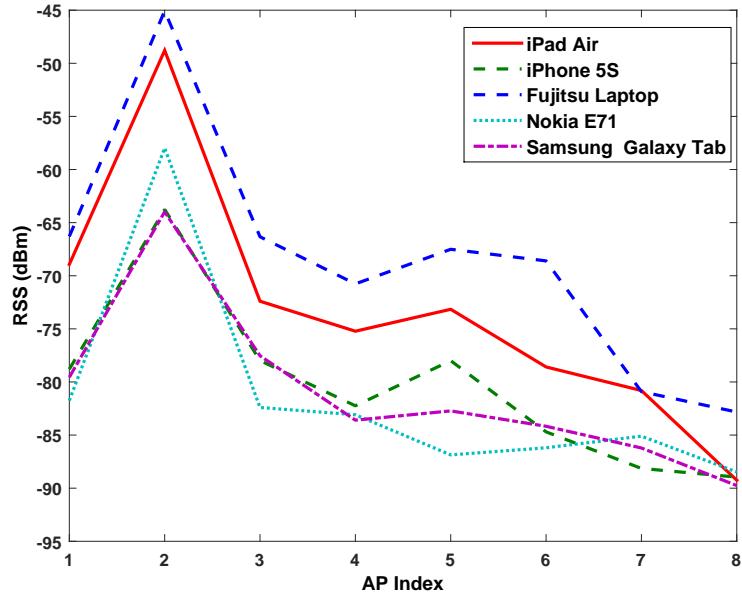


Figure 4.1: WiFi RSS values measured by different mobile devices at the same location.

4.1.2 Standardizing RSS fingerprints

Based on the analysis in the previous section, we adopt the most well-known and popular ordinary Procrustes analysis (PA) method [134] in the field of statistical shape analysis for the purpose of shape comparison. To compare the shapes of two or more objects, the PA method "superimposes" all the given objects by optimally translating, uniformly scaling and rotating them. In our case, a fingerprint (which is represented by a RSS vector, e.g. TDS) denotes an object, but due to the fact that such a fingerprint can be regarded as a one-dimensional object, only the translation and uniformly scaling operations of the ordinary PA method are involved.

Given a RSS vector of a TD, namely TDS , the translation step of the ordinary PA method will produce

$$P_1 - \overline{TDS}, P_2 - \overline{TDS}, \dots, P_n - \overline{TDS} \quad (4.1)$$

where

$$\overline{TDS} = \frac{1}{n} \sum_{j=1}^n P_j.$$

 4.1. STANDARDIZING WIFI FINGERPRINTS BASED ON THE
 PROCRUSTES ANALYSIS METHOD

Then, in the uniformly scaling step, we have

$$\widehat{TDS} = [P_1 - \overline{TDS}, P_2 - \overline{TDS}, \dots, P_n - \overline{TDS}] / \hat{\sigma}, \quad (4.2)$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{j=1}^n (P_j - \overline{TDS})^2}.$$

The vector \widehat{TDS} is thus the transformed object for superimposition, namely the standardized RSS fingerprint. Similarly, the transformed objects of all the RSS vectors collected by the reference device are derived and stored in the database. Suppose that one of the standardized RSS vector stored in the database, namely \widehat{RDS} , is chosen for matching with the RSS vector \widehat{TDS} from the TD. To evaluate the similarity between the two original curves in terms of their shapes, the Procrustes distance between the two vectors \widehat{TDS} and \widehat{RDS} , termed signal tendency index (STI), is computed as follows

$$s = \|\widehat{TDS} - \widehat{RDS}\| \quad (4.3)$$

where $\|\cdot\|$ denotes the Euclidean norm.

After elementary mathematical operations we can obtain

$$s = \sqrt{2n(1 - \rho)}, \quad (4.4)$$

where ρ denotes the sample Pearson product-moment correlation coefficient (PPM-CC) [137] between the vectors TDS and RDS . Regarding \widehat{RDS} , we define

$$\widehat{RDS} = [P_1^R - \overline{RDS}, P_2^R - \overline{RDS}, \dots, P_n^R - \overline{RDS}] / \hat{\sigma}, \quad (4.5)$$

where P_j^R is mean RSS value collected by RD from AP_j and

$$\overline{RDS} = \sum_{i=1}^n P_i^R, \quad (4.6)$$

$$\hat{\sigma}^R = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i^R - \overline{RDS})^2}.$$

Then, we can have

$$\begin{aligned}
 & \|\widehat{TDS} - \widehat{RDS}\|^2 \\
 &= \sum_{j=1}^n \left(\frac{P_j - \overline{TDS}}{\hat{\sigma}} - \frac{P_j^R - \overline{RDS}}{\hat{\sigma}^R} \right)^2 \\
 &= \sum_{j=1}^n \left(\frac{P_j - \overline{TDS}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \overline{TDS})^2}} - \frac{P_j^R - \overline{RDS}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i^R - \overline{RDS})^2}} \right)^2 \\
 &= \sum_{j=1}^n \left(\frac{(P_j - \overline{TDS})^2}{\frac{1}{n} \sum_{i=1}^n (P_i - \overline{TDS})^2} + \frac{(P_j^R - \overline{RDS})^2}{\frac{1}{n} \sum_{i=1}^n (P_i^R - \overline{RDS})^2} \right. \\
 &\quad \left. - \frac{2(P_j - \overline{TDS})(P_j^R - \overline{RDS})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \overline{TDS})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i^R - \overline{RDS})^2}} \right) \\
 &= \frac{\sum_{j=1}^n (P_j - \overline{TDS})^2}{\frac{1}{n} \sum_{i=1}^n (P_i - \overline{TDS})^2} + \frac{\sum_{j=1}^n (P_j^R - \overline{RDS})^2}{\frac{1}{n} \sum_{i=1}^n (P_i^R - \overline{RDS})^2} \\
 &\quad - \frac{2 \sum_{j=1}^n (P_j - \overline{TDS})(P_j^R - \overline{RDS})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \overline{TDS})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i^R - \overline{RDS})^2}} \\
 &= 2n - \frac{2n \sum_{j=1}^n (P_j - \overline{TDS})(P_j^R - \overline{RDS})}{\sqrt{\sum_{i=1}^n (P_i - \overline{TDS})^2} \sqrt{\sum_{i=1}^n (P_i^R - \overline{RDS})^2}} \\
 &= 2n(1 - \rho)
 \end{aligned}$$

Although (4.4) establishes an equivalence relation between the STI and sample PPMCC, it does not imply that the proposed STI method can be replaced by the simple PPMCC. To be specific, due to the ability of standardizing RSS fingerprints from heterogeneous and anonymous devices in an online fashion, the STI method can be applied to preprocess RSS fingerprints, so as to alleviate the effect of device heterogeneity on any RSS fingerprints based treatment that follows, whereas the simple PPMCC cannot be used in this way. For example, the STI method can be integrated with existing WiFi fingerprinting localization schemes, such as KNN based schemes [17, 18, 54] ELM based schemes [14, 19], and so on, to improve

their robustness to heterogeneous devices; moreover, in any practical crowdsourcing based IPS (e.g. [117, 131–133]), RSS fingerprints, which are normally collected at different times and from heterogeneous devices, can be firstly standardized by the STI method and then used for building the fingerprint database, which is helpful in mitigating the negative impact of device heterogeneity.

4.1.3 Theoretical analysis

In the first place, we investigate the reason why the STI method is robust to heterogeneous MDs from a theoretical perspective.

Without loss of generality, let the APs be transmitters and MDs be receivers. Suppose that $P(d_j)$ denotes the RSS by a MD at an arbitrary distance d_j from the transmitter of the j -th AP. According to the LDPL model [138], we have

$$P(d_j)(\text{dBm}) = 10 \log \left(\frac{\tau_j^2 G_j G_{MD} T_j}{16\pi^2} \right) - 10\alpha \log d_j + Z_j \quad (4.7)$$

where τ_j is the wavelength of the propagating signal in meter, G_j and G_{MD} are the transmitter and receiver antenna gains at the AP and MD, respectively, T_j is the signal transmission power, α is the path loss exponent, and Z_j is a random variable representing the shadowing effect in dBm which is assumed to be normally distributed with mean zero and variance σ_j^2 . It is acknowledged that (4.7) holds only if d_j is beyond a closed-in reference distance. Accordingly, the mean RSS value can be expressed as follows:

$$\bar{P}(d_j)(\text{dBm}) = 10 \log \left(\frac{\tau_j^2 G_j G_{MD} T_j}{16\pi^2} \right) - 10\alpha \log d_j \quad (4.8)$$

Since the values of parameters G_j , T_j and G_{MD} depend on the hardware of the AP and MD, if the same pair of AP and MD is considered, the relationship between the mean RSS value and the distance from the AP to the MD is one-to-one. However,

if APs or MDs with different hardware are adopted, the corresponding relationship becomes many-to-one; that is to say, given one mean RSS value, there are multiple possible distances. Hence, with the WiFi fingerprinting technique, there exist certain discrepancies between a location and its fingerprint (i.e. the corresponding RSS values) if heterogeneous APs or MDs are used, which will degrade the IPS performance. This explains why device heterogeneity degrades the performance of the fingerprinting-based IPS.

Signal strength difference (SSD) is a location signature which leverages the differences of signals perceived at APs from a MD [15]. With the SSD method, if the first AP is used as reference AP, then the SSD associated with the j -th AP is produced as new fingerprints as follows

$$\begin{aligned} & P(d_j)(\text{dBm}) - P(d_1)(\text{dBm}) \\ &= 10 \log \frac{\tau_j^2 G_j T_j}{\tau_1^2 G_1 T_1} - 10\alpha \log \frac{d_j}{d_1} + Z_j - Z_1 \end{aligned} \quad (4.9)$$

with $j = 2, \dots, n$. As suggested in [15], since the parameter G_{MD} depending on the MD hardware does not exist in (4.9), the SSD is entirely free from the influence of the device heterogeneity caused by using different MDs. It is noticeable that the SSD variance is $\sigma_j^2 + \sigma_1^2$.

Using the STI method, the average RSS from all the APs, which is denoted by \overline{TDS} can be formulated as follows:

$$\begin{aligned} \overline{TDS}(\text{dBm}) &= \frac{10}{n} \sum_{p=1}^n \log(\tau_p^2 G_p T_p) + 10 \log \frac{G_{MD}}{16\pi^2} \\ &\quad - \frac{10\alpha}{n} \sum_{p=1}^n \log d_p + \frac{1}{n} \sum_{p=1}^n Z_p \end{aligned} \quad (4.10)$$

and the translating RSS associated with the j -th AP is

$$\begin{aligned} & P(d_j)(\text{dBm}) - \overline{TDS}(\text{dBm}) \\ &= 10 \log(\tau_j^2 G_j T_j) - \frac{10}{n} \sum_{p=1}^n \log(\tau_p^2 G_p T_p) - 10\alpha \log d_j \end{aligned}$$

$$+ \frac{10\alpha}{n} \sum_{p=1}^n \log d_p + Z_j - \frac{1}{n} \sum_{p=1}^n Z_p. \quad (4.11)$$

It follows from (4.11) that the translating RSS in the STI method is uncorrelated with G_{MD} (i.e., the parameter depending on the MD hardware). In addition, according to (4.3) the scaling parameter s is uncorrelated with G_{MD} as well. Hence, it can be concluded that location fingerprints standardized by the STI method are immune to the device heterogeneity induced by MDs like the SSD method.

Furthermore, the variance of the translating RSS is equal to $\sigma_j^2 - 2\sigma_j/n + \sum_p \sigma_p^2/n^2$, which is generally small in comparison with the SSD method; e.g., if $\sigma_1 = \dots = \sigma_n$ and $n \gg 1$, the variance in the STI method is much smaller than that in the SSD method. Since a small variance indicates a narrow range of fingerprints (e.g., translating RSS and SSD) associated with each physical location, it is accordingly easy to discriminate these locations through fingerprints. Therefore, it reveals that STI is superior to SSD, which is further verified by the experimental study in Section 4.3.

Next, we analyze how the STI method improves the robustness to indoor environmental changes under certain conditions.

As pointed out in [138], in some environments, such as buildings, stadiums and other indoor environments, the path loss exponent (PLE) can take values in the range of 4 to 6. In a given indoor environment, if the number and distribution of objects (e.g. people, furniture, and so on) change over time, one constant PLE cannot accurately characterize the path attenuation at all times; that is to say, the fingerprint collected at one location during the site survey procedure will most likely deviate from its counterparts in the online phase due to indoor environmental changes, which inevitably impairs the robustness of the fingerprinting-based IPS.

On these grounds, define $\alpha + \Delta\alpha$ to be the real PLE in the online phase, where $\Delta\alpha$ reflects indoor environmental changes. Provided that the n APs are homogeneous

or have similar hardware parameters (i.e. G_1, \dots, G_n and T_1, \dots, T_n), (4.11) can be simplified as

$$\begin{aligned} P(d_j)(\text{dBm}) - \overline{TDS}(\text{dBm}) &= Z_j - \frac{1}{n} \sum_{p=1}^n Z_p \\ &+ (\alpha + \Delta\alpha) \left(-10 \log d_j + \frac{10}{n} \sum_{p=1}^n \log d_p \right). \end{aligned} \quad (4.12)$$

Equation (4.12) indicates that, given $G_1 = \dots = G_n$ and $T_1 = \dots = T_n$, the translating RSS $P(d_j)(\text{dBm}) - \overline{TDS}(\text{dBm})$ is just scaled by $\alpha + \Delta\alpha$ if ignoring the noise terms, and as a result, the shape of the fingerprint at this location is scaled in the same way as well. Considering the fact that the PA method is able to compare the shapes of objects under different scales, the scaling issue caused by the indoor environmental changes is thus mitigated when using the standardized fingerprint. However, when using the original fingerprinting technique and SSD, the scaling issue cannot be addressed, and the IPS performance will be degraded.

To sum up, the theoretical analysis reveals that the translating and scaling operations adopted in the STI method are able to alleviate the effect of device heterogeneity and indoor environmental changes. Moreover, it is also convenient to deduce similar conclusions as above if we let the MD be the transmitter and APs be receivers.

4.2 Proposed STI-WELM Algorithm

In this section, we first introduce preliminaries on WELM, and then describe the structure of the proposed algorithm combining STI and WELM.

4.2.1 Preliminaries on WELM for indoor localization

Data in real applications such as RSS from different APs usually have imbalanced class distribution, which means that some data are more important than others in

Table 4.1: Training inputs and training target for WiFi RSS fingerprint database

Training input \mathbf{x} , $RSS(dBm)$								Training target \mathbf{t}_i , (m)
AP_1	AP_2	AP_3	AP_4	AP_5	AP_6	AP_7	AP_8	Locations
-32	-95	-63	-53	-79	-47	-69	-49	(2.55 10.68)
-90	-84	-73	-65	-58	-43	-59	-37	(12.16 23.73)
\vdots				\ddots				
RSS_i^1	RSS_i^2	RSS_i^3	RSS_i^4	RSS_i^5	RSS_i^6	RSS_i^7	RSS_i^8	(t_i^1, t_i^2)
\vdots				\ddots				
RSS_M^1	RSS_M^2	RSS_M^3	RSS_M^4	RSS_M^5	RSS_M^6	RSS_M^7	RSS_M^8	(t_M^1, t_M^2)

the database. WELM proposed in [136] is to tackle the regression or classification tasks with imbalanced class distribution. It inherits the advantages from the original ELM [96] and is able to deal with data of imbalanced distributions by incorporating the information of imbalance dataset. For the scenario of IPS, assume that there are M WiFi RSS fingerprints in total collected at RPs. These WiFi RSS fingerprints and their physical coordinates are adopted as training inputs and training targets respectively to build up the WELM model. As demonstrated in Table. 4.1, each training sample can be represented as $(\mathbf{x}_i, \mathbf{t}_i) \in \mathbb{R}^n \times \mathbb{R}^2$, where the training input $\mathbf{x}_i = [RSS_i^1, RSS_i^2, \dots, RSS_i^n]$ is a vector of RSS received from n APs in the environment, and training target $\mathbf{t}_i = (t_i^1, t_i^2)$ is the 2-D physical coordinates of the RP. Assume that a SLFN with L hidden nodes can approximate these M samples with zero error, then there exist β_u , \mathbf{a}_u and b_u such that

$$\mathbf{t}_i = \sum_{u=1}^L \beta_u G(\mathbf{a}_u, b_u, \mathbf{x}_i), i = 1, 2, \dots, M, \quad (4.13)$$

where \mathbf{a}_u and b_u are the learning parameters of the hidden nodes, β_u is the output weight, and $G(\mathbf{a}_u, b_u, \mathbf{x}_i)$ is the activation function which gives the output of the u th hidden node with respect to the input \mathbf{x}_i . Given M arbitrary distinct training samples $(\mathbf{x}_i, \mathbf{t}_i)$, $i = 1, 2, \dots, M$, by substituting \mathbf{x} with \mathbf{x}_i in (4.13) we obtain $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$ where $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1)^T, \mathbf{h}(\mathbf{x}_2)^T, \dots, \mathbf{h}(\mathbf{x}_M)^T]_{M \times L}^T$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_L]_{L \times 2}^T$ and $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_M]_{M \times 2}^T$. \mathbf{H} is the hidden layer output matrix, $\boldsymbol{\beta}$ is the output weight matrix

and \mathbf{T} is the training target matrix of WELM. The hidden layer output matrix \mathbf{H} remains unchanged once the input weights and hidden layer biases are randomly given [136]. After that, an $M \times M$ diagonal matrix \mathbf{W} is defined which is associated with every training sample \mathbf{x}_i .

Regarding the IPS, we apply WELM to solve the localization problem by regression, namely minimizing the weighted cumulative localization error with respect to each training sample $(\mathbf{x}_i, \mathbf{t}_i)$, which can be mathematically written as:

$$\begin{aligned} \min_{\boldsymbol{\xi}, \boldsymbol{\beta} \in \mathbb{R}^{L \times 2}} \quad & L_P = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + W \frac{C}{2} \sum_{i=1}^M \xi_i \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} = t_i^T - \xi_i^T \quad i = 1, 2, \dots, M \end{aligned} \quad (4.14)$$

where ξ_i is the training error of \mathbf{x}_i , which is caused by the difference of the output $\mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta}$ and desired output t_i , and C is a hyper-parameter for better generalization performance [139]. The solution of the output weight vector $\boldsymbol{\beta}$ is analytically determined using the Moore-Penrose generalized inverse $\hat{\mathbf{H}}$. Dependent on the size of training samples, there are two versions of solutions of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \begin{cases} \mathbf{H}^T (\frac{\mathbf{I}}{C} + \mathbf{W} \mathbf{H} \mathbf{H}^T)^{-1} \mathbf{W} \mathbf{T}, & M < L \\ (\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{T}, & M > L \end{cases} \quad (4.15)$$

It can be seen from the above two formulas that a positive definite matrix \mathbf{I}/C is added to the diagonal of $\mathbf{W} \mathbf{H} \mathbf{H}^T$ or $\mathbf{H}^T \mathbf{W} \mathbf{H}$. Since the weight matrix $\mathbf{W} = \text{diag}(W_{ii}), i = 1, \dots, M$ is significant in WELM, two weighting schemes are proposed in [136]. One weighting scheme assigns a unified W_{ii} to each sample, the other adopts the value of golden standard that represents the perfection in nature. However, both of these weighting schemes are static and have not considered the importance of each training sample. We shall propose a new weight scheme in Section 4.2.2, which assigns different weight to each sample according to its significance.

In summary, the training process of WELM is conducted in the following three main steps:

Step 1: Randomly assign the hidden neuron parameters: input weights \mathbf{a}_u and hidden layer biases b_u , $u = 1, \dots, L$.

Step 2: Calculate the hidden layer output matrix \mathbf{H} and the weight matrix \mathbf{W} .

Step 3: Calculate the output weight β .

4.2.2 STI-WELM

The proposed STI-WELM algorithm inherits the merits of both STI and WELM, and consists of two main phases: online construction phase and online localization phase.

4.2.2.1 Online construction phase

The skeleton of the online construction procedure is depicted in Algorithm 3.2.

Algorithm 4.1: STI-WELM

1. Construct TDS_{new} based on STI measurements
 2. Build up the WiFi RSS fingerprint database for STI-WELM training
 3. Construct the weight matrix \mathbf{W} of STI-WELM
 4. Establish the trained STI-WELM model for online localization
-

First of all, we calculate the STI value s_i between TDS and each RDS_i . Since a smaller s_i indicates that RDS_i is more similar with TDS , we further define a weight value w_i for each RDS_i , which is calculated as follows:

$$w_i = \frac{\frac{1}{s_i}}{\sum_{i=1}^m \frac{1}{s_i}} \quad (4.16)$$

Then, the m RPs are sorted according to their w_i in a descending order. For the next step, we reconstruct the TDS according to the RSS values RDS_i collected by the RD at the Q RPs, with their corresponding w_i weight values being larger than a threshold w_{th} (setting $w_{th} = 0.01$ in general), to alleviate the negative effect of device heterogeneity between TD and RD. For those RPs whose corresponding w_i

weight values are smaller than w_{th} , their RSS values are discarded since they are not critical for localization anymore.

The TDS_{new} is calculated based on the following formula?

$$TDS_{new} = \frac{1}{\sum_{q=1}^Q \frac{1}{s_q}} \sum_{q=1}^Q RDS_q \cdot \frac{1}{s_q} \quad (4.17)$$

It can be seen from the above equation that TDS_{new} is constructed from $RDS_q, q = 1, \dots, Q$ at the Q RPs, which means that it is more relevant to the RSS vectors in the database than the raw TDS .

The next step is to build up a WiFi RSS fingerprint database for the STI-WELM training process. Unlike conventional fingerprinting-based IPS which requires to put the RSS fingerprints collected at all m RPs into a database for training, STI-WELM only requires to leverage the RSS fingerprints from the Q RPs whose corresponding w weight values are larger than w_{th} to build up the database. It largely reduces the computational burden for the training process. Suppose that f RSS fingerprints are collected at each RP. Therefore, the training set of STI-WELM becomes a $fQ \times n$ matrix, where the order of the f RSS fingerprints collected at each RP is based on its w_i weight values from the largest to the smallest.

After that, construct the weight matrix \mathbf{W} for the STI-WELM training process. Note that the two weighting schemes proposed in [136] are generated regardless of the special property of training samples. On the contrary, in our scheme, the weight vector for each RSS fingerprint is designed elaborately based on its corresponding STI value s_q , namely

$$\mathbf{W} = \frac{1}{\sum_{q=1}^Q \frac{1}{s_q}} \text{diag}\left(\frac{1}{s_1}, \dots, \frac{1}{s_Q}\right) \otimes I_f, \quad (4.18)$$

where I_f is the identity matrix of order f and \otimes denotes the Kronecker product. Then, these fQ RSS fingerprints and their corresponding physical locations are

adopted as the training inputs \mathbf{x} and the training targets \mathbf{t} respectively for STI-WELM offline training. Similarly to WELM, the STI-WELM model will be trained as mentioned in Section 4.2.1. The detailed steps are illustrated below:

Step 1: Randomly assign the hidden neuron parameters: input weights \mathbf{a}_u and biases b_u , $u = 1, \dots, L$.

Step 2: Calculate the hidden layer output matrix \mathbf{H} and the weight matrix \mathbf{W} .

Step 3: Calculate the output weight β .

Additionally, the activation function G and the number of hidden nodes L will be selected carefully in order to guarantee the performance of STI-WELM. A guideline for these parameter selections is presented in Section 4.3. The STI-WELM model can be obtained quickly due to the fast training speed of WELM.

4.2.2.2 Online localization phase

When a user sends a location query with the real-time TDS measurement, by feeding the TDS to the trained STI-WELM model, the output of the model is the estimated location of the user.

4.3 Experimental Study

4.3.1 Testbed

In order to validate the performance of our developed WiFi based IPS and proposed localization algorithms, we deployed our system in the Internet of Things (IoT) Laboratory of the School of Electrical and Electronic Engineering, Nanyang Technological University, named as IoT testbed for all the experiments mentioned throughout Chapters 4–7. The layout of the testbed is depicted in Figure 4.2. It is a 600 m^2 ($35\text{ m} \times 17\text{ m}$) multi-functional lab, which includes two workspaces for 50 undergraduates, two cubical offices for 40 graduates, a personal office for lab manager, an open space for Unmanned Aerial Vehicle (UAV) testing and a conference room.

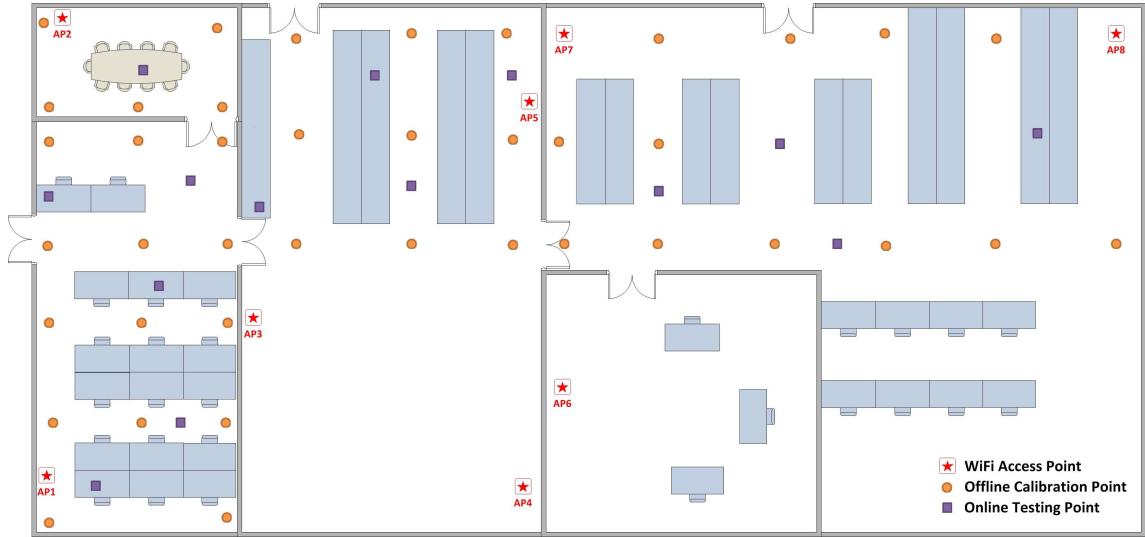


Figure 4.2: Layout of the testbed

4.3.2 System setup and data collection procedure

Extensive experiments were conducted to evaluate the performance of the proposed STI-WELM localization scheme in the IoT testbed as introduced in Section 4.3.1. In this experiment, as illustrated in Figure 4.2, 8 Linksys WRT54GS WiFi routers were installed as APs for our experiments. All the APs were fixed on 1.9-meter-high tripods to keep them on the same height level.

To examine the influence of device heterogeneity, we employed 5 different mobile devices in our experiments, including iPhone 5S (Phone), iPad Air (Tablet), Nokia E71 (Phone), Samsung Galaxy Tab (Tablet) and Fujitsu LifeBook T4220 (Laptop). Table 4.2 summarizes the detailed information of these devices. A script program was developed and ran on all the APs in order to collect RSS fingerprints associated with each mobile device from multiple APs. By leveraging this program, the APs were able to scan the 802.11 packets transmitted between mobile devices and APs so as to retrieve the RSS information of each packet, and then send the RSS value and the MAC address of the corresponding mobile device to a master AP which is connected to a central server. The server will store the RSS fingerprints and further apply our proposed STI-WELM algorithm to estimate the location of each device.

Table 4.2: Mobile devices used for data collection

Device	WiFi Module	OS
iPhone 5S	Broadcom BCM4334	iOS 8
iPad Air	Broadcom BCM43241	iOS 8
Nokia E71	Wi-Fi 802.11 b/g	Symbian 9.2
Samsung Tablet	TI OMAP4430	Android 4.2
Fujitsu Laptop	Intel Wireless WiFi Link 4965AGN	Windows 7

Specifically, we collected RSS fingerprints of the five mobile devices at 54 different points, including 40 offline calibration points and 14 online testing points, as shown in Figure 4.2. For each mobile device, 500 RSS fingerprints were collected at each point. The grid spacing between two adjacent locations of the calibration points was chosen to be larger than 1.25 m based on the analysis in [140]. At each point, the mobile device was put on a 1.65-meter-high plastic cart for collecting WiFi RSS fingerprints.

4.3.3 Comparison between RSS, SSD and STI as location fingerprints

First of all, we evaluate the performance of the location fingerprints coming from the RSS, SSD and STI. Take the RSS for example. We first include all the RSS fingerprints of five mobile devices at each point into a fingerprint set of size 2500, calculate the sample standard deviation associated with each AP, and evaluate the average standard deviation from the 8 APs to measure the stability of the location fingerprints at this point. Likewise, we can calculate the average standard deviations for the location fingerprints used in SSD and STI. Note that, to make a fair comparison, the location fingerprints in the STI method only involve the translating operation.

Figure 4.3 demonstrates the distribution histograms of the average standard deviations at the given 54 points with respect to the RSS, SSD and STI. As can be seen, the average standard deviation of STI is basically within 6 dBm while those of the

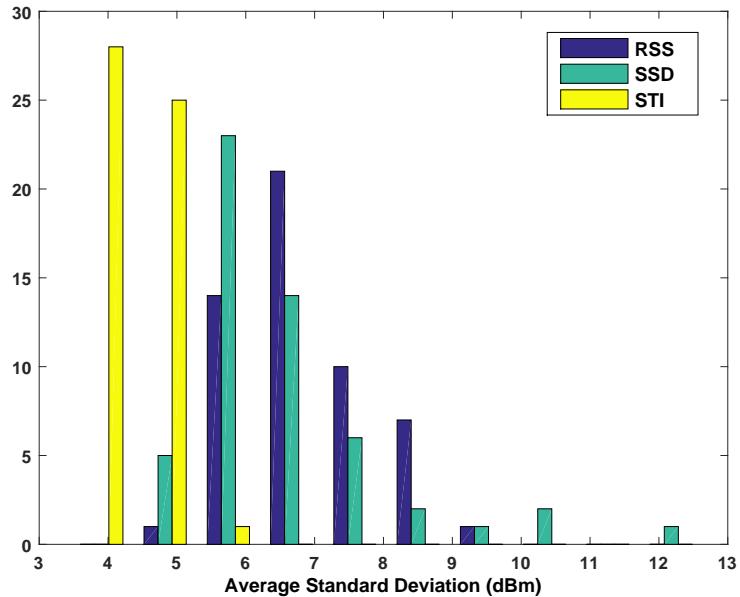


Figure 4.3: Distribution histogram of the average standard deviations of different location fingerprints

RSS and SSD are large and much widely scattered, which means that STI results is more stable and reliable location fingerprints than the others.

4.3.4 Experimental results and evaluation

Two well-known localization algorithms, namely KNN [65] and ELM [96], are chosen in order to further compare their performance when RSS, SSD and STI are applied as the location fingerprint. We include KNN into the comparison because of its wide usage as one of the classical localization algorithms. Furthermore, it has been shown in [110] that KNN-based approaches are superior to Bayesian Inference (BI) based approaches [107] when the reference device and the testing devices are different. Therefore, we include KNN instead of BI into our performance comparison and evaluation. For KNN, the value of K is determined empirically case-by-case according to the related WiFi RSS fingerprints database of the reference device. During the online phase, by matching the measured WiFi RSS fingerprints with the K closest WiFi RSS fingerprints in the database, the location of the target will be calculated. The algorithm is the same as in [65].

It has been shown in [19] that the performance of ELM in terms of the offline training time, the online testing time and the average localization accuracy are better than classical machine learning algorithms such as back-propagation (BP) algorithm and support vector machine for regression (SVR) algorithm. Therefore we also choose ELM as the localization algorithm when RSS, SSD and STI are applied as the localization fingerprint respectively. This methodology is introduced in [19].

In practice, it is more likely for the users to carry different devices from the reference device. Therefore, we only analyze the situations that the testing device and reference device are distinct in our experiments.

By leveraging the 500×40 WiFi RSS fingerprints at the 40 offline calibration points of each device, the offline RSS, SSD and STI location fingerprint databases are established. The 500×14 WiFi RSS fingerprints at the online testing points of each device are utilized for the performance evaluation of each localization algorithm. The distance error is used to measure the localization accuracy of each approach. We define the location estimation error e to be the distance between the real location coordinates (x_0, y_0) and the system estimated location coordinates (x, y) , i.e.:

$$e = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (4.19)$$

Since we utilize 5 mobile devices in our experiments, there are 20 different combinations of reference device and testing device.

4.3.4.1 Comparison between STI-KNN without scaling and STI-KNN

In order to evaluate the influence of the uniform scaling step of STI on the localization accuracy, we compare the performance of STI-KNN with and without scaling step in the first place. The value of K is chosen to be 13 by five-fold cross-validation for this experiment.

The specific average localization errors given different combinations of reference devices and testing devices are demonstrated in Table 4.3. As can be seen, STI-KNN

Table 4.3: Detailed average localization errors (in meter) of STI-KNN without scaling and STI-KNN

Testing Dataset	STI-KNN without scaling	STI-KNN
Training Dataset: iPhone 5S		
iPad Air	3.107	2.959
Nokia E71	3.717	3.716
Samsung Tablet	3.836	3.777
Fujitsu Laptop	2.983	2.927
Average	3.411	3.345
<i>iPhone 5S</i>	2.424	2.398
Training Dataset: iPad Air		
iPhone 5S	3.445	3.339
Nokia E71	3.653	3.639
Samsung Tablet	3.706	3.612
Fujitsu Laptop	2.877	2.875
Average	3.420	3.366
<i>iPad Air</i>	2.849	2.813
Training Dataset: Nokia E71		
iPhone 5S	3.630	3.552
iPad Air	3.093	3.029
Samsung Tablet	3.850	3.746
Fujitsu Laptop	3.132	3.126
Average	3.426	3.363
<i>Nokia E71</i>	2.798	2.736
Training Dataset: Samsung Tablet		
iPhone 5S	3.555	3.483
iPad Air	3.050	2.989
Nokia E71	3.728	3.673
Fujitsu Laptop	3.079	3.019
Average	3.353	3.291
<i>Samsung Tablet</i>	2.761	2.743
Training Dataset: Fujitsu Laptop		
iPhone 5S	3.758	3.385
iPad Air	3.268	2.950
Nokia E71	3.844	3.611
Samsung Tablet	4.056	3.653
Average	3.731	3.400
<i>Fujitsu Laptop</i>	2.832	2.734

can provide higher localization accuracy in every situation when the uniform scaling step is involved in the procedure. In general, the uniform scaling step enhances the precision of indoor positioning of STI-KNN by 3.09%. Therefore, we can conclude that the uniform scaling step could facilitate STI to mitigate the effect of indoor environmental dynamics.

4.3.4.2 Comparison among RSS-KNN, SSD-KNN and STI-KNN

Two location fingerprints: RSS and SSD are leveraged and integrated with the KNN localization algorithm to compare with STI-KNN. Since the value of K is critical for the performance of KNN approaches when the reference device is altered, we analyze the performance of RSS-KNN, SSD-KNN and STI-KNN with all the possible values of K , and compare their best performance in each scenario (20 different combinations of reference device and testing device). Table 4.4 demonstrates the specific average localization errors of each combination of reference device and testing device of these three approaches with their best performances given the optimal K value.

It is evident from Table 4.4 that STI-KNN provides higher localization accuracy than RSS-KNN and SSD-KNN in every situation. Figure 4.4 depicts the distance error distribution of the three approaches when each mobile device is leveraged as the reference device. Similar to the results shown in Table 4.4, STI-KNN has the best performance among the three approaches in terms of localization accuracy.

To summarize, as shown in Table 4.5, STI-KNN can enhance the precision of indoor positioning by 26.71% over RSS-KNN and 22.46% over SSD-KNN respectively. Thus, when KNN is employed as the localization algorithm, the proposed STI method can largely alleviate the effect of device heterogeneity and provide robust and high indoor positioning service consistently even the testing devices are different from the reference device.

Table 4.4: Detailed average localization errors (in meter) under various situations (KNN)

Testing Dataset	RSS-KNN	SSD-KNN	STI-KNN
Training Dataset: iPhone 5S			
<i>Best K for KNN</i>	13	12	13
iPad Air	4.738	3.778	2.959
Nokia E71	4.263	4.326	3.716
Samsung Tablet	4.447	4.723	3.777
Fujitsu Laptop	4.892	3.344	2.927
Average	4.585	4.043	3.345
Training Dataset: iPad Air			
<i>Best K for KNN</i>	13	15	13
iPhone 5S	4.154	4.388	3.339
Nokia E71	4.698	4.404	3.639
Samsung Tablet	4.651	4.549	3.612
Fujitsu Laptop	3.576	3.235	2.875
Average	4.270	4.144	3.366
Training Dataset: Nokia E71			
<i>Best K for KNN</i>	14	13	13
iPhone 5S	3.914	4.576	3.552
iPad Air	4.434	4.287	3.029
Samsung Tablet	4.001	4.909	3.746
Fujitsu Laptop	4.586	3.810	3.126
Average	4.234	4.395	3.363
Training Dataset: Samsung Tablet			
<i>Best K for KNN</i>	15	12	13
iPhone 5S	4.074	4.450	3.483
iPad Air	4.957	4.197	2.989
Nokia E71	4.152	4.545	3.673
Fujitsu Laptop	5.154	3.705	3.019
Average	4.584	4.224	3.291
Training Dataset: Fujitsu Laptop			
<i>Best K for KNN</i>	13	11	13
iPhone 5S	5.209	4.856	3.385
iPad Air	3.972	4.277	2.950
Nokia E71	5.986	4.946	3.611
Samsung Tablet	5.644	5.187	3.653
Average	5.203	4.816	3.400

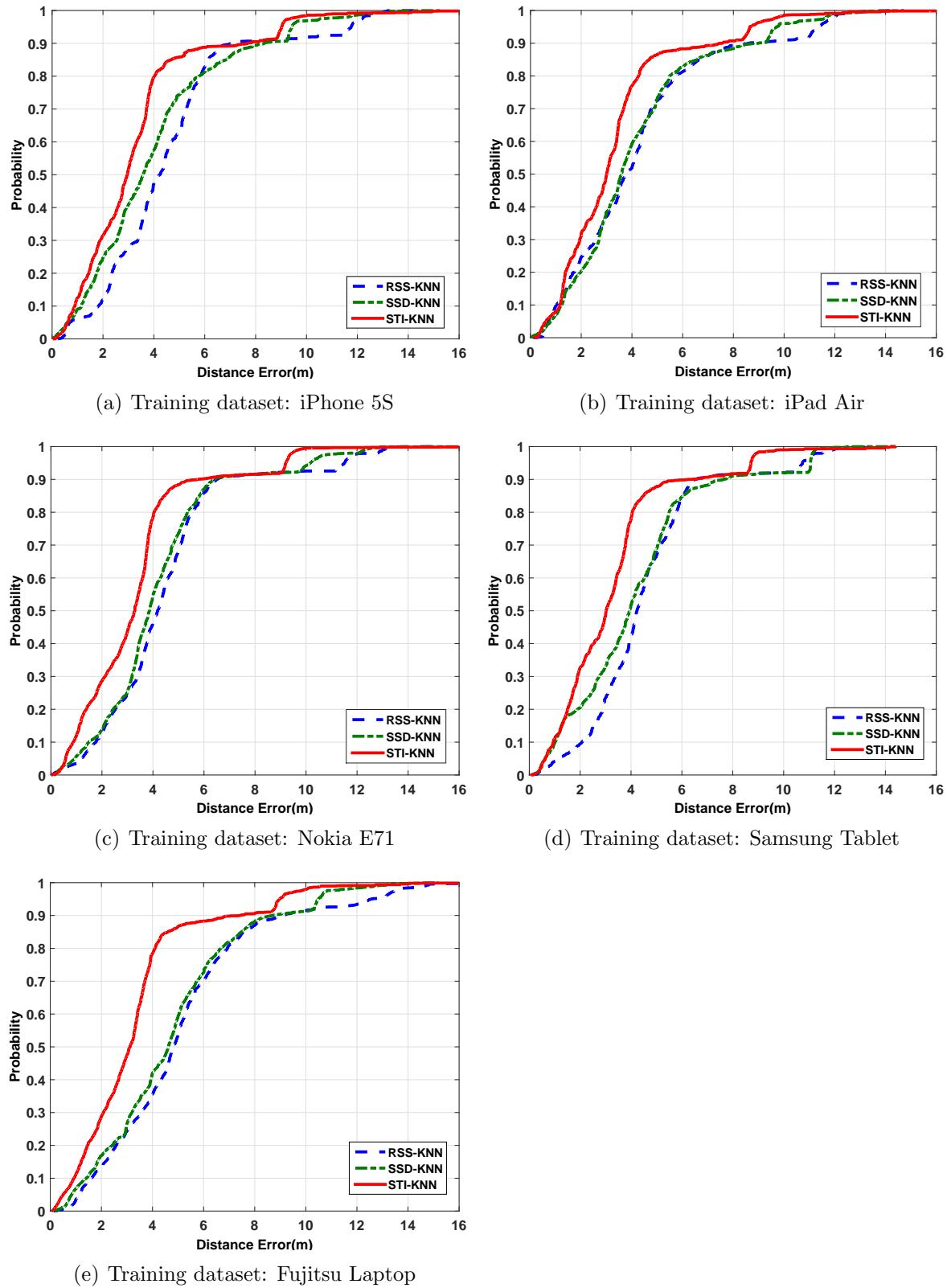


Figure 4.4: Comparison of distance error distributions for different methods (KNN)

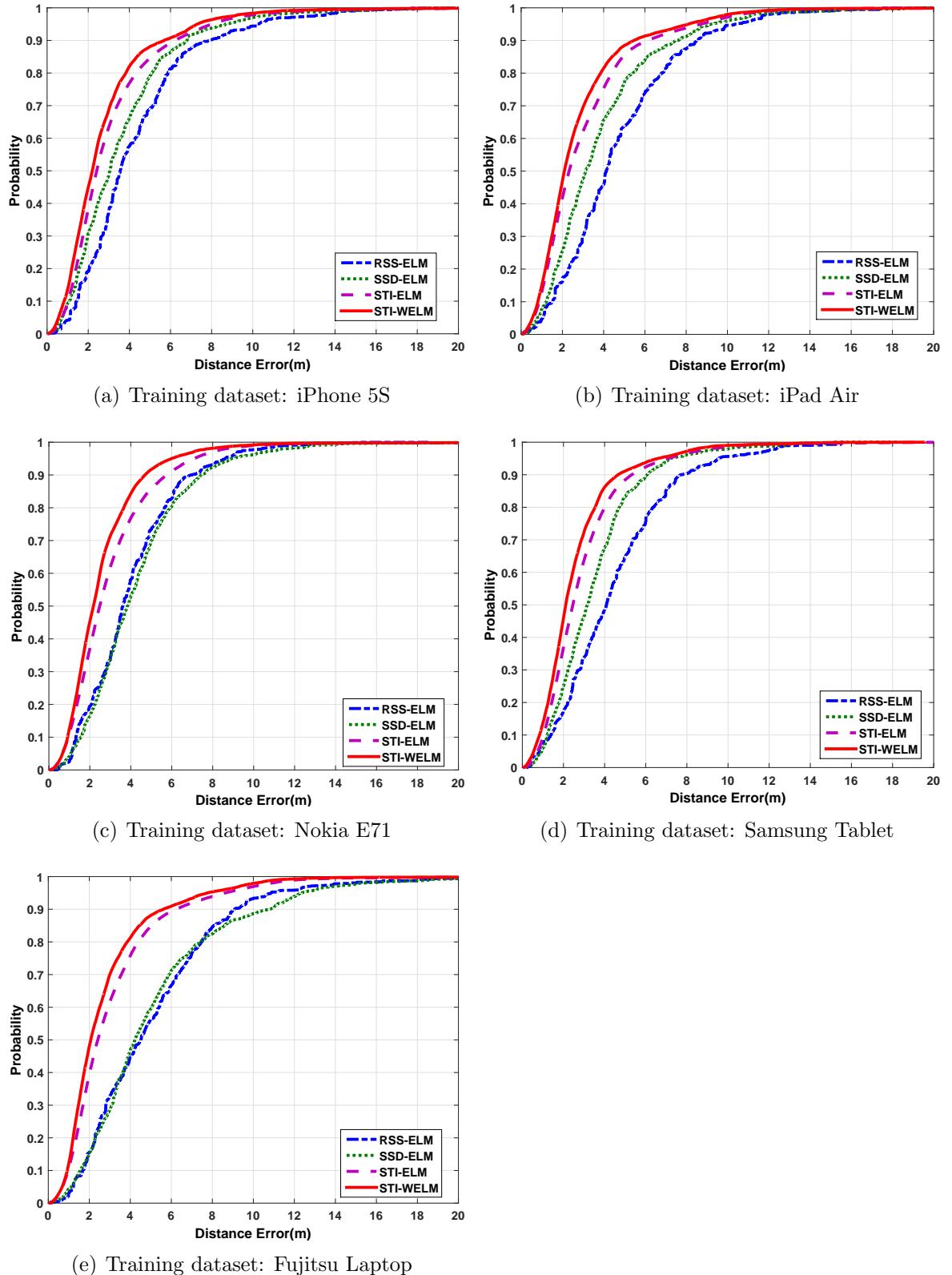


Figure 4.5: Comparison of distance error distributions for different methods (ELM and WELM)

Table 4.5: Summary of average localization errors (in meter) (KNN)

Training Dataset	RSS-KNN	SSD-KNN	STI-KNN
iPhone 5S	4.585	4.043	3.345
iPad Air	4.270	4.144	3.366
Nokia E71	4.234	4.395	3.363
Samsung Tablet	4.584	4.224	3.291
Fujitsu Laptop	5.203	4.816	3.400
Average	4.575	4.324	3.353

4.3.4.3 Comparison between RSS-ELM, SSD-ELM, STI-ELM and STI-WELM

As a fingerprinting-based IPS, the relevant ELM models for online localization are required to be built up during the online construction phase. For RSS-ELM, the RSS-ELM model is built up by adopting the 500×40 WiFi RSS fingerprints collected at the 40 offline calibration points. These WiFi RSS fingerprints and their physical locations are adopted as training inputs and training targets respectively to build up the model. For the construction of SSD-ELM model, the 500×40 WiFi RSS fingerprints collected at the offline calibration points are transferred into SSD format first. Then the model is built up in a similar process as RSS-ELM. During the online testing phase, the raw RSS vectors measured by TD are reconfigured into the SSD format and put into the trained SSD-ELM model, then the estimated location of the TD will be calculated.

For STI-ELM and STI-WELM, as mentioned in Section 4.2.2, only WiFi RSS fingerprints collected by the RD at the Q RPs are adopted for building up the STI-ELM model and STI-WELM model. Therefore, the training process of these two approaches are much faster than RSS-ELM and SSD-ELM, which require to train the WiFi RSS fingerprints at all the 40 offline calibration points. In our experiments, we select Q to be 13 because the w weight values of these Q RPs are larger than $w_{th} = 0.01$. Therefore, the size of training database is largely reduced from 500×40 to 500×13 . The performance of three activation functions: radial basis function (RBF) $G(a, b, x) = e^{-b\|x-a\|^2}$, sine function $G(a, b, x) = \sin(ax + b)$ and hard-limit

Table 4.6: Detailed average localization errors (in meter) under various situations (ELM)

Testing Dataset	RSS-ELM	SSD-ELM	STI-ELM	STI-WELM
Training Dataset: iPhone 5S				
iPad Air	4.368	2.927	2.494	2.084
Nokia E71	4.240	4.529	3.825	3.543
Samsung Tablet	5.221	4.372	3.689	3.474
Fujitsu Laptop	3.374	2.616	2.476	2.070
Average	4.301	3.611	3.121	2.793
<i>iPhone 5S</i>	4.260	3.367	2.507	2.237
Training Dataset: iPad Air				
iPhone 5S	4.378	3.909	3.039	2.732
Nokia E71	5.567	4.808	3.689	3.344
Samsung Tablet	5.146	3.874	3.331	3.136
Fujitsu Laptop	3.669	2.718	2.295	1.953
Average	4.690	3.827	3.089	2.791
<i>iPad Air</i>	3.684	3.552	2.176	1.842
Training Dataset: Nokia E71				
iPhone 5S	4.181	4.328	3.410	3.090
iPad Air	4.148	3.849	2.497	1.988
Samsung Tablet	4.336	5.180	3.569	3.293
Fujitsu Laptop	3.414	3.793	2.542	2.025
Average	4.020	4.288	3.005	2.599
<i>Nokia E71</i>	3.965	4.113	2.761	2.334
Training Dataset: Samsung Tablet				
iPhone 5S	4.140	3.775	3.207	2.933
iPad Air	4.718	3.097	2.392	2.001
Nokia E71	5.390	4.158	3.651	3.267
Fujitsu Laptop	3.603	2.960	2.534	2.160
Average	4.463	3.498	2.946	2.590
<i>Samsung Tablet</i>	4.126	2.238	2.592	2.439
Training Dataset: Fujitsu Laptop				
iPhone 5S	6.770	5.473	3.128	2.687
iPad Air	3.365	4.090	2.459	2.087
Nokia E71	5.811	5.931	3.726	3.362
Samsung Tablet	4.504	5.309	3.375	3.076
Average	5.113	5.201	3.172	2.803
<i>Fujitsu Laptop</i>	3.816	4.449	2.242	2.081

Table 4.7: Summary of average localization errors (in Meter) (ELM)

Training Dataset	RSS-ELM	SSD-ELM	STI-ELM	STI-WELM
iPhone 5S	4.301	3.611	3.121	2.793
iPad Air	4.690	3.827	3.088	2.791
Nokia E71	4.020	4.288	3.005	2.599
Samsung Tablet	4.463	3.498	2.946	2.590
Fujitsu Laptop	5.113	5.201	3.172	2.803
Average	4.517	4.085	3.066	2.715

transfer (hardlim) function $G(a, b, x) = \text{hardlim}(ax + b)$ are analyzed by leveraging the offline WiFi RSS fingerprints database. The hardlim function is chosen since it provides the best performance among the three activation functions. Another critical parameter for the performance of the ELM based approaches is the number of hidden nodes L . The five-fold cross-validation method is employed with a range from 1 to 100 and a step size of 2 in order to determine the optimal L . After comprehensive evaluations on both localization accuracy and repeatability, L is selected to be 23 for both STI-ELM and STI-WELM. Based on our experimental results, these two approaches only spend 0.056s on average to calculate the output weights β for the $(500 \times 13 = 6500)$ WiFi RSS fingerprints during the training process.

A guideline for selecting the type of activation function and the number of hidden nodes in the STI-WELM hidden layer, both of which are the critical parameters for the performance of the STI-WELM approach, is listed as follows: the suggested default activation function for the STI-WELM approach is the hardlim function, whose performance is better than others in general; as for the optimal number of hidden nodes, the five-fold cross-validation method is employed with a range from 0 to 100 and a step size of 2 based on the empirical tuning.

The weight matrix \mathbf{W} of STI-WELM is calculated according to our proposed weighting scheme as introduced in Section 4.2.2. Similar to the KNN experiments, there are 20 different combinations of reference device and testing device because 5 mobile devices are employed in the ELM experiments.

The specific average localization errors of each combination of reference device and

testing device when RSS-ELM, SSD-ELM, STI-ELM and STI-WELM are adopted are demonstrated in Table 4.6 respectively. It can be seen from Table 4.6 that its localization performance trumps other three approaches significantly in every combination. It is also noteworthy that the performance of STI-ELM is better than RSS-ELM and SSD-ELM. The mean localization accuracies of RSS-ELM and SSD-ELM are almost the same. The distance error distributions of the four approaches when each mobile device is leveraged as the reference device are presented in Figure 4.5. As observed in Figure 4.5, STI-WELM provides the most accurate indoor positioning service among the four approaches, which is consistent with the results demonstrated in Table 4.6.

In summary, STI-WELM enhances the precision of indoor positioning by 39.89% over RSS-ELM, 33.53% over SSD-ELM and 11.46% over STI-ELM respectively. Table 4.7 summarizes the performance of each approach. Therefore, the proposed STI-WELM can provide more robust, fast and accurate indoor positioning service than other approaches consistently, and alleviate the effect of heterogeneous issue among different devices remarkably. Furthermore, another noteworthy point is that the performance of both ELM and WELM is better than KNN provided that the same type of the location fingerprint is adopted. This claim is supported by Figure 4.4 and Figure 4.5, in which the curves produced by ELM and WELM based algorithms are smoother than KNN based ones, i.e., the ELM and WELM based approaches are more robust to outliers.

In addition, iPad Air and Fujitsu Laptop obtain the best overall localization accuracy among all the devices considered, which is due to their relative high transmission powers; see the resulting RSS values in relation to different devices in Figure 4.1.

4.3.4.4 Performance of STI-WELM vs the number of APs

The aforementioned section has demonstrated the superiority of STI-WELM to alleviate the effect of heterogeneous devices for indoor localization when all the APs

Table 4.8: Average localization errors (m) under the influence of the No. of APs

Number of APs	RSS-ELM	STI-ELM	STI-WELM
3	5.615	4.711	4.636
4	5.307	4.337	4.157
5	5.254	4.289	3.562
6	5.149	4.107	3.092
7	4.883	3.518	2.756
8 (All)	4.517	3.066	2.715

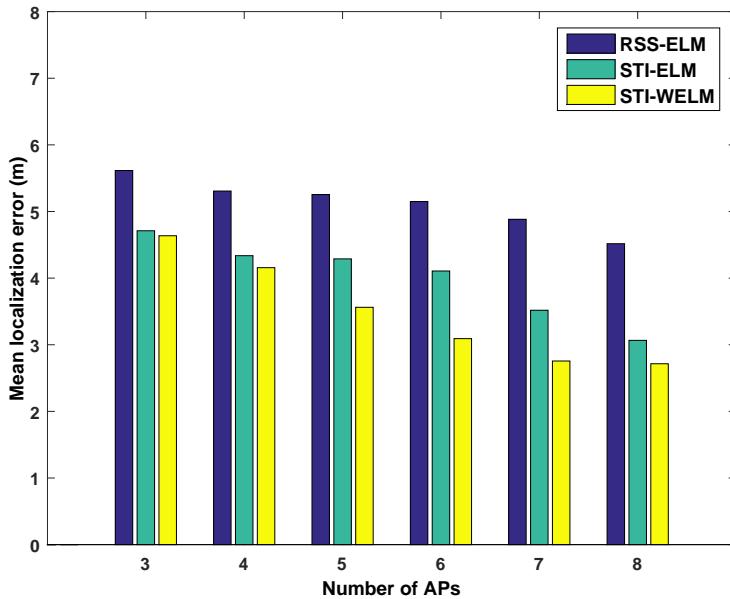


Figure 4.6: Comparison of mean localization error between different approaches under the influence of the number of APs.

in the testbed were leveraged. In this subsection, we further analyze the performance of STI-WELM with respect to the number of APs.

We compare the performance of STI-WELM with RSS-ELM and STI-ELM when the number of APs is altered. The training processes of these approaches are the same as in Section 4.2. We consider all the 20 different combinations of reference devices and testing devices for this experiment as well, since 5 different mobile devices are utilized in total. The overall performance comparison in terms of mean localization error for these approaches under different numbers of APs is demonstrated in Figure 4.6 and Table 4.8.

As shown in Table 4.8, the mean localization errors of all the three approaches

decrease as the number of APs in-use increases. It can be easily observed that STI-WELM outperforms RSS-ELM and STI-ELM in every situation. The performance of RSS-ELM is the worst in all situations since it leverages the raw RSS data of the testing device and the large localization error is caused by device heterogeneity. On the contrary, after the operations of translation and uniformly scaling in STI, the newly constructed TDS_{new} is more relevant to the RSS fingerprints stored in the reference device database. This is the main reason why both STI-ELM and STI-WELM are superior to RSS-ELM. By considering the relative importance of each RSS sample according to its corresponding STI value in the reference device database and leveraging our proposed weighting scheme, the localization accuracy of STI-WELM is higher than that of STI-ELM in general. To be specific, the mean localization error of STI-WELM is at the same level with STI-ELM when only three APs are leveraged. However, as shown in Figure 4.6, it reduces significantly when the required number of APs is between 4 and 6, and keeps the mean localization error at a low level when the number of APs is more than 7.

In summary, as long as no less than three APs are available in indoor environments, the STI-based localization approaches outperform RSS-based localization approaches. Furthermore, STI-WELM, which integrates the advantages of both STI and WELM, can overcome the heterogeneity issue of mobile devices for indoor localization and provide high localization accuracy consistently even only a few APs are available in indoor environments.

4.3.5 Implementation of the proposed approach for real location-based service

Since both the theoretical and experimental analysis have verified the superiority of the proposed IPS in terms of accuracy and robustness, we have implemented our IPS in the following four different indoor environments: Internet of Things Lab ($600m^2$) in Nanyang Technological University (NTU), Lecture Theater 22 ($500m^2$) in NTU, the Center for Berkeley Education Alliance for Research in Singapore (BEARS)

headquarter ($1500m^2$), and the Center for Research in Energy Systems Transformation (CREST) Lab ($400m^2$) in University of California, Berkeley. It turns out that our system is able to provide satisfactory LBS across heterogeneous devices in these places, including indoor positioning, indoor navigation, real-time occupancy distribution monitoring and indoor geo-fencing, and has been fully operational for more than one year. For instance, [141] provides a video demo about our indoor navigation service on Google Glass, which is the mobile device (distinct from the reference device: iPad) to be localized. As shown in the video, by leveraging the proposed STI-WELM localization algorithm, our IPS offers high localization accuracy and seamless indoor navigation across heterogeneous devices.

4.4 Conclusion

In this chapter, we presented a robust and precise IPS by introducing STI, which is a new type of fingerprints and embodies more reliable and robust location signatures compared to traditional location fingerprints in the presence of heterogeneous devices and changing indoor environments. We also proposed a novel weighting scheme by taking into consideration of the relative importance of each RSS sample according to its corresponding STI value, for the WELM training process. On these grounds, we proposed the STI-WELM scheme which inherits the advantages of both STI and WELM. According to our experimental results, the STI-WELM scheme enhances the precision of indoor positioning by 39.89% over RSS-ELM, 33.53% over SSD-ELM and 11.46% over STI-ELM, respectively, which confirms the superiority of the STI approach to the traditional RSS fingerprints as well as the recently developed SSD approach. On the other hand, one potential limitation of STI is that when the number of available APs is limited (less than two), it may not be able to precisely determine the importance of each samples because its similarity metric depends on the shape of RSS vectors. However, this situation rarely happens since majority of indoor environments are covered by multiple APs now.

Chapter 5

Robustness of IPS Against Environmental Dynamics

In Chapter 4, a new type of fingerprints which embodies more reliable and robust location signatures compared to traditional location fingerprints in the presence of heterogeneous devices has been proposed. In this chapter, we further explore how to enable the WiFi based IPS to be more robust to various environmental dynamics.

Since the WiFi RSS fingerprint database is usually built up during the offline phase, it is unable to reflect the real-time radio map of the WiFi signals well once the environment is altered during the online localization phase. Environmental factors, such as presence of humans, opening and closing of doors and variations of humidity, can interfere with the propagation of WiFi signals severely [142]. This will lead to serious localization errors if the same WiFi RSS fingerprint database is adopted.

In this chapter, we propose an indoor localization algorithm based on an online sequential extreme learning machine (OS-ELM) to address the above two problems accordingly. Originating from the batch ELM, OS-ELM inherits the advantage of ELM, which can provide good generalization performance at an extremely fast learning speed [96]. In addition, OS-ELM has an online sequential learning ability that does not require retraining when new data are received [143]. Another noteworthy

point of OS-ELM is that, different from other online sequential learning algorithms, such as stochastic gradient descent back-propagation (SGBP) [144] and growing and pruning RBF network (GAP-RBF) [145], which require specific types of hidden nodes and can only handle data one-by-one (learning only one training sample at each time), OS-ELM is able to adapt to various types of hidden nodes and can learn data one-by-one or chunk-by-chunk with a varying chunk size. Therefore, WiFi RSS fingerprints can be collected and updated more flexibly for OS-ELM online sequential learning. In addition, the fast learning speed of OS-ELM greatly reduces time consumption and manpower costs for the offline site survey. More importantly, the online sequential learning ability of OS-ELM permits the entire system to provide reliable indoor LBS under various environmental dynamics. Comprehensive simulations and experiments have been conducted to validate the performance of OS-ELM. Furthermore, experiments under specific environmental changes such as variations of occupancy distribution and events of opening or closing of doors are also conducted.

The rest of the chapter is organized as follows. Section 5.1 provides the proposed localization algorithm. The simulation results and evaluation of OS-ELM are presented in Section 5.2. In Section 5.3, a system overview of our WiFi based IPS is provided firstly, followed by the experimental results and performance evaluation of the proposed localization algorithm. We conclude the work in Section 5.4.

5.1 An OS-ELM based Indoor Localization Algorithm

5.1.1 Preliminary on OS-ELM

The OS-ELM algorithm contains two phases: an initialization phase and a sequential learning phase. One special property of OS-ELM is that it can learn data one-by-one or chunk-by-chunk (a block of data) with a fixed or varying chunk size [143]. Suppose

that the network has L hidden nodes and the data $\mathfrak{N} = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m, i = 1, \dots, N\}$ are presented to the network sequentially. In the initialization phase, $\text{rank}(\mathbf{H}_0) = L$ is required to ensure that OS-ELM can achieve the same learning performance as ELM, where \mathbf{H}_0 denotes the hidden output matrix for the initialization phase. The number of training data required in the initialization phase, N_0 , has to be equal to or greater than L , *i.e.*, $N_0 \geq L$. If $N_0 = N$, the performance of OS-ELM is the same as batch ELM. Therefore, batch ELM is a special case of OS-ELM when all of the data are used in the initialization phase.

Initialization phase: a small chunk of training data \mathfrak{N}_0 is used to initialize the learning, where $\mathfrak{N}_0 = \{\mathbf{x}_i, \mathbf{t}_i\}_{i=1}^{N_0} \subseteq \mathfrak{N}$ and $N_0 \geq L$.

Step 1: Randomly assign the hidden neuron parameters: *e.g.* input weights \mathbf{a}_i and bias b_i , $i = 1, \dots, L$.

Step 2: Calculate the initial hidden layer output matrix $\mathbf{H}_0 =$

$$\begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_1) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_1) \\ \vdots & \dots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_{N_0}) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_{N_0}) \end{bmatrix}_{N_0 \times L} \quad (5.1)$$

Step 3: Estimate the initial output weight $\beta^{(0)}$. Since $\mathbf{T}_0 = [\mathbf{t}_1, \dots, \mathbf{t}_{N_0}]_{N_0 \times m}^T$, the problem is equivalent to minimizing $\|\mathbf{H}_0\beta - \mathbf{T}_0\|$. Noticing that $\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ [96], the optimal solution is given by $\beta^{(0)} = \mathbf{P}_0 \mathbf{H}_0^T \mathbf{T}_0$, where $\mathbf{P}_0 = (\mathbf{H}_0^T \mathbf{H}_0)^{-1}$ and $\mathbf{K}_0 = \mathbf{P}_0^{-1} = \mathbf{H}_0^T \mathbf{H}_0$.

Step 4: Set $k = 0$, where k is a parameter indicating the number of chunks of data that are presented to the network.

Sequential learning phase: present the $(k + 1)$ -th chunk of new observations

$$\mathfrak{N}_{k+1} = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=(\sum_{j=0}^k N_j)+1}^{\sum_{j=0}^{k+1} N_j}$$

where N_{k+1} denotes the number of observations in the $(k + 1)$ -th chunk.

Step 1: Compute the partial hidden layer output matrix $\mathbf{H}_{k+1} =$

$$\begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_{(\sum_{j=0}^k N_j)+1}) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_{(\sum_{j=0}^k N_j)+1}) \\ \vdots & \dots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_{\sum_{j=0}^{k+1} N_j}) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_{\sum_{j=0}^{k+1} N_j}) \end{bmatrix}_{N_{k+1} \times L} \quad (5.2)$$

Step 2: Calculate the output weight $\beta^{(k+1)}$. We have $\mathbf{T}_{k+1} = [\mathbf{t}_{(\sum_{j=0}^k N_j)+1}, \dots, \mathbf{t}_{\sum_{j=0}^{k+1} N_j}]^T_{N_{k+1} \times m}$.

Moreover,

$$\mathbf{K}_{k+1} = \mathbf{K}_k + \mathbf{H}_{k+1}^T \mathbf{H}_{k+1} \quad (5.3)$$

$$\beta^{(k+1)} = \beta^{(k)} + \mathbf{K}_{k+1}^{-1} \mathbf{H}_{k+1}^T (\mathbf{T}_{k+1} - \mathbf{H}_{k+1} \beta^{(k)}) \quad (5.4)$$

In order to avoid inverting matrices, such as \mathbf{K}_{k+1}^{-1} in Equation (5) in the recursive process, the Woodbury formula [146] is applied to transform the equations as follows:

$$\mathbf{K}_{k+1}^{-1} = \mathbf{K}_k^{-1} - \mathbf{K}_k^{-1} \mathbf{H}_{k+1}^T (\mathbf{I} + \mathbf{H}_{k+1} \mathbf{K}_k^{-1} \mathbf{H}_{k+1}^T)^{-1} \mathbf{H}_{k+1} \mathbf{K}_k^{-1} \quad (5.5)$$

Since $\mathbf{P}_{k+1} = \mathbf{K}_{k+1}^{-1}$,

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \mathbf{P}_k \mathbf{H}_{k+1}^T (\mathbf{I} + \mathbf{H}_{k+1} \mathbf{P}_k \mathbf{H}_{k+1}^T)^{-1} \mathbf{H}_{k+1} \mathbf{P}_k \quad (5.6)$$

$$\beta^{(k+1)} = \beta^{(k)} + \mathbf{P}_{k+1} \mathbf{H}_{k+1}^T (\mathbf{T}_{k+1} - \mathbf{H}_{k+1} \beta^{(k)}) \quad (5.7)$$

Step 3: Set $k = k + 1$. Go to Step 2 in this online sequential learning phase.

5.1.2 OS-ELM based indoor localization algorithm

The proposed OS-ELM approach considers the localization problem as a regression problem. For the WiFi RSS fingerprint calibration process, OS-ELM only requires a relative sparse radio map of WiFi RSS during the offline phase, which trumps

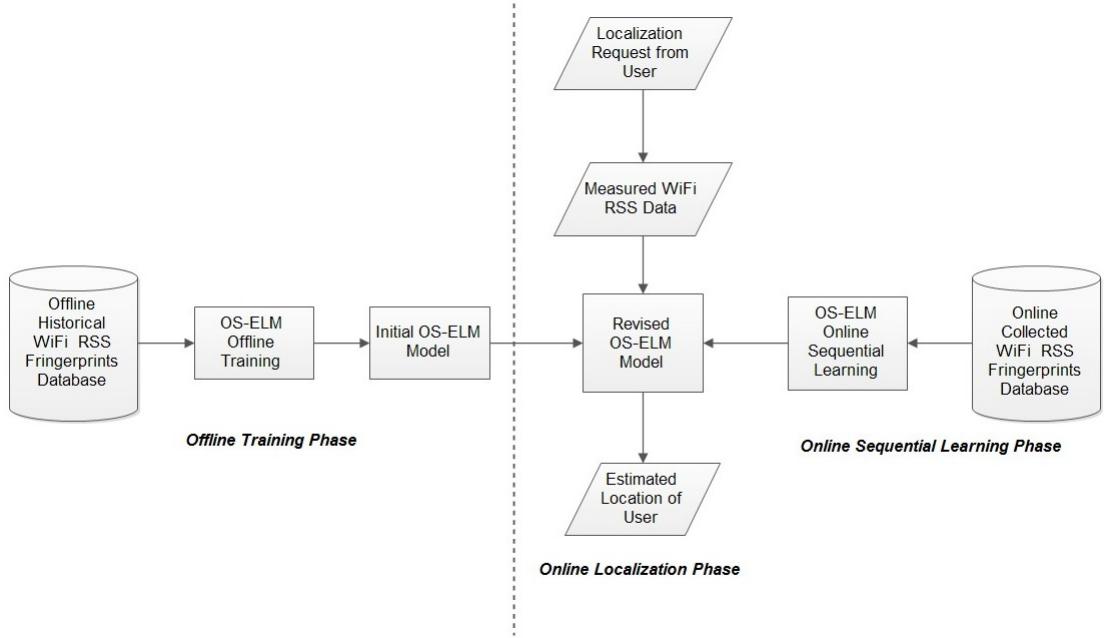


Figure 5.1: Methodology of the online sequential extreme learning machine (OS-ELM) approach.

the traditional fingerprinting-based localization algorithms. These WiFi RSS fingerprints and their physical locations are adopted as training inputs and training targets, respectively, to build up an initial OS-ELM model for online localization.

Furthermore, by leveraging the online learning ability of OS-ELM, WiFi RSS fingerprints can also be collected at some known locations during the online phase to reflect the environmental dynamics. Once new WiFi RSS fingerprints have been collected, they will be integrated into the initial OS-ELM model to update and generate a revised OS-ELM model. During the online phase, when a user sends a location request with his or her current WiFi RSS fingerprint, the fingerprint will be fed into the latest revised OS-ELM model, and then the estimated location will be calculated.

The methodology and framework of the proposed OS-ELM approach are demonstrated in Figure 5.1. The processes of the three main phases of the OS-ELM approach are presented as follows:

5.1.2.1 Offline calibration phase

Suppose that N_0 WiFi RSS fingerprints are collected at some known locations during the offline calibration phase. These WiFi RSS fingerprints and their corresponding physical locations are adopted as the training inputs \mathbf{x} and the training targets \mathbf{t} , respectively, for OS-ELM offline training. Similar to the initialization phase of OS-ELM, the initial OS-ELM model will be trained as mentioned in Section 5.1. The detailed steps are illustrated below:

Step 1: Randomly assign the hidden neuron parameters: input weights \mathbf{a}_i and input bias b_i .

Step 2: Calculate the initial hidden layer output matrix \mathbf{H}_0 .

Step 3: Estimate the initial output weight $\beta^{(0)}$.

Step 4: Set $k = 0$, where k indicates the number of updating times of WiFi RSS fingerprints that are collected during the online calibration phase.

Additionally, the activation function G and the number of hidden nodes L need to be carefully tuned in order to guarantee that the initial OS-ELM model can provide sufficient localization accuracy for the users before any WiFi RSS fingerprints are collected at any online calibration points during the online phase. The five-fold cross-validation method with a range from zero to 1000 and a step size of 50 is employed to select the optimal number of hidden nodes L . A guideline for selecting the type of activation functions and the number of hidden nodes was presented in Section 5.3.2.

5.1.2.2 Online sequential learning phase

The main purpose of the online sequential learning phase of OS-ELM is to make the localization algorithm more robust and adapt to various environmental changes.

Unlike other online sequential learning algorithms, which can only handle data one-by-one, OS-ELM can learn data one-by-one or chunk-by-chunk with a varying chunk size.

Therefore, WiFi RSS fingerprints can be collected and updated more flexibly for OS-ELM online sequential learning. These newly collected WiFi RSS fingerprints and their corresponding physical locations will be adopted as training samples, and they will be updated chunk-by-chunk to revise the initial OS-ELM model during the online sequential learning phase. Suppose that N_{k+1} WiFi RSS fingerprints have been collected during the $(k + 1)$ -th online calibration; the revised OS-ELM model will be obtained by the following steps:

Step 1: Calculate the partial hidden layer output matrix H_{k+1} .

Step 2: Calculate the output weight $\beta^{(k+1)}$.

Step 3: Set $k = k + 1$ for the next online calibration.

5.1.2.3 Online localization phase

Before any WiFi RSS fingerprints are collected at any known locations during the online phase, the initial OS-ELM model will be used to provide estimated locations of users when they send location queries with their real-time WiFi RSS fingerprints.

When more WiFi RSS fingerprints are collected at some known locations during the online calibration phase, the OS-ELM model is updated and revised. It will be used to provide the estimated location of the user.

5.2 Simulation Results and Evaluation

We develop a simulation environment using MATLAB R2013a in order to evaluate the performance of the proposed OS-ELM approach before any experiment is conducted. The simulation is running on a PC, which has an Intel Core i5-2400

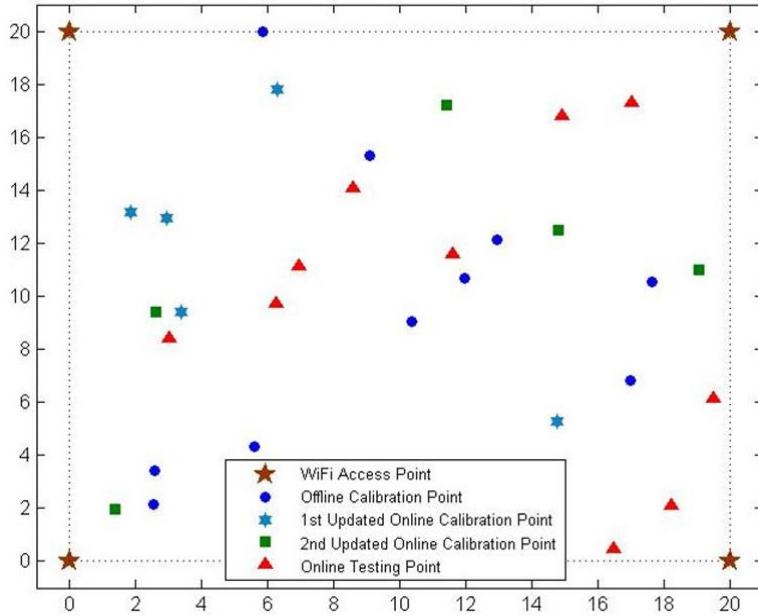


Figure 5.2: Positions of the WiFi APs, offline calibration points, online calibration points and online testing points in the simulated field.

3.10-GHz CPU and 8 GB RAM. As shown in Figure 5.2, we assume a $20\text{ m} \times 20\text{ m}$ room, where four WiFi APs are installed at the four corners of the room. The most commonly-used path loss model for indoor environment is the International Telecommunication Union (ITU) indoor propagation model [126]. Since it provides a relation between the total path loss PL (dBm) and distance d (m), it is adopted to simulate the WiFi signals generated from each WiFi AP. The indoor path loss model can be expressed as:

$$PL(d) = PL_0 - 10\alpha \log(d) + X_\sigma \quad (5.8)$$

where PL_0 is the pass loss coefficient, and it is set to be -40 dBm in our simulation. X_σ represents a zero-mean normal random noise with standard deviation $\sigma = 0.5$, and α is the path loss exponent.

During the offline calibration phase, α is set to be two in Scenario I, and simulated WiFi RSS fingerprints from the four WiFi routers are collected at 10 randomly selected offline calibration points for OS-ELM offline training, with 200 WiFi RSS

fingerprints collected at each point. The hardlim function $G(a, b, x) = \text{hardlim}(ax + b)$ is chosen as the activation function, and 380 hidden nodes are selected and put in the hidden layer during the offline training. The initial OS-ELM model is obtained after a 0.219-s training process. In order to imitate the environmental dynamics in the room, we set α to be 2.5 in Scenario II and 3.5 in Scenario III, respectively. WiFi RSS fingerprints are collected at five online calibration points and five online testing points under each scenario. Two hundred WiFi RSS fingerprints are collected at each point, and the positions of these points are distinct. The positions of the offline calibration points, the online calibration points and the online testing points are demonstrated in Figure 5.2. The updated WiFi RSS fingerprints at the online calibration points are adopted as online training samples and leveraged to revise the initial OS-ELM model.

We evaluate the performance of OS-ELM with each online sequential learning updated based on the WiFi RSS fingerprints collected at the 10 online testing points. The distance error is used to measure the localization accuracy of the proposed OS-ELM approach. We define the location estimation error e to be the distance between the real location coordinates (x_0, y_0) and the system estimated location coordinates (x, y) , *i.e.*,

$$e = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (5.9)$$

Table 5.1 illustrates the performance of OS-ELM with each online sequential learning update in terms of training time, testing time and average localization accuracy. The comparison of the cumulative percentile of error distances between the initial OS-ELM and the two updated OS-ELM is presented in Figure 5.3. It can be easily spotted from Table 5.1 that the average localization accuracy of OS-ELM becomes better when more WiFi RSS fingerprints at various online calibration points have been learned online. The latest updated OS-ELM can provide 1.794 m, which enhances the precision of indoor localization by 42.18% over the performance of the initial OS-ELM. Furthermore, the online sequential learning of OS-ELM is very efficient. It only spends 0.014 s on average to calculate the output weights β for 1000

Table 5.1: Simulation results of OS-ELM.

Number of Calibration Points (Offline + Online)	Training Time (s)	Testing Time (s)	Accuracy (m)
10 + 0	0.219	0.015	3.103
10 + 5	0.148	0.014	2.563
15 + 5	0.139	0.014	1.794

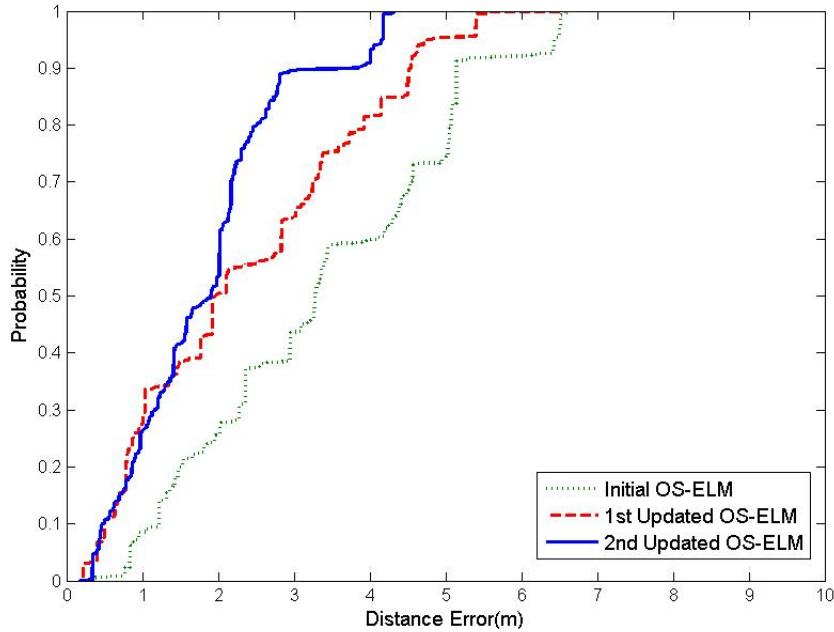


Figure 5.3: Cumulative percentile of error distance.

newly-received WiFi RSS fingerprints in each online sequential learning update. Based on the simulation results and evaluation, we can conclude that OS-ELM can provide higher localization accuracy due to its efficient online sequential learning ability when the indoor environment is altered during the online phase.

5.3 Experimental Results and Performance Evaluation

5.3.1 System overview

We conducted extensive experiments to evaluate the performance of the proposed OS-ELM approach in the IoT testbed as introduced in Section 4.3.1. As shown

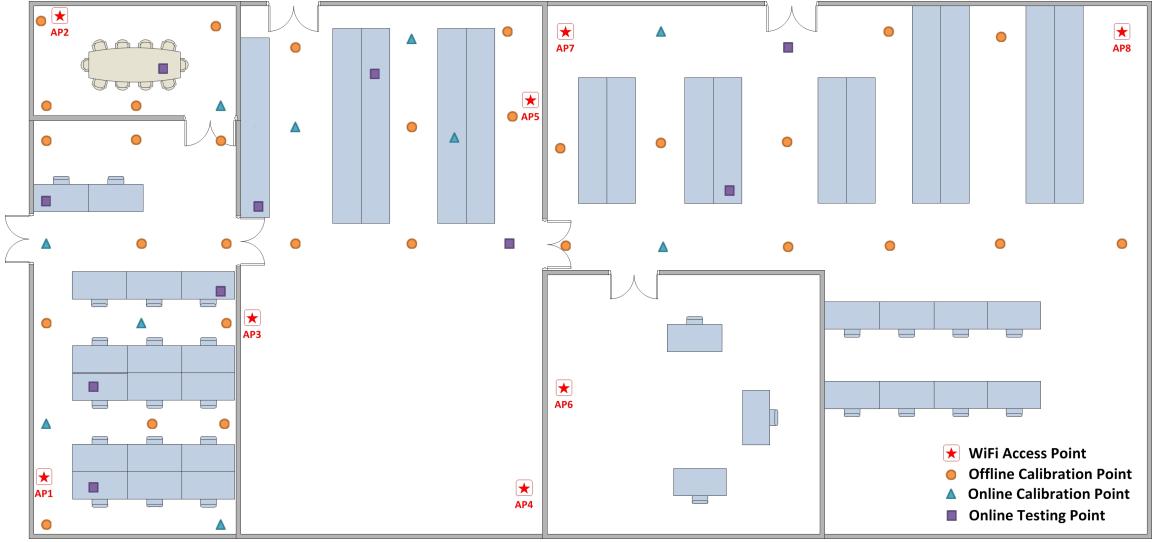


Figure 5.4: Positions of the WiFi APs, offline calibration points, online calibration points and online testing points in the test-bed.

in Figure 5.4, 8 D-Link DIR-605L WiFi Cloud Routers are utilized as WiFi APs in the experiments. In order to collect WiFi RSS fingerprints from multiple APs simultaneously, an Android application that can collect data once per second is developed. As shown in Figure 5.5, this Android app is installed on a Samsung I929 Galaxy SII mobile phone. We leveraged this mobile phone to collect the WiFi RSS fingerprints at the offline calibration points, online calibration points and online testing points. After that, all of this information was sent to a process server running on a PC. The performance evaluation was conducted on the server side.

The initial OS-ELM model was built by the following steps. During the offline phase, 30 offline calibration points were selected, and 1000 WiFi RSS fingerprints were collected at each point. The positions of these 30 offline calibration points are demonstrated in Figure 5.4. The grid spacing between two adjacent locations of the calibration points was chosen to be larger than 1.25 m based on the analysis in [140]. By leveraging these WiFi RSS fingerprints and their physical positions as training inputs and training targets accordingly, the initial OS-ELM model was constructed. During the online phase, we continued to collect WiFi RSS fingerprints at several online calibration points and online testing points for five days. In each



Figure 5.5: The test mobile device Samsung I929 and developed Android apps.

day, two distinct online calibration points and two distinct online testing points were selected in order to reflect the environmental dynamics. The positions of these total 10 online calibration points and 10 online testing points are also presented in Figure 5.4. One thousand WiFi RSS fingerprints are collected at each point. Unlike [22], which only tested the performance of OS-ELM in the corridors in the IoT testbed, as shown in Figure 5.4, the positions of the online calibration points and online testing points were selected on the tables and cubicles, as well as in the testbed to acquire a more comprehensive performance evaluation of the proposed OS-ELM approach. In order to obtain the online RSS measurements to update the OS-ELM localization model without introducing any extra infrastructure, we upgrade the software of COTS WiFi APs so they could capture data packets in the existing WiFi traffic, and retrieve RSS readings and MAC address from the packets for RSS data acquisition. Since the upgraded AP can obtain the real-time RSS readings of other APs as well, all the APs are becoming natural online reference points. The detailed methodology is illustrated in Chapter 7.

5.3.2 Selection of parameters for the initial OS-ELM model

Based on the analysis in Section 5.1, the type of activation function and the number of hidden nodes in the OS-ELM hidden layer are two key parameters affecting the

performance of OS-ELM during the initialization phase, which is accordingly the offline calibration phase in our case.

5.3.2.1 Selection of the type of activation function G for the initial OS-ELM model

By using the 30,000 WiFi RSS fingerprints we collected during the offline calibration phase, we evaluate the performance of three different activation functions: radial basis function (RBF) $G(a, b, x) = e^{-b||x-a||^2}$, sine function $G(a, b, x) = \sin(ax+b)$ and hard-limit transfer (hardlim) function $G(a, b, x) = \text{hardlim}(ax + b)$, with different numbers of hidden nodes.

It can be seen from Figure 5.6 that as the number of hidden nodes increases, the performances in terms of the mean localization accuracy of the sine function and hardlim function become better. On the contrary, the performance of RBF is the worst and appears to be irrelevant with respect to the number of hidden nodes. Since the performance of the hardlim function is the best among the three activation functions, it is chosen as the activation function for the initial OS-ELM model and also for the online sequential learning of OS-ELM.

5.3.2.2 Selection of the number of hidden nodes L for the initial OS-ELM model

Another critical parameter for the performance of the OS-ELM approach is the number of hidden nodes L in the initial OS-ELM hidden layer, after the hardlim function is chosen as the activation function. The five-fold cross-validation method is employed with a range from 0 to 1000 and a step size of 50 in order to determine the optimal L . As observed in Figure 5.6, the performances in terms of the localization accuracy of the hardlim function become relatively stable when L is increased to 600.

In addition to the evaluation of the localization accuracy, the repeatability (REP) of the initial OS-ELM with a different number of hidden nodes is also evaluated

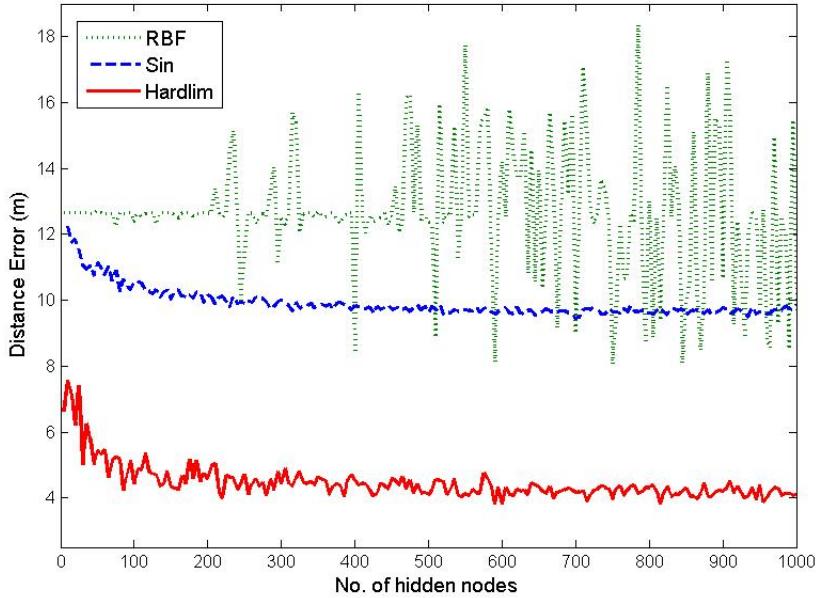


Figure 5.6: Localization accuracy regarding different activation functions and different numbers of hidden nodes.

by leveraging the 30,000 WiFi RSS fingerprints that we collected during the offline calibration phase. REP is measured by the standard deviation of localization errors over the r repeated realizations, and this measure is proposed based on the fact that ELM with the same parameters and in the same training dataset may draw quite different results. A smaller value of REP is desired in general. The mean localization error \bar{e} and REP is calculated based on the following equations:

$$\bar{e} = \frac{1}{r} \sum_{m=1}^r e_m \quad (5.10)$$

$$REP = \sqrt{\frac{1}{r} \sum_{m=1}^r (e_m - \bar{e})^2} \quad (5.11)$$

where r in our experiment is selected as 50 by using fivefold cross-validation. Figure 5.7 demonstrates the repeatability of the initial OS-ELM, which is measured by the standard deviation of localization error with a range of hidden nodes from zero to 1000 and a step size of 50. As observed from Figure 5.7, the standard deviation

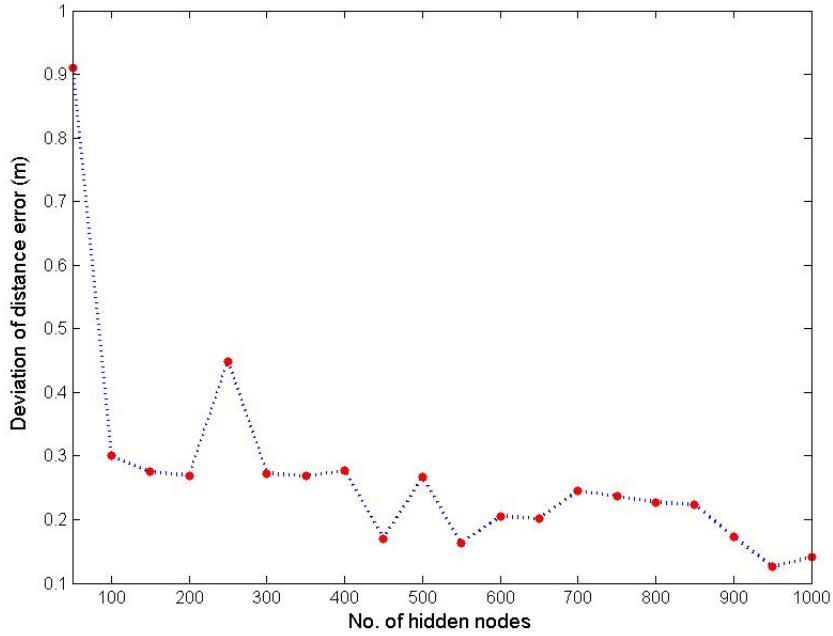


Figure 5.7: Deviation of distance error with different numbers of hidden nodes.

of localization errors decreases when the number of hidden nodes is increased. After comprehensive evaluations on both localization accuracy and repeatability, L is selected to be 950.

A guideline for selecting the type of activation function and the number of hidden nodes in the OS-ELM hidden layer, both of which are the critical parameters for the performance of the OS-ELM approach, is listed as follows: the suggested default activation function for the OS-ELM approach is the hardlim function, whose performance is better than others generally in terms of simulation and experimental results. As for the optimal number of hidden nodes, the five-fold cross-validation method is employed with a range from 0 to 1000 and a step size of 50 based on the empirical tuning.

5.3.3 Comparison between OS-ELM and other methods

Three well-known algorithms, KNN [65], Fuzzy KNN [147] and batch ELM [96], are chosen to compare with the proposed OS-ELM approach. It has been shown

in [17, 19] that the performance of batch ELM in terms of the offline training time, the online testing time and the average localization accuracy is better than classical machine learning algorithms such as the back-propagation (BP) algorithm and the support vector machine for regression (SVR) algorithm. Therefore, we choose batch ELM to be compared with the proposed OS-ELM. Considering the wide usage of KNN and fuzzy KNN as classical localization algorithms, we also include them in the comparison.

Unlike OS-ELM, which can update and revise the initial OS-ELM model sequentially during the online phase, KNN, fuzzy KNN and batch ELM can only utilize the data collected during the offline phase. In order to make a fair comparison, we collected WiFi RSS fingerprints, not only at the 30 offline calibration points, but also at the 10 online calibration points, to build up the WiFi RSS fingerprints offline database during the offline calibration phase. One thousand WiFi RSS fingerprints were collected at each point. The hardlim function and 950 hidden nodes in the hidden layer were chosen for batch ELM offline training, which are the same as for the initial OS-ELM offline training. The batch ELM model was obtained by leveraging the WiFi RSS fingerprints and their corresponding locations stored in the database. During the online phase, after feeding the WiFi RSS fingerprints into the batch ELM model, this model will output the estimated location of the target. For KNN and fuzzy KNN, the value of K was chosen empirically based on the data stored in the database. During the online phase, by matching the measured WiFi RSS fingerprints with the K closest WiFi RSS fingerprints in the database, the location of the target will be calculated. The detailed methodologies of these are presented in [65, 147].

The comparison of performance will be conducted as follows. First of all, we evaluate the performance of KNN, fuzzy KNN, batch ELM and OS-ELM without online sequential learning based on the WiFi RSS fingerprints we collected at the 10 online testing points during the online localization phase. Table 5.2 demonstrates the performance comparison among KNN, fuzzy KNN, batch ELM and OS-ELM in terms of the training time, the testing time and the average localization accuracy.

Table 5.2: Comparison between OS-ELM and other methods.

Approaches	Number of Calibration Points (Offline + Online)	Training Time (s)	Testing Time (s)	Accuracy (m)
KNN	40 + 0	-	0.108	3.098
Fuzzy KNN	40 + 0	-	0.117	2.728
Batch ELM	40 + 0	11.574	0.113	2.581
OS-ELM	30 + 0	8.101	0.113	2.615
	30 + 2	1.204	0.111	2.487
	32 + 2	1.213	0.114	2.346
	34 + 2	1.182	0.109	2.219
	36 + 2	1.226	0.115	2.091
	38 + 2	1.229	0.117	1.973

As shown in Table 5.2, although the localization accuracy of the initial OS-ELM is slightly worse than that of batch ELM, the offline training time of OS-ELM is less than that of batch ELM by 30.01%, which evidently reduces the time and manpower costs for the offline site survey. The testing times of KNN, fuzzy KNN, batch ELM and OS-ELM are almost the same.

During the online phase, since we collected WiFi RSS fingerprints at two different online calibration points at each time, the performance of OS-ELM with each online sequential learning update is also presented in Table 5.2. As observed from Table 5.2, the average localization accuracy of OS-ELM becomes better when more WiFi RSS fingerprints at different online calibration points have been learned online. In addition, another noteworthy point is that the online sequential learning time of OS-ELM is quite fast. It only takes 1.2108 s on average to calculate the output weights β for 2000 newly-received WiFi RSS fingerprints in each online sequential learning update.

As observed from Table 5.2, the average localization accuracy by using KNN, fuzzy KNN and batch ELM is, respectively, 3.098 m, 2.728 m and 2.581 m. In contrast, with online sequential learning of WiFi RSS fingerprints at 10 different online calibration points, OS-ELM can provide a localization accuracy of 1.973 m, which enhances the precision of indoor localization by 36.31%, 27.68% and 23.56% over KNN, fuzzy KNN and batch ELM, respectively. The comparison of the cumulative

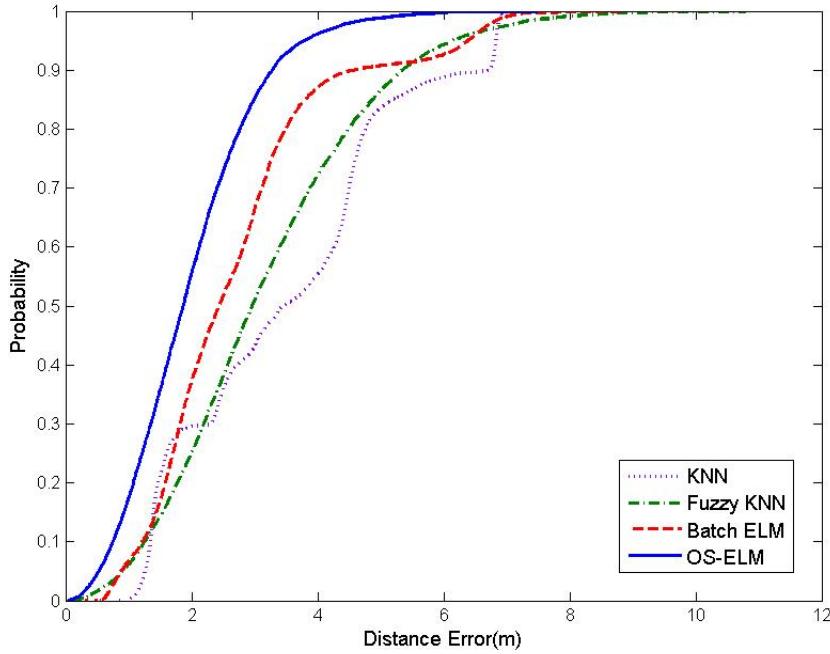


Figure 5.8: Cumulative percentile of error distance for different methods.

percentile of error distances between KNN, fuzzy KNN, batch ELM and the latest updated OS-ELM is presented in Figure 5.8. Figure 5.9 demonstrates the distance error distribution of the four approaches. The distance error distribution of OS-ELM, as shown in Figure 5.9(d), ranges mainly within 2.5 m. On the contrary, the distance error distributions of KNN in Figure 5.9(a), fuzzy KNN in Figure 5.9(b) and batch ELM in Figure 5.9(c) are much more scattered.

In summary, based on our experimental results and analysis, we can conclude that OS-ELM can provide higher localization accuracy consistently with a fast online sequential learning speed to adjust to various environmental dynamics than existing approaches, such as KNN, fuzzy KNN and batch ELM.

5.3.4 Performance evaluation of OS-ELM under specific environmental dynamics

Variations of occupancy distribution and events of opening and closing of doors have been observed as two major environmental factors that could severely affect

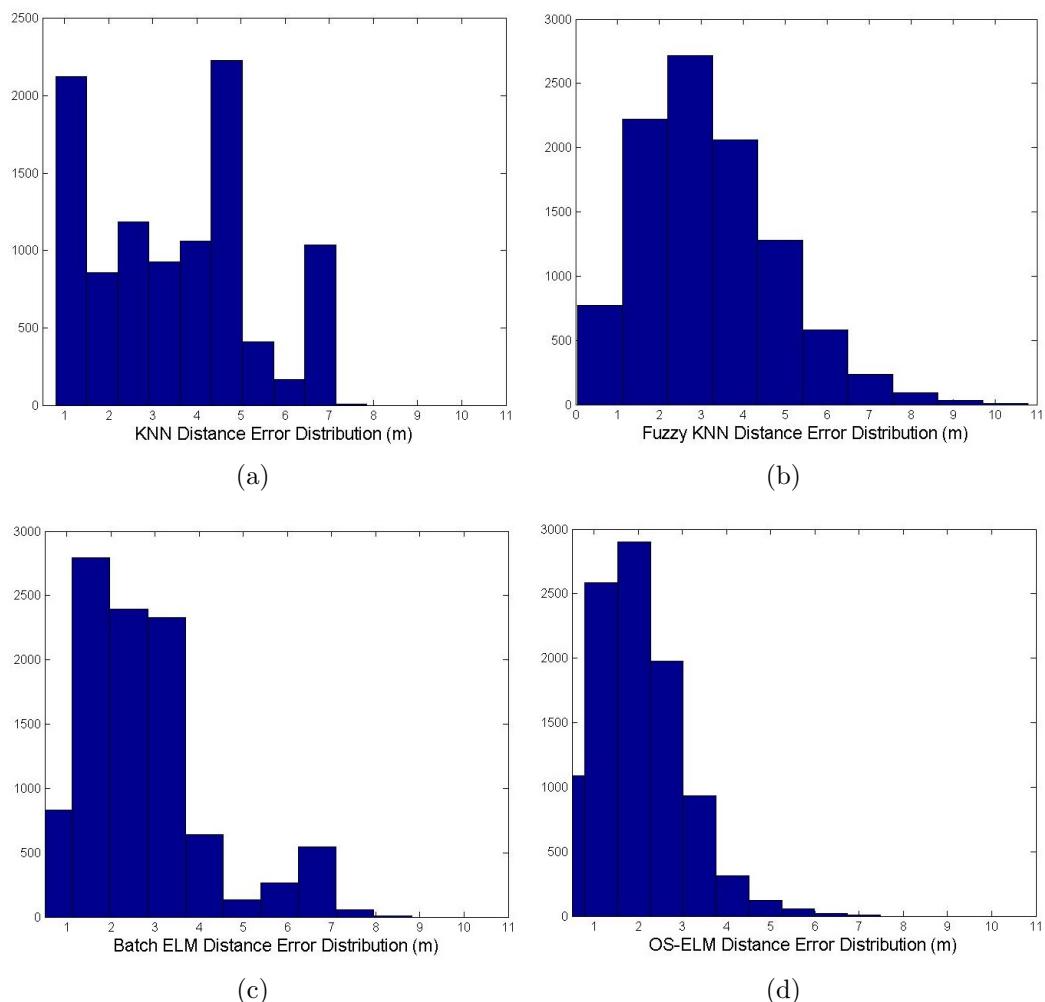


Figure 5.9: Comparison of the distance error distribution for different methods. (a) KNN; (b) fuzzy KNN; (c) batch ELM; (d) OS-ELM.

t the localization accuracy of WiFi based IPS, since they change frequently over time in the indoor environment [142]. Besides the general performance evaluation of the proposed OS-ELM approach, as shown in [22] and the above sections, we also conducted experiments to evaluate how well OS-ELM adapts to some specific environmental dynamics. During each experiment, we measured the localization accuracy of the system when only one environmental factor was altered, while others remained unvaried.

5.3.4.1 Impact of human presence and movements

This experiment aims to determine how well the proposed OS-ELM approach can adjust to the interferences of human presence and movements while other environmental factors are kept unchanged. The experiment was conducted in the conference room in the testbed, since it is easier to create and manage the variation of the occupancy distribution for performance evaluation in this relatively small space. As shown in Figure 5.4, the location of the conference room is the left upper corner of the testbed. The area of the conference room is around 22 m^2 ($6.3\text{ m} \times 3.5\text{ m}$).

The experiment was conducted based on the following steps. First of all, we continued to collect WiFi RSS fingerprints at five offline calibration points for one day when nobody was in the conference room. After that, the batch ELM model and the initial OS-ELM model were built up based on the collected WiFi RSS fingerprints and their corresponding positions. We put the testing mobile device on the conference table to evaluate the performances of the batch ELM and the OS-ELM under the non-human interference condition in the first place. As shown in Table 5.3, the batch ELM and the OS-ELM provide similar localization accuracy under the non-human interference condition, which are 2.203 m and 2.183 m, respectively.

Then, we evaluate the performance of both batch ELM and OS-ELM with human interference. The position of the online testing point was unchanged. However, five people intentionally sat around the table for 6 h. The purpose is to use human bodies to disturb or block the WiFi signal transmission between the test mobile device and

Table 5.3: Impact of human presence and movements on localization accuracy.

Localization Accuracy (m)	Non-Human Interference	With Human Interference
Batch ELM	2.203	3.106
OS-ELM	2.183	2.197

the multiple WiFi APs. Meanwhile, WiFi RSS fingerprints were collected at three online calibration points to update and revise the OS-ELM model.

After conducting the experiment for 6 hours, the average localization accuracies of the two approaches are calculated and demonstrated in Table 5.3. It can be seen from Table 5.3 that the performance of batch ELM becomes worse with human interference. Its performance decay is 41.99% compared with non-human interference. The main reason is that the batch ELM model was constructed under non-human interference, which cannot reflect the variation of occupancy distribution adaptively during the online phase. On the contrary, by leveraging the information collected at the online calibration points, OS-ELM can nicely capture the event of human presence and movements and take this environmental change into the online localization process sequentially. As shown in Table 5.3, the average localization accuracy provided by OS-ELM under the human interference circumstance is 2.197 m, almost the same as the non-human interference scenario. It enhances the precision of indoor localization by 29.27% over batch ELM under the same condition.

To conclude, OS-ELM can provide high localization accuracy consistently when the occupancy distribution is altered due to its fast adaptation.

5.3.4.2 Impact of opening/closing of doors

Besides the variation of occupancy distribution, another environmental factor that affects the performance of WiFi based IPS badly is the events of opening and closing of doors. Therefore, we also conducted an experiment to evaluate how well the proposed OS-ELM approach can adjust to opening and closing of doors while other environmental factors were kept unchanged.

Table 5.4: Impact of opening/closing doors on localization accuracy.

Localization Accuracy (m)	All Doors are Opened	All Doors are Closed
Batch ELM	2.214	4.203
OS-ELM	2.107	2.085

The experiment was conducted based on the following steps. We collected WiFi RSS fingerprints at 15 offline calibration points, which were near the four doors inside the testbed for one day under the all-door-open condition firstly. The batch ELM model and the initial OS-ELM model were built up by leveraging the collected WiFi RSS fingerprints and their corresponding positions. Then, we determined the performance of these two approaches near the door area. It can be seen from Table 5.4 that both batch ELM and OS-ELM can provide nearly 2.15-m localization accuracy on average when all doors are opened.

After that, we continued the experiment to assess the performance of these two approaches under the all-door-close condition. We closed the four doors within the testbed for 6 hours and collected WiFi RSS fingerprints at the same online testing points as under the all-door-open condition. WiFi RSS fingerprints were also collected at the 10 online calibration points in the testbed simultaneously for updating the OS-ELM model.

The average localization accuracy of batch ELM and OS-ELM under the all-door-close condition is presented in Table 5.4. As shown in Table 5.4, the average localization accuracy of batch ELM is 4.203 m when all doors are closed, which is 47.32% worse than the all-door-open condition. This reveals that batch ELM is not capable of providing a sufficient and satisfactory indoor positioning service when the status of doors is different from the offline calibration phase.

In contrast, the performance of OS-ELM is more stable and robust when the status of doors is altered. It can be seen in Table 5.4 that the average localization accuracy provided by OS-ELM under the all-door-close condition is 2.085 m, which is roughly the same as the all-door-open condition. Moreover, it enhances the precision of

indoor localization by 50.39% over batch ELM. The main reason that OS-ELM can consistently provide the indoor positioning service under different status of doors is due to its online sequential learning ability. By leveraging the information collected at the online calibration points, OS-ELM can sequentially update its model to adjust the influence from opening or closing doors.

Thus, the experimental results leads to the same conclusion that OS-ELM can provide higher localization accuracy steadily under various environmental dynamics, regardless of the impact from human presence or movements or the influence from changing the status of doors.

5.4 Conclusion

In this chapter, we proposed an indoor localization algorithm based on OS-ELM to address the two challenging problems of the existing WiFi based IPS: the intensive costs of manpower and time for offline site survey and the robustness to environmental dynamics. Both our simulation analysis and experimental studies have shown that OS-ELM can tackle the problems satisfactorily. The fast learning speed of OS-ELM obviously reduced the time consumption and manpower costs for the site survey during the offline calibration phase. In addition, the online sequential learning ability of OS-ELM made it possible to reflect and adapt to the environmental changes in a timely manner. Furthermore, WiFi RSS fingerprints can be collected and updated more flexibly, since OS-ELM can learn data with a varying chunk size. Experiments under specific environmental dynamics such as variations of the occupancy distribution and events of opening or closing doors were also conducted. In summary, OS-ELM can provide high localization accuracy with a fast online sequential learning speed under various environmental changes and achieve superior performance to the existing approaches. Meanwhile, since OS-ELM requires online measurements to be collected for updating the localization model, mobile devices at known locations are needed. Feasible crowdsourcing methods are desired for online data acquisition.

Chapter 6

Access Point Selection Strategy

In Chapter 4 and Chapter 5, we have proposed effective solutions to overcome the device heterogeneity issue and the vulnerability to environmental dynamics for WiFi based IPS. Another issue that limits the performance of WiFi based IPS is the instability of WiFi RSS. In this chapter, a novel AP selection strategy that can select the optimal subset of APs that provides the most valuable information for improving the localization accuracy of WiFi based IPS is proposed.

The large variation of WiFi RSS is caused by multiple factors such as multi-path effect, reflection and interference of signal propagation, variations of humidity and even variations of occupancy distribution. One straightforward and common solution to address this issue is to leverage as many available APs as possible in indoor environments to improve the localization accuracy since the deployment of APs is dense in indoor environments and each AP may provide different location related information. However, in this case the localization algorithm needs to handle high computational burden due to the usage of large numbers of APs. One major challenge for existing WiFi based IPSs is how to reduce this computational burden while preserving or even improving the localization accuracy.

Some existing approaches to address this issue are leveraging feature extraction methods [99, 100], which aim to extract the most valuable feature components to

be used for localization. However, the larger variation of WiFi RSS limits the improvement of these approaches. Another category of approaches to handle this issue is the AP selection method. This method is to select a subset of APs out of all the available APs to reduce the computational cost [101,102]. Nevertheless, there are two main drawbacks in existing approaches. One is that the dependence between APs is not considered in the AP selection process. Assuming the independence between APs, the discriminate ability of APs is measured and sorted based on the individual importance. The lack of comprehensive analysis and consideration of the dependence among APs may lead to a biased solution. Another drawback is that all these traditional methods select APs during the offline phase and these selected APs cannot provide real-time information to ameliorate the localization accuracy during the online phase due to various environmental dynamics. Furthermore, the discriminate ability of each AP varies with locations. Therefore, the AP selection process should be conducted online and correlated with the real-time location of the mobile device adaptively.

In this chapter, we focus on how to reduce the computational burden for WiFi based IPS by selecting a subset of APs that provide the most useful information for localization. Existing approaches select and measure the discriminate ability of APs individually during the offline phase without the consideration of the dependence among them. We propose a novel online mutual information (OnlineMI) AP selection strategy that measures the collective discriminate ability among APs based on their mutual information. Furthermore, the proposed AP selection process is conducted online to adapt to various environmental dynamics. The weighted K nearest neighbor (WKNN) approach is further employed as the localization algorithm after the OnlineMI AP selection. Simulation and experiments are carried out, and performance evaluation and comparison with existing methods demonstrate the superiority of the proposed OnlineMI-WKNN approach.

The rest of the chapter is organized as follows. Section 6.1 presents the proposed OnlineMI AP selection strategy. The simulation results and evaluation are demonstrated in Section 6.2. In Section 6.3, our experimental setup and procedure are

elaborated first, and then the experimental results and performance evaluation of the proposed scheme are reported. We conclude the work in Section 6.4.

6.1 OnlineMI based AP Selection Strategy and Localization Algorithm

6.1.1 OnlineMI AP selection strategy

In contrast to the traditional approaches, in this section, we propose a novel AP selection strategy, namely OnlineMI, which selects the optimal subset of available APs according to the consideration of the dependence (e.g. mutual information MI) among them and conducts the selection process during the online phase based on the real-time location of the mobile device and other environmental dynamics.

The methodology of OnlineMI is introduced as follows: assume that N_t APs have been deployed in the indoor environment. At some specific locations, certain APs may not detect the mobile device due to long distance between them. Therefore, we define the minimum detectable WiFi RSSI value to be -100dBm in the situation. Suppose that the mobile device can receive WiFi RSSI from N_t APs and q RSS samples have been collected and stored at each online point. Each online RSS sample can be represented as $R_{ol}^t = [RSS_{ol}^1, RSS_{ol}^2, \dots, RSS_{ol}^{N_t}]$ from N_t APs. Our objective is to select the N_f APs with the best discriminate abilities from total N_t APs to reduce the computational burden and preserve or even enhance the localization accuracy of the IPS.

The 1st step of OnlineMI AP selection strategy is to calculate the mutual information between each pair of available APs. There are $N_t(N_t - 1)/2$ combinations in total. For instance, the mutual information of AP_a and AP_b is obtained as follows:

$$MI(AP_a, AP_b) = H(AP_a) + H(AP_b) - H(AP_a, AP_b)$$

where $H(AP_a)$ and $H(AP_b)$ are the entropy of the RSS received from AP_a and AP_b respectively; $H(AP_a, AP_b)$ is the joint entropy of AP_a and AP_b , which can be calculated as follows:

$$H(AP_a, AP_b) = \sum_{v_2} \sum_{v_1} [Pr(RSS_{ol}^a = v_1, RSS_{ol}^b = v_2) \times \\ \log Pr(RSS_{ol}^a = v_1, RSS_{ol}^b = v_2)]$$

where v_1 and v_2 are the possible RSS values received from AP_a and AP_b respectively; $Pr(RSS_{ol}^a = v_1, RSS_{ol}^b = v_2)$ is the joint probability when $RSS_{ol}^a = v_1, RSS_{ol}^b = v_2$. Since mutual information measures the information that is shared between random variables, smaller $MI(AP_a, AP_b)$ indicates less dependence between AP_a and AP_b , and more distinct and valuable information than other pairs. The procedure to select the first two APs s_1, s_2 of OnlineMI is presented as follows:

$$(s_1, s_2) = \arg \min_{\{a,b\} \subseteq t} MI(AP_a, AP_b)$$

The 2nd step of OnlineMI is to select one more AP_{s_3} based on the following criterion:

$$s_3 = \arg \min_{z \in t \setminus \{s_1, s_2\}} MI(AP_{s_1}, AP_{s_2}, AP_z)$$

$N_t - 2$ cases of $MI(AP_{s_1}, AP_{s_2}, AP_z)$ need to be considered. For each case, we can leverage the result of $H(AP_{s_1}, AP_{s_2})$ directly, which has been calculated in the 1st step. The remaining two terms $H(AP_z)$ and $(AP_{s_1}, AP_{s_2}, AP_z)$ can be obtained as follows:

$$H(AP_z) = - \sum_{v_z} Pr(RSS_{ol}^z = v_z) \times \\ \log Pr(RSS_{ol}^z = v_z)$$

$$MI(AP_{s_1}, AP_{s_2}, AP_z) = H(AP_{s_1}, AP_{s_2}) + H(AP_z) \\ - H(AP_{s_1}, AP_{s_2}, AP_z)$$

and

$$H(AP_{s_1}, AP_{s_2}, AP_z) = \sum_{v_z} \sum_{v_2} \sum_{v_1} [Pr(RSS_{ol}^{s_1} = v_1, RSS_{ol}^{s_2} = v_2, RSS_{ol}^z = v_z) \times \\ \log\{Pr(RSS_{ol}^{s_1} = v_1, RSS_{ol}^{s_2} = v_2, RSS_{ol}^z = v_z)\}]$$

where v_1 , v_2 and v_z are the possible RSS values received from AP_{s_1} , AP_{s_2} and AP_z respectively; $Pr(RSS_{ol}^{s_1} = v_1, RSS_{ol}^{s_2} = v_2)$ is the joint probability when $RSS_{ol}^{s_1} = v_1, RSS_{ol}^{s_2} = v_2, RSS_{ol}^z = v_z$.

The following steps of OnlineMI aim to select additional one more AP at each step. In general, the l step of OnlineMI can be conducted as follows:

$$s_{l+1} = \arg \min_{z \in t \setminus \{s_1, \dots, s_l\}} MI(AP_{s_1}, \dots, AP_{s_l}, AP_z)$$

where

$$MI(AP_{s_1}, \dots, AP_{s_l}, AP_z) = H(AP_{s_1}, \dots, AP_{s_l}) + H(AP_z) - H(AP_{s_1}, \dots, AP_{s_l}, AP_z)$$

There are $(N_t - l)$ combinations of $MI(AP_{s_1}, \dots, AP_{s_l}, AP_z)$.

For each case, since $H(AP_{s_1}, \dots, AP_{s_l})$ has been obtained in the $(l - 1)th$ step, the terms in $MI(AP_{s_1}, \dots, AP_{s_l}, AP_z)$ to be calculated in this step are only $H(AP_z)$ and $H(AP_{s_1}, \dots, AP_{s_l}, AP_z)$. $H(AP_{s_1}, \dots, AP_{s_l}, AP_z)$ can be calculated as follows:

$$H(AP_{s_1}, \dots, AP_{s_l}, AP_z) = \sum_{v_z} \dots \sum_{v_1} [Pr(RSS_{ol}^{s_1} = v_1, \dots, RSS_{ol}^{s_l} = v_l, RSS_{ol}^z = v_z) \times \\ \log\{Pr(RSS_{ol}^{s_1} = v_1, \dots, RSS_{ol}^{s_l} = v_l, RSS_{ol}^z = v_z)\}]$$

where v_1, \dots, v_l, v_z are the possible RSS values received from $AP_{s_1}, \dots, AP_{s_l}$ and AP_z ; $Pr(RSS_{ol}^{s_1} = v_1, \dots, RSS_{ol}^{s_l} = v_l, RSS_{ol}^z = v_z)$ is the joint probability when $RSS_{ol}^{s_1} = v_1, \dots, RSS_{ol}^{s_l} = v_l, RSS_{ol}^z = v_z$.

By following the procedure introduced above, N_f APs with the best discriminate abilities from total N_t available APs will be selected online. The computational cost for localization process is reduced since the rest of $(N_t - N_f)$ APs are discarded by the OnlineMI AP selection strategy.

6.1.2 OnlineMI-WKNN localization algorithm

One of the well-known and widely used classical fingerprinting localization algorithms is WKNN [65]. We further propose an indoor localization scheme, OnlineMI-WKNN that integrates the advantages of them.

After we select N_f APs with the best discriminate abilities based on OnlineMI AP selection strategy, only these N_f APs will be leveraged for the following localization process. Therefore, each online testing RSS sample can be reformatted as $R_{ol}^f = [RSS_{ol}^1, RSS_{ol}^2, \dots, RSS_{ol}^{N_f}]$ from the selected N_f APs. The dimension of the offline WiFi RSS database involved for localization process will be reduced and reformatted from $p \times N_t$ to $p \times N_f$, where p is the total number of offline calibration points. The mean RSS vector obtained at calibration point C_j can be expressed as $R_{C_j} = [RSS_{C_j}^1, RSS_{C_j}^2, \dots, RSS_{C_j}^{N_f}]$.

By leveraging the reformatted online RSS sample and reconstructed offline WiFi RSS database, the real-time location of the mobile device can be estimated as follows:

The 1st step is to calculate the Euclidean distance d_j between the online RSS sample R_{ol} and each mean offline RSS vector R_{C_j} stored in the offline WiFi RSS database. d_j is obtained as follows:

$$d_j = \|R_{ol} - R_{C_j}\|$$

In the 2nd step, the distances ($d_j, 1 \leq j \leq p$) are sorted in ascending order firstly. Then, k offline reference points and their corresponding physical coordinates with

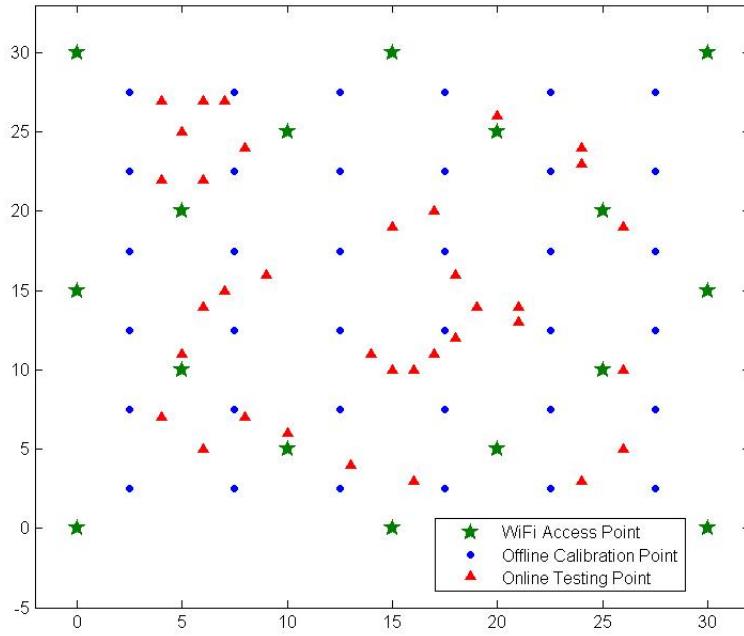


Figure 6.1: Locations of the WiFi APs, offline calibration points and online testing points in the simulated field.

k smallest d_j values are picked up to estimate the real-time location (x, y) of the mobile device. (x, y) is calculated as follows:

$$(x, y) = \frac{1}{c} \sum_{j=1}^k \frac{1}{d_j} (x_j, y_j)$$

$c = \sum_{j=1}^k \frac{1}{d_j}$ is the normalization constant and (x_j, y_j) denote the physical coordinates of the j th calibration point.

6.2 Simulation Results and Evaluation

6.2.1 Simulation setup

We conducted simulations to evaluate the performance of the proposed OnlineMI AP selection strategy and OnlineMI-WKNN localization algorithm. The simulation was conducted on a PC, which has a Intel Core i5-2400 3.10-GHz CPU and 8 GB RAM. We used MATLAB R2014a to develop a simulated indoor environment.

The simulated indoor environment is a $30\text{ m} \times 30\text{ m}$ room. As shown in Figure 6.1, 16 WiFi APs are placed in the room. The WiFi signals broadcast from these APs are simulated based on the International Telecommunication Union (ITU) indoor propagation model [126], the indoor path loss model which estimates the relation between the total path loss PL (dBm) and distance d (m) can be expressed as:

$$PL(d) = PL_0 - 10\alpha \log(d) + X_\sigma$$

where PL_0 is the pass loss coefficient which is set to be -30 dBm in our simulation; α is the path loss exponent which is set as 2.8 and X_σ represents a zero-mean normal random noise with standard deviation $\sigma = 0.5$.

During the offline calibration phase, WiFi RSS fingerprints from the 16 WiFi APs were collected at 36 offline calibration points in order to construct the offline WiFi RSS database. 100 WiFi RSS fingerprints were collected at each point. Then, 36 online testing points were randomly selected during the online phase with 100 WiFi RSS samples collected at each point. The locations of the offline calibration points and the online testing points are shown in Figure 6.1.

6.2.2 Performance evaluation

Two traditional AP selection approaches, MaxMean and InfoGain, are chosen to compare with OnlineMI. MaxMean is a common AP selection approach that employs the absolute mean RSS value as the importance index and chooses those APs with the strongest mean RSS values [101]. Apart from the RSS value, InfoGain utilizes the individual information gain of AP as the importance index to assess the discriminative ability of each AP [102]. These existing approaches do not consider the dependence between APs in the AP selection process. The discriminative ability of APs is measured and sorted based on the individual importance. The key difference between OnlineMI and these approaches is that it measures the collective discriminative ability of each group of APs based on mutual information among

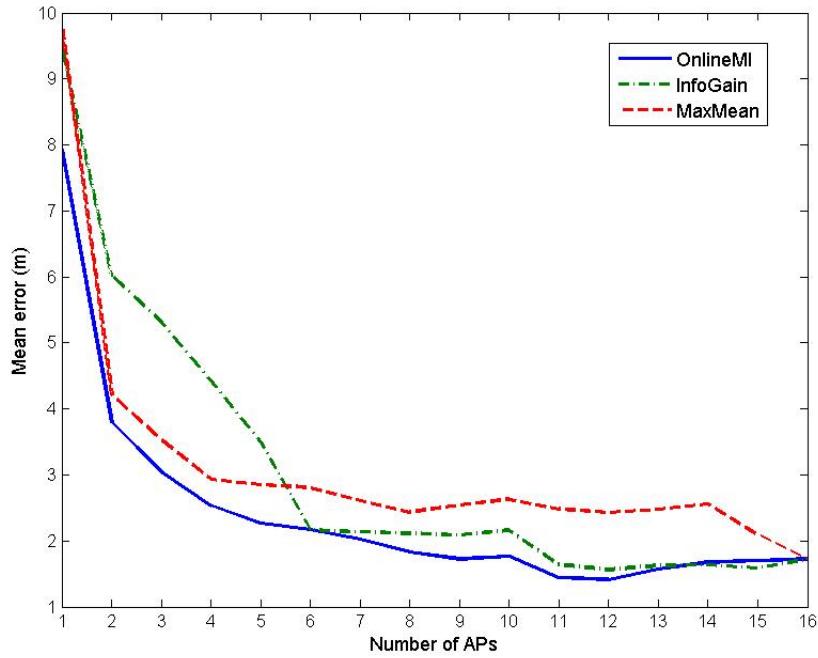


Figure 6.2: Comparison of mean error between different AP selection methods for different numbers of APs.

them instead of ranking the individual discriminative ability of each AP. After the AP selection process, the WKNN localization algorithm was adopted in conjunction with these two AP selection approaches in order to make a fair comparison with OnlineMI-WKNN. The value of k was chosen as 9 empirically and was kept unique for all the three approaches.

Figure 6.2 presents the comparison of mean errors between different AP selection methods for different numbers of APs. It can be seen from Figure 6.2 that the localization accuracy in terms of mean error improves as the number of AP increases. OnlineMI outperforms MaxMean and InfoGain under the same number of APs in most of the situations. The mean error value of OnlineMI reduces significantly from one to three APs, then gradually reduces from four to ten APs, reduces slightly from ten to twelve APs, and saturated after thirteen. The smallest mean error of OnlineMI is 1.42m when twelve APs are selected. This indicates that some APs are redundant for localization purpose in the indoor environment. We further analyze the performance of OnlineMI under two cases: 50% of APs (8 APs) and 75% of APs (12 APs) selected respectively.

Table 6.1: Performance comparison of different methods (No. of AP: 8)

Methods	Mean Error (m)	Variance of Error	Accuracy within 2 m	Accuracy within 3 m
MaxMean	2.44	1.47	46.83%	62.58%
InfoGain	2.16	1.34	55.83%	76.67%
OnlineMI	1.83	0.77	61.17%	91.58%

Table 6.2: Performance comparison of different methods (No. of AP: 12)

Methods	Mean Error (m)	Variance of Error (m)	Accuracy within 1 m	Accuracy within 2 m
MaxMean	2.43	1.35	15.17%	49.17%
InfoGain	1.56	0.71	24.08%	68.25%
OnlineMI	1.42	0.69	31.01%	82.75%
All APs	1.72	0.83	24.83%	65.17%

We present the evaluation of OnlineMI when 50% of APs (8 APs) were selected first. Table 6.1 demonstrates the performance comparison among OnlineMI, InfoGain and MaxMean in terms of mean error, variance of error, accuracy within 2m and accuracy within 3m. As shown in Table 6.1, OnlineMI can provide 1.83m localization accuracy even by employing only half of total available APs. It enhances the precision of indoor positioning by 15.28% over InfoGain and 25% over MaxMean. OnlineMI can reach the accuracy within 3m for 91.58% by using only 8 APs while other two approaches require more APs. In addition, its error variance is the smallest among the three approaches, which indicates that the APs selected through OnlineMI provide more useful and reliable information for indoor positioning. The localization accuracy of OnlineMI with half of total APs is only 6% lower than All APs situation. Therefore, OnlineMI can reduce nearly 50% computational burden while providing acceptable localization accuracy by only employing half of total available APs.

We further compare the performance of three approaches when 75% of APs (12 APs) were selected. The performance comparison among OnlineMI, InfoGain, MaxMean and All APs in terms of mean error, variance of error, accuracy within 1m and accuracy within 2m is presented in Table 6.4. As observed from Table 6.4, OnlineMI can provide 1.42m localization accuracy, which improves the precision of indoor

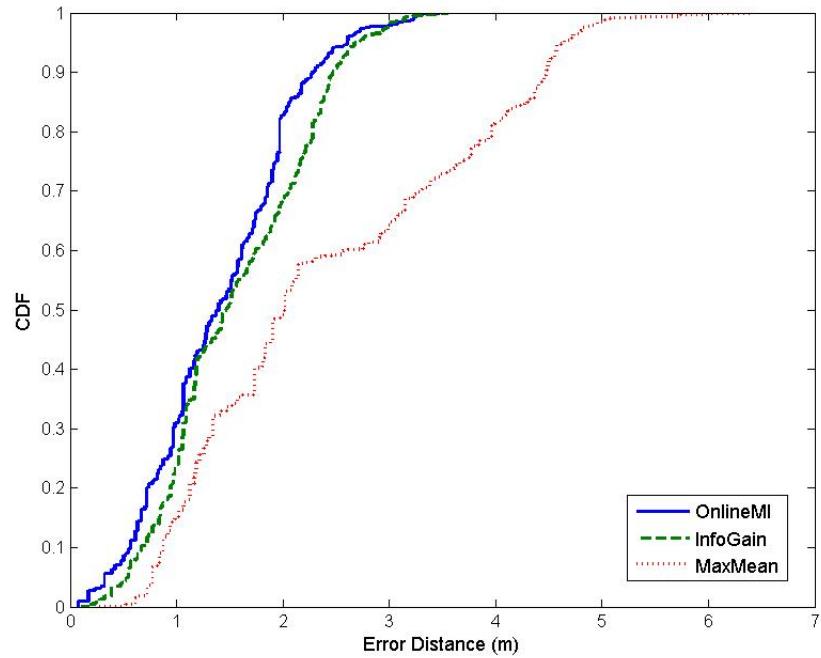


Figure 6.3: Comparison of distance error distribution for different methods, where the number of APs is twelve.

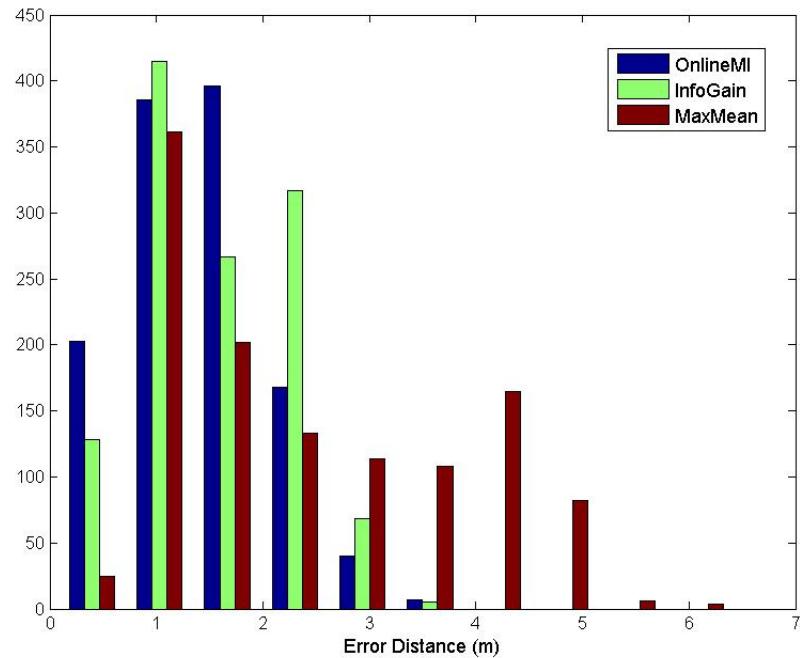


Figure 6.4: Comparison of distance error distribution for different methods, where the number of APs is twelve.

positioning by 8.97% over InfoGain and 41.56% over MaxMean respectively. Other performance metrics, including variance of error, accuracy within 1m and accuracy within 2m, of OnlineMI outperform the two traditional AP selection approaches. Its error variance is the smallest among the three approaches. After the integration with WKNN algorithm, OnlineMI-WKNN can provide the localization accuracy of 82.75% within 2m by using only 75% of total available APs.

Figure 6.3 demonstrates the cumulative distribution of distance errors in this scenario. The distance error distributions of OnlineMI, InfoGain and MaxMean are demonstrated in Figure 6.4. We can obtain the same conclusion from Figure 6.3 and Figure 6.4. As shown in Figure 6.4, the distance errors of OnlineMI are mainly within 3.2 m. By contrast, the distance error distributions of the traditional two approaches are much more scattered.

The performance of OnlineMI in this scenario is also compared with the situation where all APs are leveraged for localization. Its localization accuracy is even 17.44% higher than that of all APs, which verifies that not all APs are useful for indoor positioning, proving the existence of redundancy among them. OnlineMI AP selection strategy can pick up the most discriminant group of APs to enhance the indoor localization and reduce the computational burden simultaneously. The simulation results strongly support that the proposed OnlineMI approach can select a subset of APs that provide non-redundant information for localization consistently.

6.3 Experimental Results and Performance Evaluation

6.3.1 Performance evaluation

Extensive experiments have been conducted in the IoT testbed as introduced in Section 4.3.1 to evaluate the performance of the proposed OnlineMI AP selection strategy and OnlineMI-WKNN localization algorithm. As shown in Figure 6.5, 16

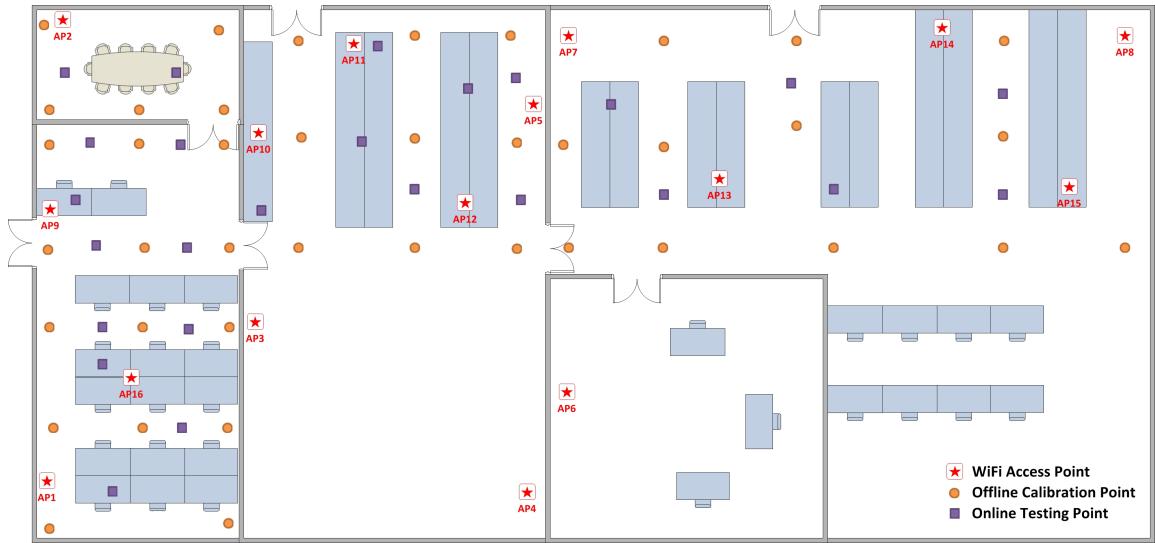


Figure 6.5: Positions of the WiFi APs, offline calibration points and online testing points in the testbed.

TP-LINK TL-WDR4300 Wireless N750 Dual Band Routers are adopted as WiFi APs in the testbed. In order to collect WiFi RSS fingerprints from multiple APs simultaneously, an Android application that can collect data is developed. This application is installed on a Samsung Galaxy S4 mobile phone. All the WiFi RSS fingerprints at offline calibration points and online testing points are collected using this phone for performance evaluation. 40 offline calibration points are selected and 100 WiFi RSS fingerprints are collected at each point during the offline phase. Then, 25 online testing points were selected during the online phase, with 100 WiFi RSS samples collected at each point. The locations of the offline calibration points, and the online testing points are demonstrated in Figure 6.5.

Two traditional AP selection approaches, MaxMean and InfoGain, are chosen to compare with OnlineMI. After the AP selection process, the WKNN localization algorithm was adopted in conjunction with these two AP selection approaches in order to make a fair comparison with OnlineMI-WKNN. The value of k was chosen as 11 empirically for all the three approaches.

The overall performance in terms of mean errors between these three AP selection methods under different numbers of APs is demonstrated in Figure 6.6. For all the three approaches, the mean errors decrease as the number of AP in-use increases. It

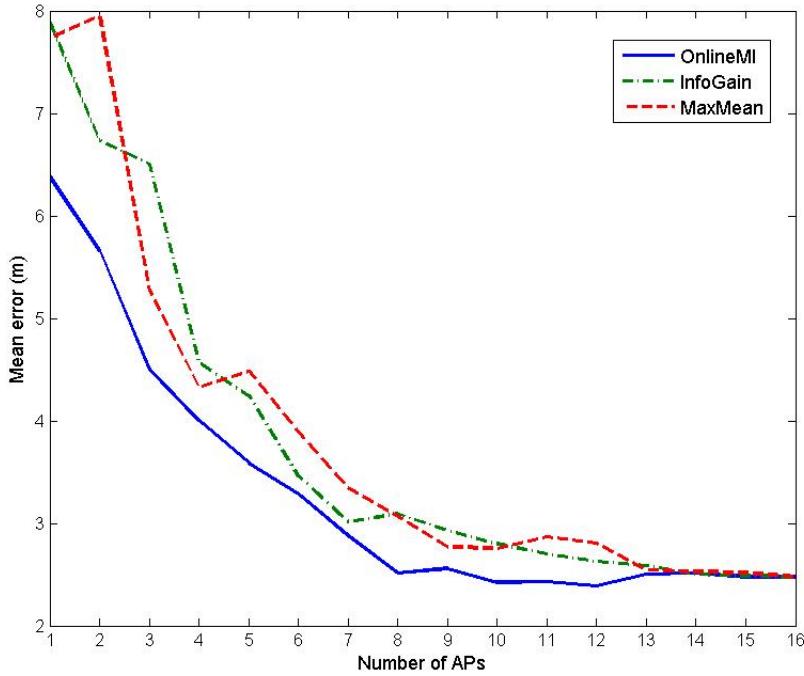


Figure 6.6: Comparison of mean error between different AP selection methods for different numbers of APs.

Table 6.3: Performance comparison of different methods (No. of AP: 8)

Methods	Mean Error (m)	Variance of Error	Accuracy within 2 m
MaxMean	3.07	1.84	38.52%
InfoGain	3.09	1.64	30.72%
OnlineMI	2.52	1.43	46.56%

can be easily observed that OnlineMI outperforms MaxMean and InfoGain under the same number of APs in nearly every situation. To be specific, the mean error value of OnlineMI reduces significantly from one to eight APs, then saturates and keeps in a low mean error level after eight APs. The smallest mean error of OnlineMI is 2.39m when twelve APs were selected. This indicates that some APs are redundant for localization purpose in the indoor environment. We further analyze the performance of OnlineMI under two cases: 50% of APs (8 APs) and 75% of APs (12 APs) were selected respectively.

The evaluation of OnlineMI when 50% of APs (8 APs) were selected is presented first. Table 6.3 demonstrates the performance comparison among OnlineMI, Info-

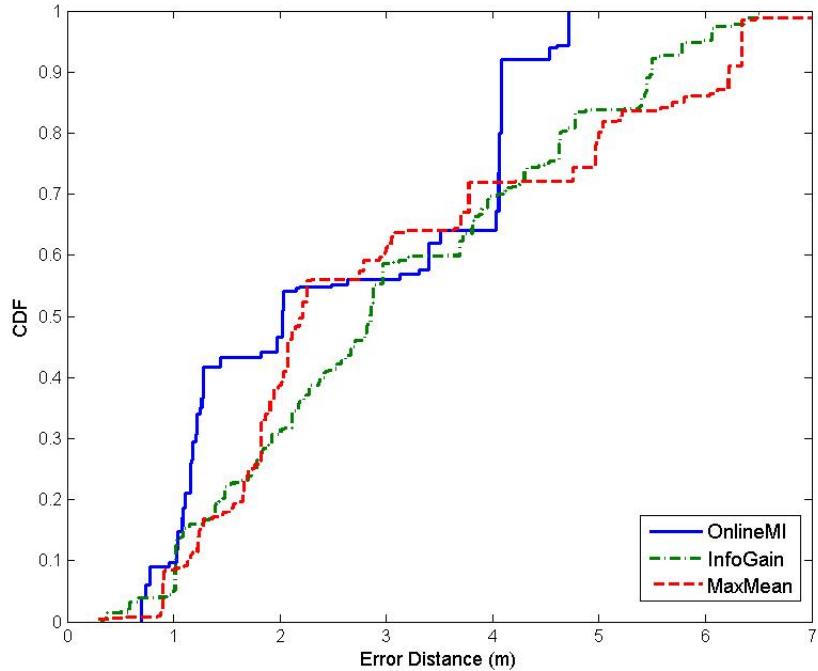


Figure 6.7: Comparison of distance error distribution for different methods, where the number of APs is eight.

Gain and MaxMean in terms of mean error, variance of error and accuracy within 2m. Figure 6.7 demonstrates the cumulative distribution of distance errors in this scenario. As shown in Table 6.3, OnlineMI can provide 2.52m localization accuracy even by employing only half of the total available APs. It enhances the precision of indoor positioning by 18.60% over InfoGain and 17.93% over MaxMean. OnlineMI can reach the accuracy of 46.56% within 2m by using only 8 APs while other two approaches require more APs. In addition, its error variance is the smallest among the three approaches, which indicates that the APs selected through OnlineMI provide more reliable information for indoor positioning. The localization accuracy of OnlineMI with half of the total available APs is only 1.5% lower than All APs situation.

The performance of three approaches when 75% of APs (12 APs) were selected is also evaluated. Figure 6.8 demonstrates the cumulative distribution of distance errors in this scenario. The performance comparison among OnlineMI, InfoGain, MaxMean and All APs in terms of mean error, variance of error, accuracy within 2m and

Table 6.4: Performance comparison of different methods (No. of AP: 12)

Methods	Mean Error (m)	Variance of Error	Accuracy within 2 m	Accuracy within 1 m
MaxMean	2.80	1.94	36.00%	9.80%
InfoGain	2.60	1.59	44.60%	3.96%
OnlineMI	2.39	1.47	47.04%	16.64%
All APs	2.48	1.54	52.00%	8.44%

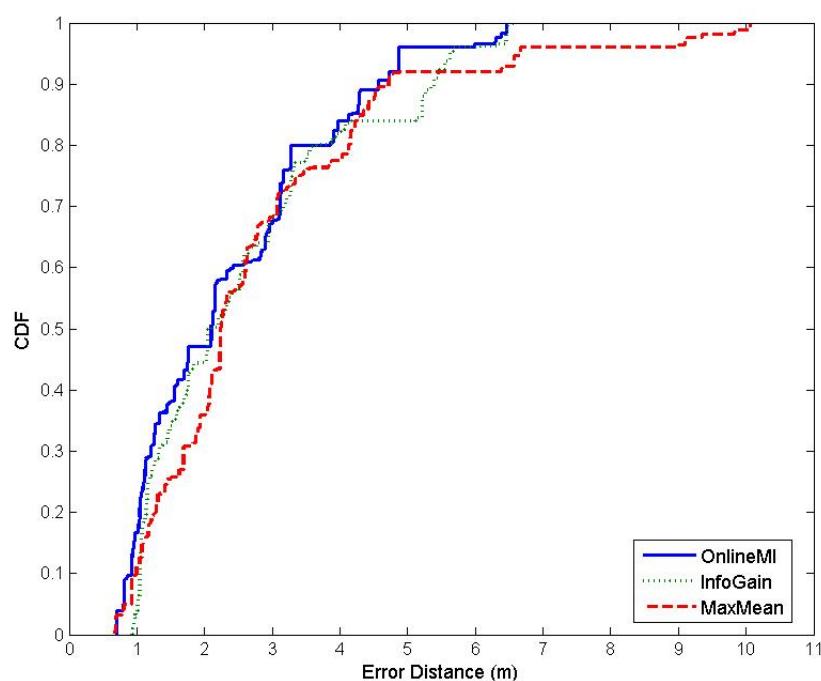


Figure 6.8: Comparison of distance error distribution for different methods, where the number of APs is twelve.

accuracy within 1m is presented in Table 6.4. As observed from Table 6.4, OnlineMI can provide 2.39 m localization accuracy, which improves the precision of indoor positioning by 9.14% over InfoGain and 14.88% over MaxMean respectively. Other performance metrics including variance of error, accuracy within 2m and accuracy within 1m, of OnlineMI outperform the two traditional AP selection approaches. Its error variance is the smallest among the three approaches.

The performance of OnlineMI in this scenario is also compared with the situation where all APs are leveraged for localization. Its localization accuracy is even 3.6% higher than that of all APs. It can provide a smaller variance of error than that of all APs, which verifies that not all APs are useful for indoor positioning. The OnlineMI AP selection strategy can pick up the most discriminant group of APs to enhance the indoor localization and reduce the computational burden simultaneously.

In summary, the performance of OnlineMI when different portions of total APs were selected for indoor localization have been evaluated comprehensively. OnlineMI outperforms existing approaches such as InfoGain and MaxMean due to two reasons. Firstly, it takes the mutual correlations between APs into consideration and measures the collective discriminative ability of each group of APs. Conventional approaches only determine the discriminative ability of APs according to its individual importance. The lack of comprehensive analysis and consideration of the dependence among APs may lead to a biased solution. Secondly, the AP selection process of OnlineMI is conducted online, which selects the APs adaptively based on the real-time location of the mobile device and the environmental dynamics. On the other hand, the traditional approaches select APs during offline phase and keep using them online statically. These offline selected APs may not provide valuable information consistently for indoor localization.

6.4 Conclusion

In this chapter, we proposed a novel AP selection strategy, OnlineMI, which is able to select the optimal subset of APs to improve the indoor localization accuracy and

reduce the computational burden. Unlike traditional AP selection methods that consider the individual discriminate ability of all the APs only, OnlineMI measures the collective discriminate ability of groups of APs based on mutual information among them. Furthermore, since its AP selection process is conducted online associated with the real-time location of the mobile device, OnlineMI can select the subset of APs that provides non-redundant information for indoor positioning consistently and adaptively even under various environmental dynamics online. In addition, WKNN approach is employed as localization algorithm after the OnlineMI AP selection. According to our simulation and experimental results, the proposed OnlineMI-WKNN approach is able to select a subset of APs online that contain most critical information for localization, meanwhile preserve or even enhance the localization accuracy of the entire IPS. Its overall performance such as mean error and variance of error outperforms traditional AP selection methods. The proposed OnlineMI approach is based on mutual information which is the basic measurement of the dependence between two variables. More advanced criteria that is able to quantify the dependence precisely can be explored for AP selection.

Chapter 7

Online Radio Map Construction

We proposed feasible solutions to address several key issues for WiFi-based IPS including the device heterogeneity issue in Chapter 4, the vulnerability to environmental dynamics in Chapter 5 and AP selection in Chapter 6. In this chapter, we further tackle another bottleneck which is the time-consuming and labor-intensive offline site survey process.

Some previous works tried to use indoor radio propagation model for online radio map construction to replace the laborious offline site survey process [77, 114]. However, the simple log-distance path loss model failed to capture the intangible RSS distribution in complex indoor environments. Some works deploy fixed reference anchors to obtain fresh RSS readings for radio map adaptation [54, 112]. Nevertheless, the requirement of extra hardware implementation is the bottleneck of these methods. Learning-based approaches are also introduced to reduce the number of reference anchors to be deployed [115, 116, 148, 149]. However, these methods still need to conduct offline initialization to collect RSS fingerprints as label data for learning purposes. Although certain crowdsourcing methods have been introduced in [72, 117] recently to tackle the mentioned issues, extra user intervention is still needed. Therefore, an efficient and non-intrusive, which is easily implemented scheme for online radio map construction and adaptation is urgently needed.

In this chapter, we propose WinIPS, a WiFi-based non-intrusive indoor positioning system, that enables automatic online radio map construction and adaptation for calibration-free indoor localization to overcome the aforementioned issues of WiFi fingerprinting-based IPS. For RSS data acquisition, we develop WinSMS, a novel intelligent wireless system that can capture data packets transmitted in the existing WiFi traffic and extract the RSS and MAC addresses of both APs and mobile devices (MDs) in a non-intrusive manner without introducing any extra hardware. Since we can obtain the real-time RSS measurements of APs, these APs are becoming natural online reference points for online radio map construction and adaptation. Therefore, we can completely avoid the tedious offline site survey process. Furthermore, in order to build up a more fine-grained radio map, we further propose Gaussian Process Regression (GPR) with Polynomial Surface Fitting Mean (PSFM-GPR), a reliable regression technique dedicated to predict RSS on virtual reference points (VRPs) which can capture the irregular RSS distribution over complex indoor environments precisely. Existing online RSS modeling techniques such as GPR with Log-Distance mean (LDM-GPR) [114] and Geography Weighted Regression (GWR) [113] assume that RSS values at the same distance from the AP are unique which is true in free space but unlikely so in real indoor environment. PSFM-GPR models RSS distribution in two dimensional surface which is closer to practical scenario. In addition, this online radio map is more adaptive and robust to environmental dynamics than traditional offline calibrated RSS database since it is up-to-date all the time. Since the online radio map is based on AP generated RSS values, it is not suitable to localize MDs directly due to the device heterogeneity issue. We leverage signal tendency index (STI) in Chapter 4, which compares the shapes of RSS vectors between RSS readings of MD and online RSS fingerprint database instead of raw RSS values. Then, we propose signal tendency index - weighted K nearest neighbor (STI-WKNN), that adopts the similarity index STI as a novel weighting scheme for WKNN, to improve the localization accuracy of WinIPS across heterogeneous devices. Extensive experiments are carried out for six months to validate the effectiveness of WinIPS in a real-world multi-functional office. The experimen-

tal results demonstrate that PSFM-GPR achieves 4.8 dBm average RSS estimation error, which enhances the accuracy of online RSS prediction by 74.24%, 28.60%, and 29.58% over ZeroM-GPR, LDM-GPR and GWR, respectively. By using the online radio map generated from PSFM-GPR, the localization accuracy of WinIPS is 1.718m on average, which improves the accuracy by 45.52% over ZeroM-GPR, 33.16% over LDM-GPR and 35.23% over GWR as well. Furthermore, STI-WKNN improves localization accuracy by 23.95% over traditional algorithm across heterogeneous MDs.

In summary, we make the following contributions:

- We develop a WiFi-based non-intrusive IPS, WinIPS, which is able to estimate locations of MDs without app-installation process on user's side.
- For online RSS data acquisition, we design WinSMS to overhear the WiFi traffic and extract RSS values and MAC address of MDs and APs from the data packets in non-intrusive manner. WinSMS can be directly implemented on COTS WiFi routers, and make these routers as natural reference points without introducing any extra hardware infrastructure.
- For online radio map construction and adaptation, we propose PSFM-GPR that is able to build up and update fine-grained radio map automatically over environmental dynamics and also discard the impractical laborious offline site survey process.
- We introduce STI-WKNN that allows WinIPS to provide high localization accuracy consistently across heterogeneous MDs.
- We prototype WinIPS and test it in real complex indoor environment. Promising results indicate that WinIPS makes a substantial progress towards fortifying WiFi fingerprint-based IPS for feasible large-scale commercialization.

The rest of the chapter is organized as follows. Section 7.1 introduces the detailed system design of WinIPS, as well as the methodologies of WinSMS, PSFM-GPR and

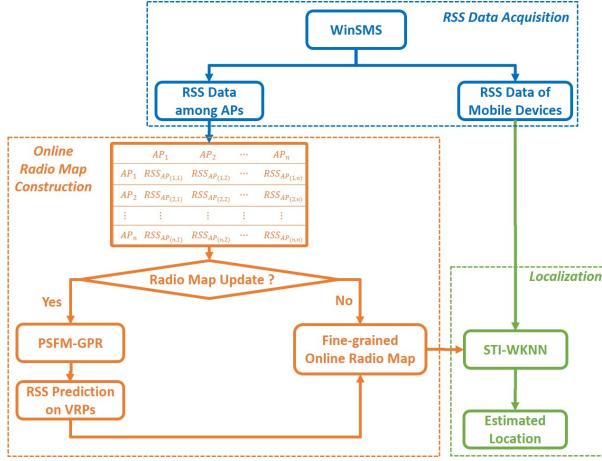


Figure 7.1: System architecture of WinIPS.

STI-WKNN. In Section 7.2, our experimental testbed and data collection procedure are elaborated at first, and then experimental results and performance evaluation of WinIPS are reported. We conclude this chapter in Section 7.3.

7.1 WiFi-based Non-Intrusive Indoor Positioning System (WinIPS)

7.1.1 System overview

We introduce, WiFi-based non-intrusive Indoor Positioning System (WinIPS), which enables automatic online radio map construction and adaptation for calibration-free indoor localization. The system architecture of WinIPS is illustrated in Fig. 7.1. It consists of three main parts: RSS data acquisition, online radio map construction and online localization. For RSS data acquisition, we develop, WiFi-based non-intrusive Sensing and Monitoring System (WinSMS), which enables COTS WiFi APs can overhear the 802.11 data packets transmitted in the existing WiFi traffic and extract RSS values in a non-intrusive manner without extra hardware infrastructure. All the data will be forwarded to a back-end server for radio map construction and localization. We propose Gaussian Process Regression with Polynomial Surface

Fitting Mean (PSFM-GPR), a reliable regression technique dedicated for RSS prediction on each VRP to construct and update a fine-grained online RSS radio map over various environmental dynamics. Moreover, signal tendency index - weighted K nearest neighbor (STI-WKNN) is adopted to estimate the locations of heterogeneous MDs with consistent high localization accuracy. The users can use any browser on their MDs to obtain the estimated location through WinIPS Web server without any troublesome App installation process. The following sections will introduce the methodologies of WinSMS, PSFM-GPR and STI-WKNN, respectively.

7.1.2 WinSMS for RSS data acquisition

Fingerprinting-based localization algorithms are widely used for existing WiFi-based IPSs since they are able to capture odd RSS distributions in complex indoor environments more accurately than model-based approaches. However, the RSS data acquisition process of traditional offline site survey is extremely tedious. A user is required to carry a MD to actively scan the RSS of nearby APs manually at numerous reference locations to obtain a fine-grained RSS radio map for online localization. The main drawbacks, the laborious offline site survey process and the vulnerability to environmental dynamics have been elaborated in Chapter 2. Unfortunately, the Apple Inc. has not provided any RSS API for third-party developers. Due to these reasons, active WiFi scanning via MD is not a practical method to establish radio map anymore. Therefore, it's urgent and indispensable to design a scheme for online RSS radio map construction and adaptation in an accurate, reliable, efficient, practical and non-intrusive manner.

We develop, WinSMS, an intelligent wireless system that enables COTS WiFi APs to overhear the data packets transmitted in the existing WiFi traffic in real-time without any intrusiveness on the user side. It is based on wireless distribution system (WDS) and can be implemented on most of the COTS WiFi routers that support OpenWrt [150] operating system. As a WDS, WinSMS can create a WiFi LAN and provide the basic Internet services for users in its wireless network coverage.

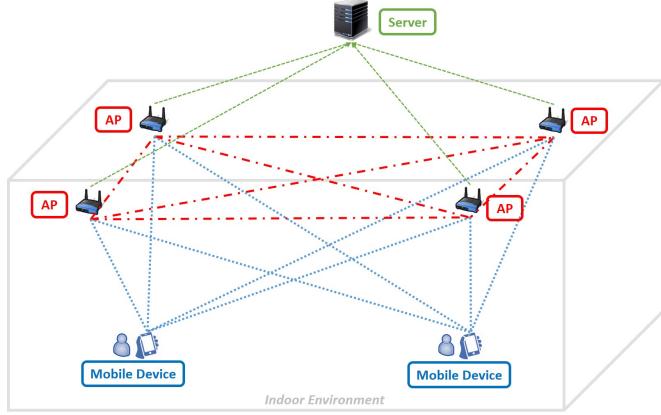


Figure 7.2: System architecture of WinSMS

More importantly, WinSMS has the ability to overhear the 802.11n data packets transmitted between each MD and WiFi routers, and precisely retrieve the RSS values and corresponding MAC addresses as identifiers. Then, all the information will be sent to a back-end server without requiring user to install any dedicated Apps for data acquisition.

Fig. 7.2 presents the system architecture of WinSMS. The main components of WinSMS includes COTS WiFi APs, a back-end server, as well as users and their MDs. All the APs in WinSMS perform the following major tasks: capture the 802.11n data packets in the network, extract relevant information from the packets, arrange them in a particular format and forward them to the back-end server. We upgrade the firmware of APs with OpenWrt and add a designed software based on Libpcap [151] to sniff the existing WiFi traffic, capture as well as analyze the data packets. Unlike traditional active RSS scanning via a MD which has a limited sampling rate, APs are able to overhear sustainable amount of data packets generated by various existing Apps on MDs, such as data stream from watching videos, push notification services and periodic email fetching, at the maximum rate around 100 packets per second in a non-intrusive manner. Furthermore, since WinSMS opportunistically captures the data packets from existing WiFi traffic, it poses no additional burden on the battery life of MD. Noticing that usually a person cannot move a significant distance in a second and the RSS value cannot change dramat-

Table 7.1: Online RSS observation among APs captured by WinSMS

	AP_1	AP_2	\dots	AP_n
AP_1	$RSS_{AP_{(1,1)}}$	$RSS_{AP_{(1,2)}}$	\dots	$RSS_{AP_{(1,n)}}$
AP_2	$RSS_{AP_{(2,1)}}$	$RSS_{AP_{(2,2)}}$	\dots	$RSS_{AP_{(2,n)}}$
\vdots	\vdots	\vdots	\vdots	\vdots
AP_n	$RSS_{AP_{(n,1)}}$	$RSS_{AP_{(n,2)}}$	\dots	$RSS_{AP_{(n,n)}}$

ically in such a short time, the RSS values received within 1 second are averaged out in the first place as a pre-filtering step. In this way, the RSS values collected by WinSMS are smoother than active scanning method. The weakest signal strength is set to be -95 dBm. If a particular data packet is received by only one AP, we set the value received by the others as -95 dBm which effectively means that the device is outside the range of that AP. After that, the retrieved RSS values of MDs with their corresponding MAC addresses will be sent to the back-end server through UDP protocol. The server is responsible to parse the data and build up the online RSS fingerprint database for localization.

For each AP, in addition to capturing the data packets sent and received by each MD, it can overhear packets of other APs as well. Therefore, the RSS measurements at these APs can be viewed as the signal strength of nearby MDs, and these measurements are also affected by the same environmental dynamics and change over time. As summarized in Table 7.1, all the APs can be used as natural online reference points for RSS radio map construction and adaptation since we have their physical coordinates and real-time RSS readings. Fig. 7.3 demonstrates the visualized pairwise RSS of 8 APs. In principle, each AP cannot sense the signal strength of itself. Therefore, we calibrate the average RSS of two APs placed side-by-side and assign -30 dBm as the self-sensed RSS to complete the pairwise RSS matrix of APs as shown in Fig. 7.3.

As shown in Fig. 7.4, the RSS values on the limited numbers of APs may not be good enough to describe a fine-grained RSS distribution of each AP. In order to obtain a more fine-grained radio map, we introduce VRPs and propose PSFM-GPR, a suitable RSS modeling scheme that is able to precisely estimate the RSS values of

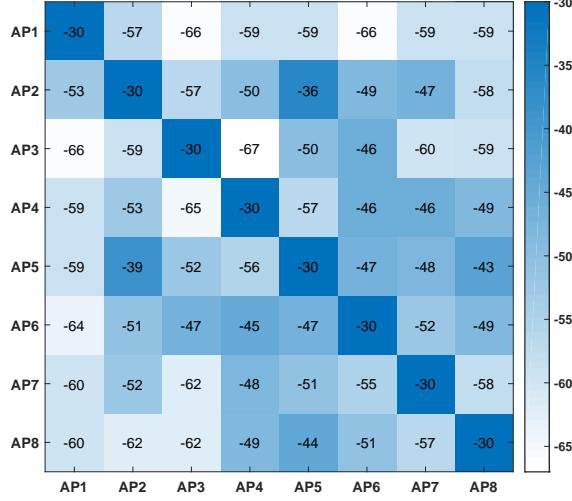


Figure 7.3: Visualization of pairwise RSS matrix of 8 APs (dBm)

each AP at predefined VRPs for fine-grained online radio map construction. The methodology of PSFM-GPR is introduced in the following section.

7.1.3 PSFM-GPR for online radio map construction and adaptation

7.1.3.1 Gaussian process regression model for RSS modeling

Admittedly, the RSS transmitted from a WiFi AP in a free space is a log linear delay function of the distance. Nevertheless, this property does not hold in practice due to the multipath effects caused by furniture, walls and moving occupants in complex indoor environments. Therefore, the ideal log-distance path loss model is not able to predict the RSS distribution precisely anymore. An efficient and powerful nonlinear approach is required to model the anomalous distribution of RSS values. As a nonparametric nonlinear regression approach, Gaussian process regression (GPR) is an appropriate method for capturing the noisy nature of RSS, and predicting RSS values for online dynamic radio map construction and adaptation. In fact, GPR has been widely employed in numerous areas, including geostatistics, spatial

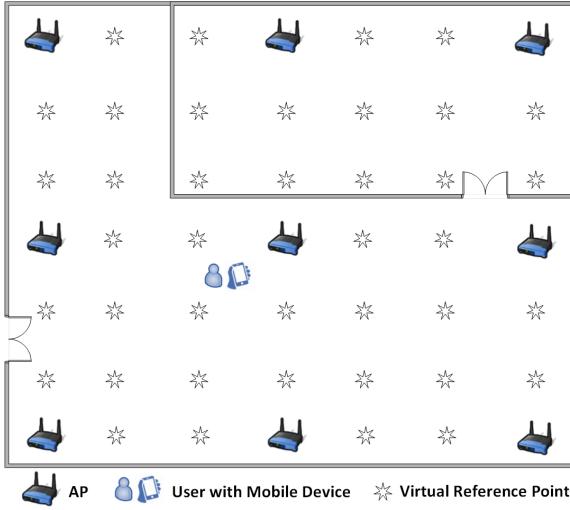


Figure 7.4: Scenario of online radio map construction in complex indoor environment.

smoothing, robotic applications and machine learning for probabilistic modeling, inference and prediction [152]. Moreover, previous works [153–155] employed GPR for RSS interpolation to reduce the number of reference points during the offline calibration phase.

A Gaussian process (GP) generates data located throughout a finite set of random variables \mathbf{Z} which follows a joint multivariate Gaussian distribution. It is characterized by its mean function $m(z) = \mathbb{E}[f(z)]$ and the covariance function $k(z, z') = \mathbb{E}[(f(z) - m(z))(f(z') - m(z'))]$, where $z \in \mathbf{Z}$ of a process $f(z)$. The marginalization property of GP [152] allows us to predict the posterior probability with an unknown input z^* according to some given inputs z and their corresponding observations.

Since the online radio map construction process of each AP is similar, we will introduce how to use GPR to predict RSS values of AP_i as an example, where AP_i is one of the total n APs. WinSMS enables each AP to scan not only the RSS of MDs, but also RSS of other APs. Therefore, all the APs are natural online reference points (training points) for radio map construction and adaptation. The corresponding dataset for each AP consist of pairs of $(\mathbf{l}_i, s_i)_{i=1}^n$, $\mathbf{l}_i \in \mathbf{L}, s_i \in \mathbf{S}$, where $\mathbf{l}_i = (x_i, y_i)$ is the two-dimensional coordinates of an AP, and s_i is an RSS value of the AP at

location \mathbf{l}_i . The relationship between the two-dimensional space \mathbf{L} and RSS \mathbf{S} can be modeled as a GP:

$$s_i = f(\mathbf{l}_i) + \epsilon_i$$

where ϵ_i is independent and identically distributed (i.i.d.) additive zero-mean Gaussian noise with variance σ_ϵ^2 . Assume the RSS observations at each AP can be drawn from the GP:

$$s \sim \mathcal{GP}(m(\mathbf{l}), k(\mathbf{l}, \mathbf{l}'))$$

where $m(\cdot)$ and $k(\cdot, \cdot)$ represent the mean and covariance function of GP respectively. GP learns the covariance of training dataset through the kernel covariance function. In our case, the input data are the two-dimensional coordinates. The value of the kernel covariance function is higher where two points are near to each other and lower where two points are far way. We utilize the most popular squared exponential kernel covariance function:

$$k(\mathbf{l}, \mathbf{l}') = \sigma_f^2 \exp\left[\frac{-\|\mathbf{l} - \mathbf{l}'\|^2}{2r^2}\right] + \sigma_\epsilon^2 \delta(\mathbf{l}, \mathbf{l}') \quad (7.1)$$

where σ_f^2 and r are the hyperparameters of GP and $\delta(\cdot, \cdot)$ stands for the Kronecker delta function. Since we have n APs in the space, we can calculate the covariance of each pair of APs according to Equation (7.1) and obtained the $n \times n$ covariance matrix $K(\mathbf{L}, \mathbf{L})$ for all pairs of training data. Suppose that we would like to predict RSS values $\{s_j\}_{j=1}^m \in \mathbf{S}^*$ of AP $_i$ at m VRPs $\{\mathbf{l}_j\}_{j=1}^m \in \mathbf{L}^*$ to build up a fine-grained radio map. The multivariate Gaussian distribution of training data and predicted RSSs with a zero mean distribution can be described as follows:

$$\begin{bmatrix} \mathbf{S} \\ \mathbf{S}^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\mathbf{L}, \mathbf{L}) + \sigma_\epsilon^2 \mathbf{I} & K(\mathbf{L}, \mathbf{L}^*) \\ K(\mathbf{L}^*, \mathbf{L}) & K(\mathbf{L}^*, \mathbf{L}^*) \end{bmatrix}\right)$$

where $K(\mathbf{L}, \mathbf{L}^*)$ is $n \times m$ covariance matrix between \mathbf{S} and \mathbf{S}^* , and \mathbf{I} is the iden-

tical matrix. The RSS value of this AP at an interested point \mathbf{l}_j can be predicted according to the posterior mean and variance of GP:

$$\bar{s}_j = K(\mathbf{l}_j, \mathbf{L})[K(\mathbf{L}, \mathbf{L}) + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{S}$$

$$cov(s_j) = K(\mathbf{l}_j, \mathbf{l}_j) - K(\mathbf{l}_j, \mathbf{L})[K(\mathbf{L}, \mathbf{L}) + \sigma_n^2 \mathbf{I}]^{-1} K(\mathbf{L}, \mathbf{l}_j)$$

where \bar{s}_j is the estimated mean RSS at this location, $cov(s_j)$ denotes the posterior variance as an estimation confidence indicator, and $K(\mathbf{l}_j, \mathbf{L})$ and $K(\mathbf{L}, \mathbf{l}_j)$ indicate $1 \times n$ and $n \times 1$ matrices of the covariance between this point and all training points.

As shown in Equation (7.2), GP model usually adopts the zero mean function (ZeroM-GPR) as the default settings, which means the estimated RSS values will tend to zero at locations that are far from any training points (APs) [155]. This is obviously impractical for RSS modeling. Previous works [114] used the Log-Distance path loss model to obtain a general mean of RSS and then make use of GPR to estimate the residual RSS errors. The estimated RSS at arbitrary location \mathbf{l}_j is calculated by

$$\bar{s}_j = m(\mathbf{l}_j) + K(\mathbf{l}_j, \mathbf{L})[K(\mathbf{L}, \mathbf{L}) + \sigma_\epsilon^2 I]^{-1} (\mathbf{S} - m(\mathbf{L}))$$

$$m(\mathbf{l}_j) = PL_0 + 10\alpha \log(\|\mathbf{l}_j - \mathbf{l}_{AP_i}\|/d_0)$$

where $\|\mathbf{l}_j - \mathbf{l}_{AP_i}\|$ indicates the distance between AP_i and location \mathbf{l}_j , PL_0 is the path loss coefficient as the RSS value at initial distance d_0 , and α is the path loss exponent. These three parameters of Log-Distance path loss model in Equation (7.3) are calculated by curve fitting with the training points. The Log-Distance mean GPR (LDM-GPR) can be used to estimate RSS distribution in open space because it describes the relationship between RSS and distance. However, in practice, as shown in Fig. 7.5, the RSS distribution are much more complicated. The RSS values at the same distance from the AP are usually distinct due to the irregular layout in complex indoor environment. Therefore, the LDM-GPR is no longer suitable since it does not consider the orientation or the surrounding environmental property on

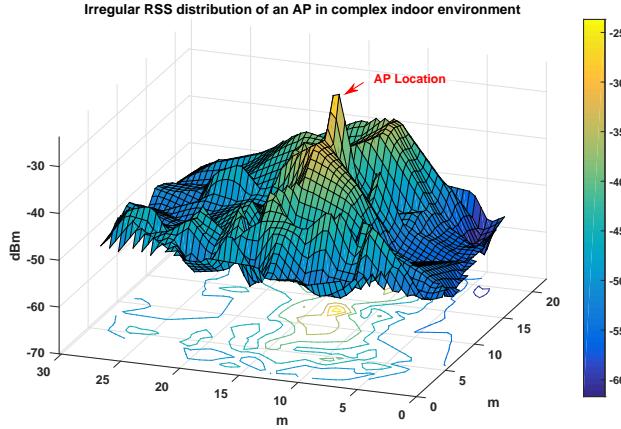


Figure 7.5: Irregular RSS distribution of an AP in complex indoor environment.

each VRP.

7.1.3.2 Online radio map construction with PSFM-GPR

In order to address this issue, we propose PSFM-GPR, which utilizes a two dimensional polynomial surface fitting model to estimate the general mean of RSS rather than the Log-Distance path loss model, and then utilizes GP to estimate the residual RSS errors. First of all, we assume the RSS distribution of AP_i to be a two dimensional polynomial function as follows:

$$m(\mathbf{l}) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \quad (7.4)$$

where $\mathbf{l} = (x, y)$ denotes the coordinates of other APs. Since the WinSMS can obtain the RSS values of AP_i at all other APs' locations, all the parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ in Equation (7.4) can be estimated and updated online using two degree polynomial surface fitting. According to our data analysis regarding the fitting accuracy and the computational overhead, we found that two degree polynomial surface fitting is good enough to capture the odd RSS distribution instead of using any higher degree polynomial functions. With this proper mean of RSS, the predicted RSS by

PSFM-GPR at any arbitrary location \mathbf{l}_j is calculated by

$$s_j = m(\mathbf{l}_j) + K(\mathbf{l}_j, \mathbf{L})[K(\mathbf{L}, \mathbf{L}) + \sigma_e^2 I]^{-1}(\mathbf{S} - m(\mathbf{L})) \quad (7.5)$$

$$m(\mathbf{l}_j) = \beta_0 + \beta_1 x_j + \beta_2 y_j + \beta_3 x_j^2 + \beta_4 y_j^2 + \beta_5 x_j y_j \quad (7.6)$$

where (x_j, y_j) is the coordinates of location \mathbf{l}_j .

After estimating the RSS values of all the n APs by PSFM-GPR at the m VRPs, we can obtain a RSS vector $\mathbf{s}_j = [s_j^1, s_j^2, \dots, s_j^n]$, where $1 \leq j \leq m$ and s_i^j ($1 \leq i \leq n$) denotes the RSS values from AP_i at each VRP \mathbf{l}_j . Therefore, a $m \times n$ RSS fingerprint database can be built up effectively online to avoid cumbersome offline site survey process.

7.1.3.3 Online radio map adaptation with PSFM-GPR

In addition to the radio map construction process introduced in the previous section, the radio map adaptation is another crucial process of WinIPS system because it keeps the radio map up-to-date automatically over various contextual dynamics including time and space. Since the WinIPS system can obtain the RSS values of all APs in real time as presented in Table 7.1, each column in $n \times n$ RSS matrix can be used as a trigger to determine whether the system should initiate the radio map adaptation process for each AP.

The detailed procedure of radio map adaptation is presented in Algorithm 7.1. First of all, we will compare the differences between the real-time RSS values s_i^t and the RSS profile s_i^{t-1} stored in the database for all the APs. If the RSS distance between these two RSS vectors $\|s_i^t - s_i^{t-1}\|$ is larger than a RSS threshold θ_{th} , it implies that the RSS profile of AP_i is outdated due to some indoor environmental dynamics, and the radio map update procedure will be initiated for this AP. According to our empirical study, we set the threshold θ_{th} to 10dBm. The RSS values from this AP at each VRP will be updated by the PSFM-GPR scheme as introduced in Section 7.1.3.2. In this way, the $m \times n$ online RSS fingerprint database will be up-to-date

Algorithm 7.1: Online radio map update algorithmInitialization:**Input:** n - The total number of APs m - The total number of VRPs \mathbf{s}^{t-1} - $n \times n$ RSS matrix of AP as shown in Table 1 s_i^{t-1} - The RSS vector of AP_i stored in the fingerprint database s_i^t - The RSS vector of AP_i at the time t θ_{th} - The RSS threshold for AP RSSI differences**Output:** \mathbf{s}_f^t - $m \times n$ Up-to-date RSS fingerprint database at time t Check RSS profile of each AP:**for** $i = 1, \dots, n$ **do** **if** $\|s_i^t - s_i^{t-1}\| > \theta_{th}$ **then** RSS profile of AP_i is required to update $AP_i \in AP_Q$ **else** RSS profile of AP_i is up-to-date **end if****end for**Update RSS fingerprint database:**for** $q = 1, \dots, Q$ **do** $AP_q \rightarrow PSFM - GPR$ to predict RSS on all VRPs **for** $j = 1, \dots, m$ **do** $s^j = s_{AP_q}^j$ **end for****end for****return** \mathbf{s}_f^t

as always and more robust to various contextual dynamics than traditional offline fingerprint database.

7.1.4 STI-WKNN Localization algorithm for heterogeneous mobile device

As introduced in the aforementioned sections, an up-to-date and fine-grained online RSS fingerprint database is obtained using WinSMS and PSFM-GPR. However, the RSS values stored in this database are based on APs. This database cannot be applied directly for localization of MDs because the RSS signatures of AP and MDs are usually different due to various heterogeneous factors, including distinct WiFi chipsets, WiFi antennas, hardware driver, and even operating systems [15, 94]. To illustrate this issue, we conduct an experiment that collected 500 RSS samples of a portable AP, as well as two different MDs: iPhone 6 and Nexus 6 at two identical locations with respect to 14 commodity WiFi APs in a complex indoor environment. As observed in Figure 7.6, each curve connects the average RSSs between one device and 14 APs. The RSSs associated with AP and MDs are significantly different, which verifies the effect of device heterogeneity. Therefore, the localization accuracy will be severely jeopardized if we employ RSS fingerprint database of AP to estimate the location of MD directly.

To handle the device heterogeneity issue, we leverage signal tendency index (STI) [20], which compares the similarities of the RSS curve shapes by using ordinary Procrustes Analysis (PA) method [134] instead of using the raw RSS for fingerprint matching. As shown in Figure 7.6, the shape of the curves displays certain similarities. Hence, one curve can be roughly recovered from another one via translation and scaling operation in ordinary PA method. To be specific, given a real-time RSS vector from a MD, s_d , the translation step of the ordinary PA method will produce

$$s_d^1 - \bar{s}_d, s_d^2 - \bar{s}_d, \dots, s_d^n - \bar{s}_d \quad (7.7)$$

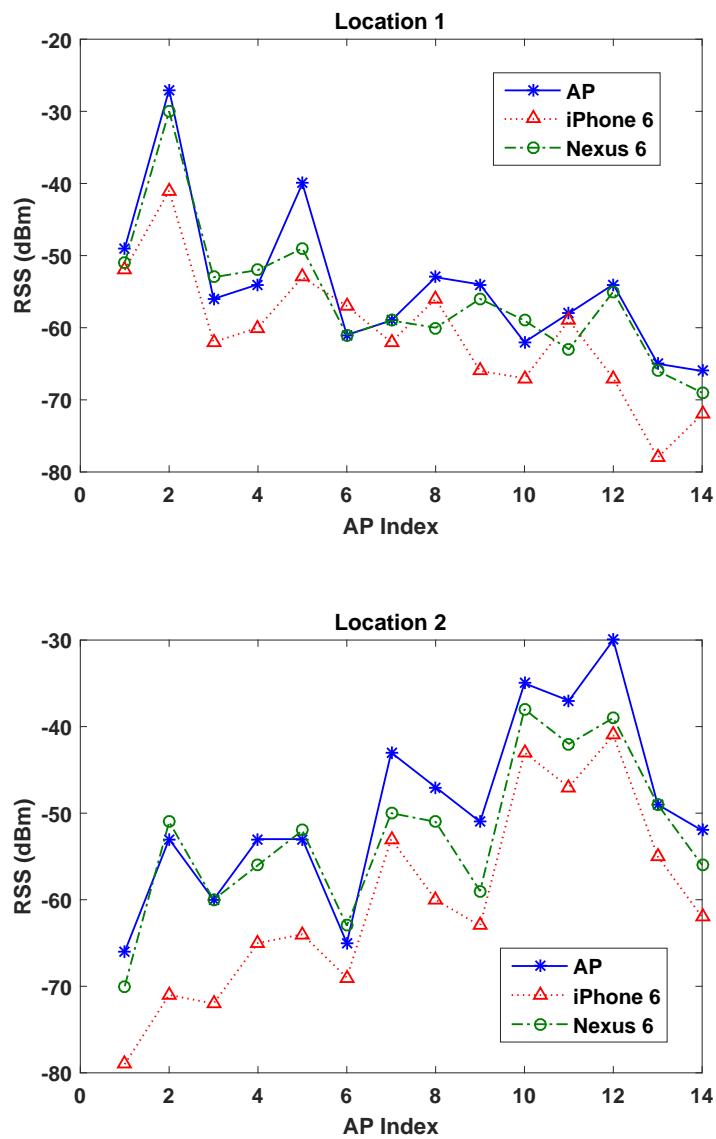


Figure 7.6: WiFi RSSs measured by different mobile devices at identical locations.

where

$$\bar{s}_d = \frac{1}{n} \sum_{i=1}^n s_d^i.$$

Then, in the uniform scaling step, we have

$$\hat{s}_d = [s_d^1 - \bar{s}_d, s_d^2 - \bar{s}_d, \dots, s_d^n - \bar{s}_d] / \hat{\sigma}, \quad (7.8)$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (s_d^i - \bar{s}_d)^2}.$$

The \hat{s}_d is the transformed object of ordinary PA method. Similarly, all the AP-based RSS vectors stored in the fingerprint database will be transformed as well. All the transformed RSS fingerprints $\{\hat{s}^j\}_{j=1}^m$ will be compared with \hat{s}_d in terms of their shape similarity. We define the Procrustes distance between the two vectors \hat{s}_d and \hat{s}^j , termed signal tendency index (STI), which is computed by

$$STI^j = \|\hat{s}_d - \hat{s}^j\| \quad (7.9)$$

where $\|\cdot\|$ denotes the Euclidean norm. After that, we introduce a new weighting scheme which involves STI and integrate it with the classical localization algorithm, weighted K nearest neighbor (WKNN), namely STI-WKNN, instead of using the distance of RSS vectors as the weights. Since we have calculated the STI value STI^j between s_d and each s^j , a smaller STI^j indicates that s^j is similar to s_d . We further define a weight value w^j for each s^j , which is calculated as follows:

$$w^j = \frac{\frac{1}{STI^j}}{\sum_{j=1}^m \frac{1}{STI^j}} \quad (7.10)$$

Then, the m VRPs are sorted according to their w^j in a descending order. Only top K VRPs and their corresponding physical coordinates are adopted to estimate the

location of MD (x_d, y_d) , which is calculated by:

$$(x_d, y_d) = \frac{1}{c} \sum_{k=1}^K (x_k, y_k) \cdot w^k \quad (7.11)$$

where (x_k, y_k) denotes the coordinates of i th VRP and $c = \sum_{k=1}^K w^k$ is the normalization constant.

In summary, the STI-WKNN localization scheme first compares the similarities of the RSS curve shapes between real-time RSS vector of a MD and those stored in the fingerprint database by ordinary PA method. Then, the similarity index STI is adopted as a novel weighting scheme for WKNN to estimate the location of heterogeneous MDs.

7.2 Experimental Results and Discussions

7.2.1 Experimental setup

The experiments were carried out in the IoT testbed as introduced in Section 4.3.1. The layout of the testbed at the beginning of the experiment (labeled as T_1) is depicted in Fig. 7.7 (a), while the layout after renovation six months later (labeled as T_2) is presented in Fig. 7.7 (b). As shown in Fig. 7.7, there are several obvious layout differences during the experiment which definitely affect the RSS distribution in the area. We leveraged these changes to verify the radio map adaptation and localization performance of WinIPS under environmental dynamics. Different from the traditional evaluation methods [113, 114] that usually adopt corridors or open spaces as testbeds, which are favorable for distance-related RSS modeling, as shown in Fig. 7.7, this complex indoor environment is much more suitable than an ideal environment for performance evaluation of WinIPS.

In our experiments, 10 COTS routers, TP-LINK TL-WR703N WiFi router, were adopted as APs of WinSMS in our experiment. TLWR703N has a 400 MHz Atheros



Figure 7.7: (a) Layout of the testbed at the beginning of the experiment (T_1). (b) Layout of the testbed six months later after renovation (T_2).

AR7240 CPU with 4 MB flash memory and 32 MB RAM. The Atheros AR9331 chipset is used in its platform working on 2.4 GHz. To implement WinSMS, we upgraded their firmware to OpenWrt and add our designed software. As shown in Fig. 7.7, TL-WR703N nano router is small size and extremely easy to be deployed. We choose this router to show that commercial routers are becoming portable and easy for installation nowadays. Moreover, with the booming development of Internet of Things (IoT), billions of IoT devices will be densely deployed in indoor environments for multiple purposes in the near future. Equipped with WiFi modules, they can be easily upgraded to serve as online reference points for dynamic radio map construction and adaptation. The locations of these 10 APs are depicted in Fig. 7.7 and they were fixed on 1.9-meter-high tripods to keep them on the same height level. One server is employed to process the RSS data sent by APs, construct and update the RSS radio map and fingerprint database by PSFM-GPR, and adopts STI-WKNN to estimate the location of each MD.

50 testing points (small red circles in Fig. 7.7) were randomly selected to evaluate the performance of WinIPS. To validate the RSS estimation accuracy of PSFM-GPR, we collected the real RSS values of a TL-WR703N router at these points as the ground truth. Furthermore, we also collected the RSS measurements of two MDs: Nexus 6 and iPhone 6, at all the testing points to evaluate the localization accuracy of STI-WKNN across heterogeneous devices.

7.2.2 RSS estimation accuracy

Firstly, we conducted an experiment to continuously monitor the distribution of RSS variations of an AP (AP10) to understand the fluctuations of RSS caused by various environmental dynamics in six months. Fig. 7.8 demonstrates the distribution of RSS variation of the AP through one week and six months respectively. As shown in Fig. 7.8, the RSS variation in six months (long-term) is much larger than in one week (short-term). Thus, it indicates that the static radio map calibrated at particular time is definitely unable to serve as the reference for consistent location estimation

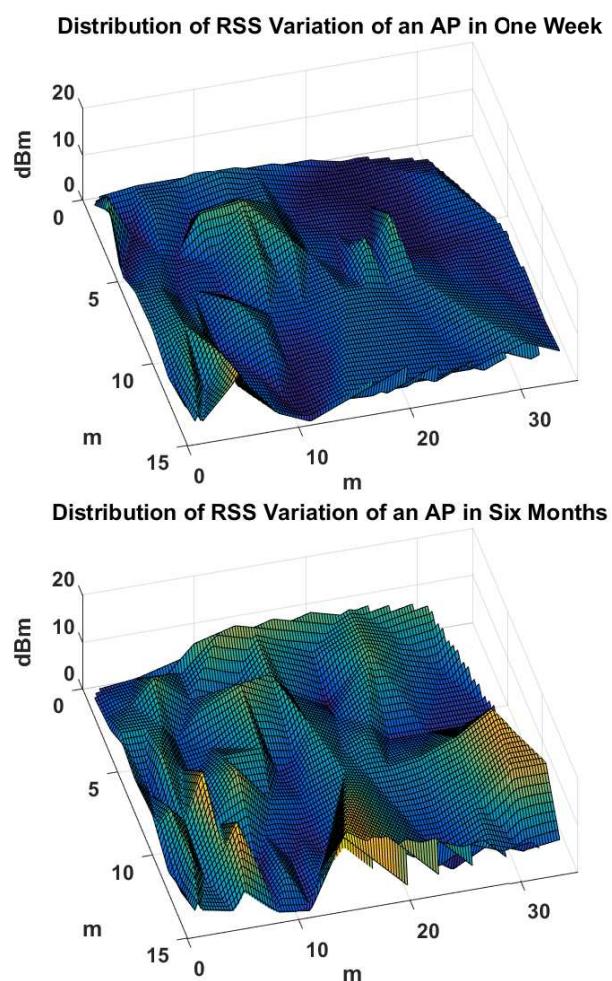


Figure 7.8: Comparison of RSS variation distribution between short-term and long-term time period.

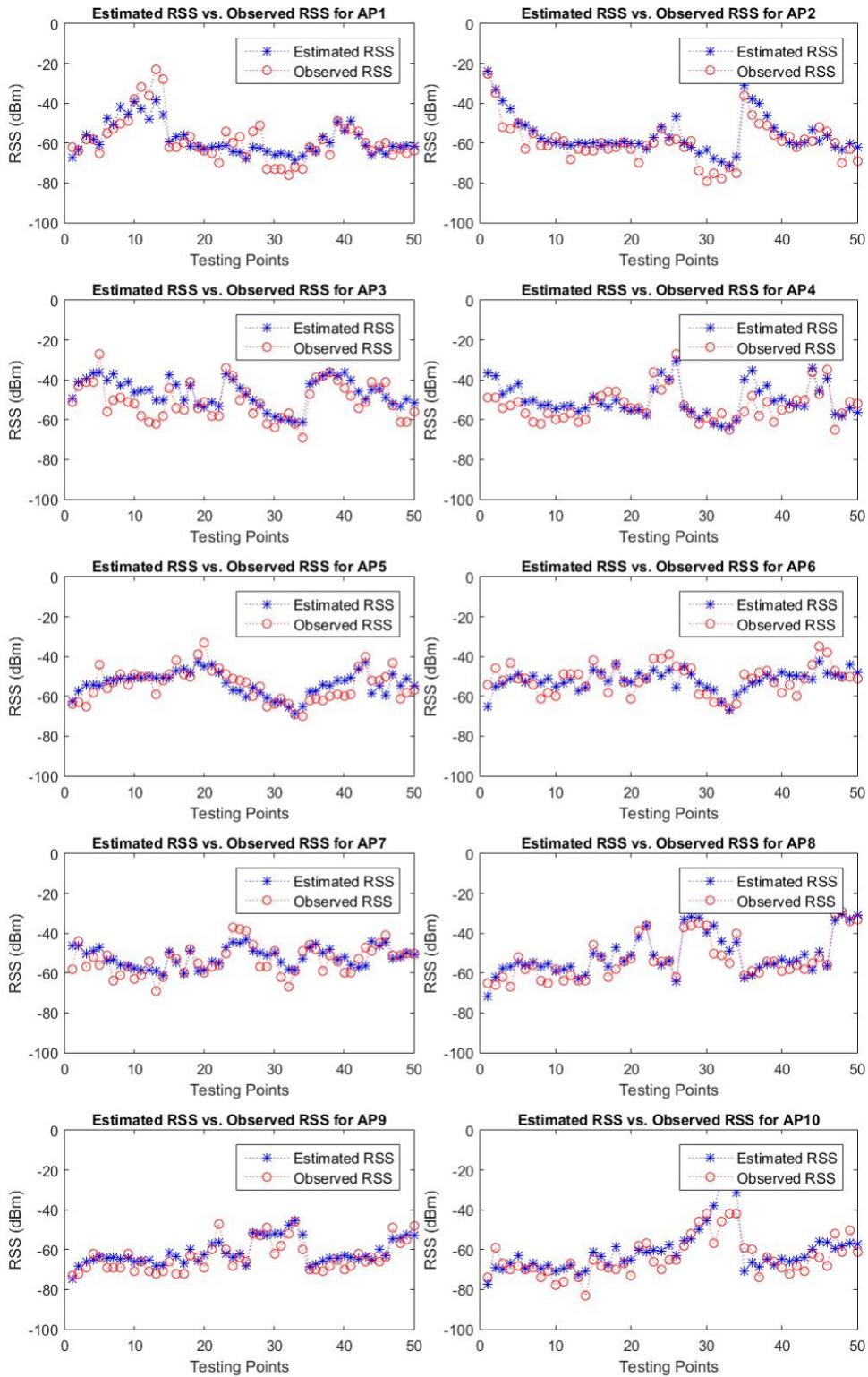


Figure 7.9: RSS estimation accuracy using PSFM-GPR.

Table 7.2: Comparison of RSS estimation accuracy between different methods (dB-m)

	Mean	Mean Improve (%)	σ	σ Improve (%)
ZeroM-GPR	18.16	74.24	6.90	49.16
LDM-GPR	6.55	28.60	5.11	31.29
GWR	6.64	29.58	9.81	64.23
PSFM-GPR	4.68	3.51		

at all times, since the real time RSS values can vary significantly, especially in long-term deployments. Radio map adaptation strategy such as WinSMS is urgently desired to make the IPS resilient to environmental dynamics.

For WinSMS, the real-time RSS measurements among the 10 APs collected by it can be summarized as a 10×10 RSS matrix, which is similar to Table 7.1. By employing these data, we predicted RSS values from APs using PSFM-GPR at the 50 testing points and compared with the observed RSS (ground truth). The comparison results of all 10 APs are demonstrated in Fig. 7.9, which displays consistent matching between RSS predicted by PSFM-GPR and observed RSS in general. We further compared the RSS estimation of PSFM-GPR with ZeroM-GPR, LDM-GPR [114] and GWR [113] in terms of mean (\bar{e}_{RSS}) and standard deviation (σ_{RSS}) of the RSS estimation error as shown in Table 7.2. The average estimated RSS error of PSFM-GPR is 4.68 dBm which is the smallest among the four methods. It is able to reduce the average RSS error by 74.24%, 28.60%, 29.58% compared to ZeroM-GPR, LDM-GPR and GWR respectively. Moreover, the standard deviation of RSS error of PSFM-GPR is also the smallest among the four methods, indicating that RSS predicted by PSFM-GPR is more stable than existing approaches.

Furthermore, we evaluated the RSS estimation accuracy of PSFM-GPR in two-dimensional space. To illustrate, Fig. 7.10 describes the estimated RSS error distribution of AP8 from four different RSS modeling methods. As illustrated in Fig. 7.10(d), most of RSS errors of PSFM-GPR are smaller than 10 dBm and distributed evenly in a low RSS error level. The reason for such an outstanding

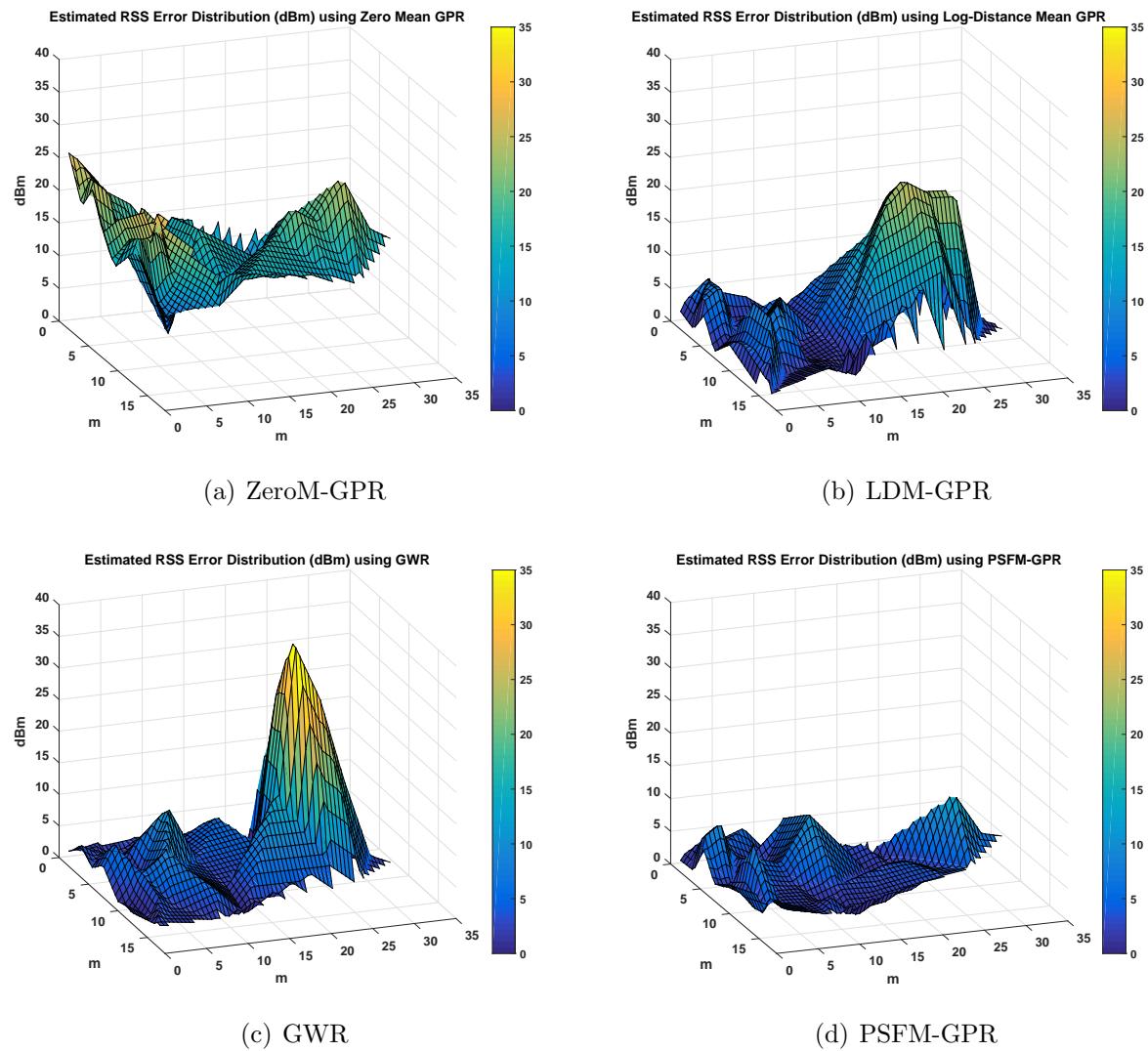


Figure 7.10: Comparison of estimated RSS error distribution of AP8 using different methods.

Table 7.3: Comparison of localization accuracy (m) between online RSS prediction methods

	\bar{e}_{LA} (m)	Improve (%)	σ_{LA} (m)	Improve (%)
ZeroM-GPR	3.153	45.52	2.514	68.06
LDM-GPR	2.570	33.16	1.405	42.87
GWR	2.652	35.23	1.390	42.25
PSFM-GPR	1.718	0.803		

performance is that PSFM-GPR performs two dimensional surface fitting for RSS prediction which precisely captures odd RSS distribution in different orientation. In contrast, the RSS errors of ZeroM-GPR is highest especially at the locations far away from any AP (online reference points). For LDM-GPR and GWR, they failed to capture the intangible RSS distribution in complex indoor environments because only the relationship between RSS and distance in RSS modeling is considered.

7.2.3 Localization estimation Accuracy

The aforementioned section illustrates the RSS estimation evaluation of WinIPS. We present the localization accuracy evaluation of WinIPS in this section. To have a fine-grained online RSS fingerprint database, the back-end server virtually divided our testbed into a 1.5×1.5 m grid and adopted PSFM-GPR to predict RSS values from all APs on the 288 grid points (VRPs). The grid spacing between two adjacent VRPs was chosen to be 1.5m according to the analysis in [140]. For the evaluation purpose, we collected 500 RSS samples of iPhone 6 at each testing point, and use the average location estimation by STI-WKNN to compare with physical location of each testing point (ground truth).

7.2.3.1 Comparison between different online RSS prediction methods

First of all, we evaluate the impacts of different online RSS prediction methods on localization accuracy. In the back-end server, we established three online RSS fingerprint databases using ZeroM-GPR, LDM-GPR and GWR, respectively in a

Table 7.4: Comparison of localization accuracy using offline site survey and PSFM-GPR (m)

Method	\bar{e}_{LA} (m)	σ_{LA} (m)
Offline Site Survey (T_1)	2.327	1.471
Offline Site Survey (T_2)	1.569	1.097
PSFM-GPR (T_2)	1.718	0.803

similar process as PSFM-GPR. STI-WKNN was used as the localization algorithm for all the schemes in this evaluation to make a fair comparison. The statistical attributes (i.e.mean (\bar{e}_{LA}) and standard deviation (σ_{LA}) of localization accuracy via PSFM-GPR is compared with other three existing approaches. The overall performance is summarized in Table 7.3 and Fig. 7.12. It is evident from Table 7.3 that the localization accuracy of WinIPS is much higher when PSFM-GPR is adopted for RSS prediction on VRPs. Fig. 7.12 depicts the distance error distribution of the four approaches. Similar to the results shown in Table 7.3, PSFM-GPR has the best performance among the four approaches. PSFM-GPR + STI-WKNN can provide 1.718 m average localization accuracy with the smallest $\sigma_{LA} = 0.803$. It enhances the precision of indoor positioning by 45.52% over ZeroM-GPR, 33.16% over LDM-GPR and 35.23% over GWR respectively. Furthermore, the smallest σ_{LA} indicates that online RSS fingerprint database generated by PSFM-GPR can provide more useful information for reliable localization service than the other approaches.

We also explored whether there is any potential correlation between RSS estimation accuracy and localization accuracy using PSFM-GPR. Fig. 7.11 demonstrates the RSS estimation error vs. the localization error when different number of APs are utilized. It can be seen from Fig. 7.11, smaller RSS estimation errors lead to smaller localization errors. Then, we can infer that there is a positive correlation between RSS estimation accuracy and localization accuracy for PSFM-GPR. Another noteworthy point is that the results in Table 7.3 are comparable to those reported in [66, 153] which rely on cumbersome offline calibrated RSS fingerprint database.

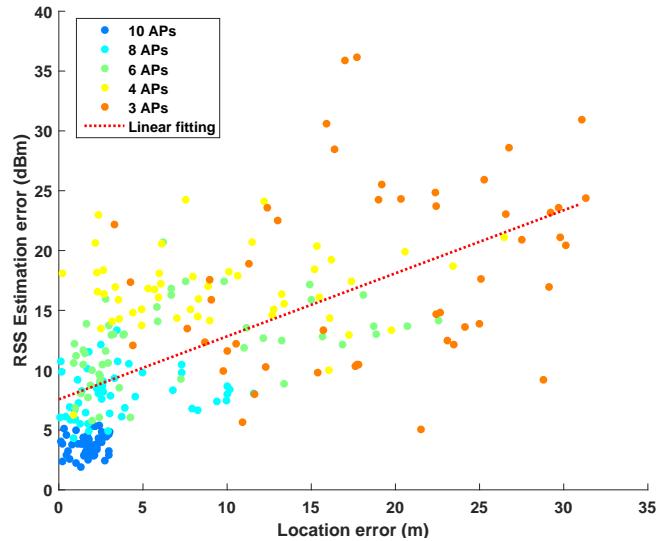


Figure 7.11: Plots of the RSS estimation error vs. the localization error, color coded by the number of APs in use.

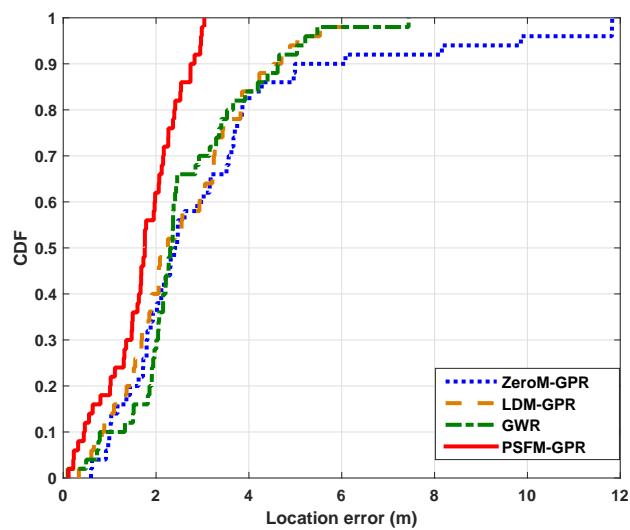


Figure 7.12: Comparison of localization accuracy between different online RSS Prediction Methods.

7.2.3.2 Comparison between traditional offline site survey and PSFM-GPR

In this section, we compare the localization performance of PSFM-GPR with traditional offline site survey method. We collected real RSS measurements of the MD on the physical coordinates of each VRP to build up the offline RSS fingerprint database. Although WinSMS is able to collect RSS values at a fast speed (0.5 second/sample), we still spent 5 hours to complete the offline site survey process which is truly time-consuming and labor-intensive. We performed two offline site survey and constructed the corresponding fingerprint databases at the beginning of the experiment (T_1) and six months later (T_2). The testing data was collected on T_2 . The overall performance is presented in Table 7.4. When up-to-date online RSS fingerprint database generated by PSFM-GPR is compared with offline database constructed at the same day (T_2), the average localization accuracy of it is only a little worse by 8.67% than offline calibrated RSS fingerprint database. However, it is impractical to build up an offline radio map every day for localization purposes. To illustrate the vulnerability of transitional offline site survey method to environmental dynamics, we compared the performance of out-of-date offline RSS fingerprint database (T_1) with PSFM-GPR as well. Under this situation, PSFM-GPR reduces the localization error by 26.17% than outdated offline RSS fingerprint database. In summary, PSFM-GPR can construct and update RSS fingerprint database automatically that enables WinIPS to provide consistent high localization accuracy over various environmental dynamics. Furthermore, it avoids the cumbersome offline site survey process which is the major bottleneck for large-scale commercialization of WiFi-based IPS.

7.2.3.3 Impact of device heterogeneity

To validate the effectiveness of WinIPS under the impact of device heterogeneity, we collected RSS measurements of one more device, Nexus 6, at the 50 testing points for this evaluation. The overall results are summarized in Table 7.5. As

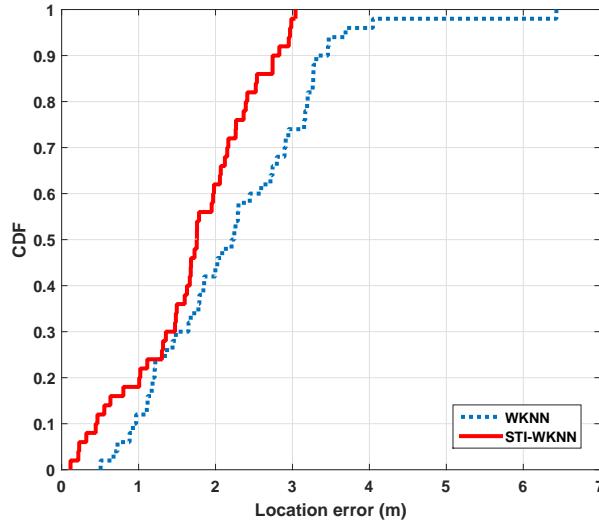


Figure 7.13: Comparision of cumulative distribution of location error between WKN-N and STI-WKNN.

Table 7.5: Comparison of localization accuracy among different mobile devices (m)

Mobile device	\bar{e}_{LA} (m)	σ_{LA} (m)
iPhone 6	1.718	0.803
Nexus 6	1.963	0.912

observed from Table 7.5, WinIPS can provide high localization accuracy (within 2m on average) consistently across heterogeneous MDs using STI-WKNN. Although the online RSS fingerprint database established by PSFM-GPR is based on data among APs, the device heterogeneity issue can be largely alleviated by comparing the similarities of RSS curve shapes (STI) rather than raw RSS values for WKNN fingerprint matching. Fig. 7.13 depicts the distance error distribution of original WKNN and STI-WKNN. STI-WKNN has much better performance in terms of localization accuracy than original WKNN. It improves localization accuracy by 23.95% over original WKNN across heterogeneous MDs. In summary, the merit of STI-WKNN enhances the robustness of WinIPS to device heterogeneity issue for indoor localization.

7.3 Conclusion

In this chapter, we proposed, WinIPS, a WiFi-based non-intrusive indoor positioning system that enables automatic online radio map construction and adaptation for calibration-free indoor localization to overcome the major drawbacks of WiFi fingerprinting-based IPS. For RSS data acquisition, we developed WinSMS, a novel wireless system that can capture data packets transmitted in the existing WiFi traffic and extract the RSS and MAC addresses of both APs and mobile devices in a non-intrusive manner without introducing any extra hardware infrastructure. Since we can obtain the real-time RSS measurements of APs, the APs are becoming natural online reference points for online radio map construction and adaptation. Therefore, we can completely get rid of the tedious offline site survey process. Furthermore, in order to construct a more fine-grained radio map, we further proposed PSFM-GPR, a reliable regression technique dedicated to predict RSS on VRPs which can precisely capture the irregular RSS distribution over complex indoor environment. The online radio map is more adaptive and robust with respect to environmental dynamics than traditional offline calibrated RSS database since it keeps updated with new measurements. The online radio map is generated based on AP generated RSS values so it's not appropriate to employ directly for localization of mobile devices due to the device heterogeneity issue. We introduced STI-WKNN, which compares the shapes of RSS vectors between RSS readings of mobile device and online RSS fingerprint database rather than raw RSS values, to improve the localization accuracy of WinIPS across heterogeneous devices. Extensive experiments have been carried out to validate the effectiveness of WinIPS in a real-world multi-functional office. The experimental results show that PSFM-GPR achieves 4.8 dBm average RSS estimation error, which enhances the accuracy of online RSS prediction by 74.24%, 28.60%, 29.58% compared to ZeroM-GPR, LDM-GPR and GWR respectively. By leveraging the online radio map generated from PSFM-GPR, the localization accuracy of WinIPS is 1.718m on average, which improves the accuracy by 45.52% over ZeroM-GPR, 33.16% over LDM-GPR and 35.23% GWR as well.

Furthermore, STI-WKNN improves the localization accuracy by 23.95% over the original WKNN across heterogeneous mobile devices.

In summary, WinIPS overcomes the bottlenecks of current WiFi-based IPS, making it promising for large-scale practical implementation. One potential improvement for WinIPS is using more reliable and fine-grained localization signature such as Channel State Information (CSI) for radio map construction and adaptation because RSS is noisy, unstable and coarse.

Chapter 8

Seamless Integration of IPS and GPS

In Chapters 4–7, we have proposed systematic solutions to overcome the longstanding challenges for WiFi based IPS, including the device heterogeneity issue, the vulnerability to environmental dynamics, the large computational burden and the intensive costs of manpower and time for offline site survey. By solving these problems, our WiFi based IPS is able to outperform other state-of-the-art IPSs in terms of localization accuracy, reliability and efficiency. However, another noteworthy issue is how to integrate our IPS with GPS to provide seamless LBS while reducing the power consumption on the user side.

GPS is widely used and delivers outstanding outdoor LBS. Extensive research and studies have been conducted in IPS, and some feasible solutions have been proposed. However, the majority of existing research adopts the common assumption that the operating environment is either indoor or outdoor and is known, which does not necessarily hold in reality. Some areas adjacent to buildings (covered corridor, connections between buildings) or semi-open buildings (parking garage) have partial characteristics of both indoor and outdoor environments. In this case, a sole dependence on GPS or IPS is unable to deliver a precise indoor-outdoor (IO) detection. In the meanwhile, an accurate and effective IO detection scheme provides basic,

but critical information for upper-layer applications to serve individual users, e.g., mobile applications leveraging reliable IO status to give better services and alleviate battery consumption. Nonetheless, the research towards IO detection itself is severely lacking.

In this chapter, we propose BlueDetect as an accurate, fast response and energy-efficient scheme for IO detection and seamless LBS running on a mobile device based on the emerging low-power iBeacon technology. iBeacon is an advanced Bluetooth protocol proposed by Apple [61], which is built upon Bluetooth Low Energy (BLE). In our system, only a few small-sized, low-cost and battery-powered BLE beacons are required by BlueDetect, which are placed at landmarks, such as the boundary of covered corridors and entrances/exits of buildings, with a sparse density in intermediate regions between indoor and outdoor environments (classified as a semi-outdoor environment). The GPS module is turned on for LBS only in an outdoor environment. When it comes from outdoors to semi-outdoors, the decrease of the mean GPS signals is utilized as a trigger to turn off GPS and turn on Bluetooth, and the iBeacon mode of BlueDetect is responsible for providing LBS within semi-outdoor environments. Transitions between semi-outdoor and indoor environments are achieved seamlessly by comparing the signals of two BLE beacons placed on both sides of the entrance of the building.

We implement and evaluate BlueDetect on the Android platform using different mobile devices. We test it in a university campus where various environment scenes are included. The experimental results show that BlueDetect provides a higher context detection accuracy than GPS-IO and IODetector. Furthermore, it also delivers accurate and reliable positioning service in semi-outdoor environments. The power consumption of BlueDetect is evaluated as well. Owing to the energy-efficient property of iBeacon, BlueDetect consumes the minimum power among the three approaches. A video demo is presented to demonstrate the seamless navigation service using BlueDetect [156].

The rest of the chapter is organized as follows. We present the system design of BlueDetect in Section 8.1. A comprehensive evaluation of BlueDetect in terms of

IO detection accuracy, localization accuracy and power consumption is reported in Section 8.2. We conclude this chapter in Section 8.3.

8.1 System Design

In this section, we present an overview of the BlueDetect system firstly. After that, we introduce the design details of transition methodologies between outdoors/semi-outdoors/indoors and the iBeacon mode of BlueDetect for semi-outdoor environments, respectively.

8.1.1 System overview

Figure 8.1 illustrates the representative scenes of outdoor, semi-outdoor and indoor environments, as well as the localization techniques employed by BlueDetect in these three environments. For pure outdoor environments, only the GPS module will be turned on when users query their location information. When it comes from outdoor to semi-outdoor, the decrease of mean GPS signals is utilized as a trigger to turn on Bluetooth. Once the semi-outdoor state is confirmed, GPS is turned off for energy saving, and then, the iBeacon mode of BlueDetect will provide localization service in semi-outdoor environments. Only a few of the BLE beacons are required to be deployed to facilitate context detection and LBS in this area. In the transition from semi-outdoors to outdoors, the signals of BLE beacons deployed at the boundary of semi-outdoor environments will be detected. GPS will be turned on, and Bluetooth will be switched off for energy saving when no beacon signal can be detected by users' mobile devices. The transitions between semi-outdoor and indoor are achieved seamlessly and easily by deploying two BLE beacons at two sides of each entrance of indoor environments as landmarks. Whether the user is going into or out from an indoor environment can be inferred by comparing the RSSs of the two beacons. Since WiFi based IPS [14, 20] is the most practical system for indoor localization, which reuses existing WiFi infrastructures and can be applied



Figure 8.1: Representative scenes and corresponding localization technologies of three different environments.

directly for most mobile devices, we leverage it to provide indoor LBS when the indoor status is confirmed. The detailed transition methodologies between outdoors and semi-outdoors, semi-outdoors and indoors, and the localization algorithm of BlueDetect in the semi-outdoor environment are introduced in the following sections.

8.1.2 Seamless transition between outdoors and semi-outdoors

For pure outdoor environments, GPS is capable of providing sufficient positioning accuracy. When it comes to semi-outdoor or indoor environments, the number of visible satellites is decreased, and a significant decrease of the GPS signals is expected due to the block of the line-of-sight between the satellites and mobile device. The performance of GPS is jeopardized dramatically while draining the battery at a high power rate under this circumstance. Therefore, BlueDetect will turn off the GPS module in both semi-outdoor and indoor environments.

We conducted experiments to analyze the variation of GPS signals when a client with a mobile device was moving from outdoor to indoor environments (three experiments were conducted in a covered corridor, a connection between buildings and a semi-open parking garage). Figure 8.2 illustrates the maximum, mean and minimum of GPS SNR readings from visible satellites in outdoor, semi-outdoor and indoor environments. As shown in Figure 8.2, the value of mean SNR dropped more than 20% when coming from the outdoor to the semi-outdoor environment. Clearly,

it is more suitable than the other two values to be used as a trigger indicating the switching of environments. Algorithm 8.1 shows the transition methodology between outdoors and semi-outdoors of BlueDetect.

Algorithm 8.1: Bluedetect IO detection and localization algorithm (outdoor \Leftrightarrow semi-outdoor)

Input:

B - Bluetooth signal (iBeacon), G - GPS signal, σ_s - switching threshold, σ_t - duration threshold

Output:

Location of the mobile device

Outdoor \Rightarrow Semi-outdoor

if $G_{current} < \sigma_s * G_{previous}$ for σ_t **then**

 Turn on Bluetooth;

if No less than 2 beacons' $B > B_{min}$ **then**

 Turn off GPS;

 Utilize B for localization;

Environment Type \leftarrow Semi-outdoor

else

 Turn off Bluetooth;

 Utilize G for localization;

Environment Type \leftarrow Outdoor

end if

end if

Semi-outdoor \Rightarrow Outdoor

if $B \in \mathbf{B}_{boundary}$ **then**

 Turn off Bluetooth, Turn on GPS;

 Utilize G for localization;

Environment Type \leftarrow Outdoor

else

 Utilize B for localization;

Environment Type \leftarrow Semi-outdoor

end if

If the decline of mean GPS SNRs is detected and the value of the decrease is larger than the switching threshold $\sigma_s = 80\%$ for three consecutive samples (sampling rate: 645 ms/sample), the Bluetooth module will be turned on searching for BLE beacons. When fewer than two beacons are detected and their RSSs are larger than $B_{min} = -70$ dBm, it is confirmed that the mobile device is in the semi-outdoor environment, followed by turning off the GPS to save energy and activating the iBeacon mode of

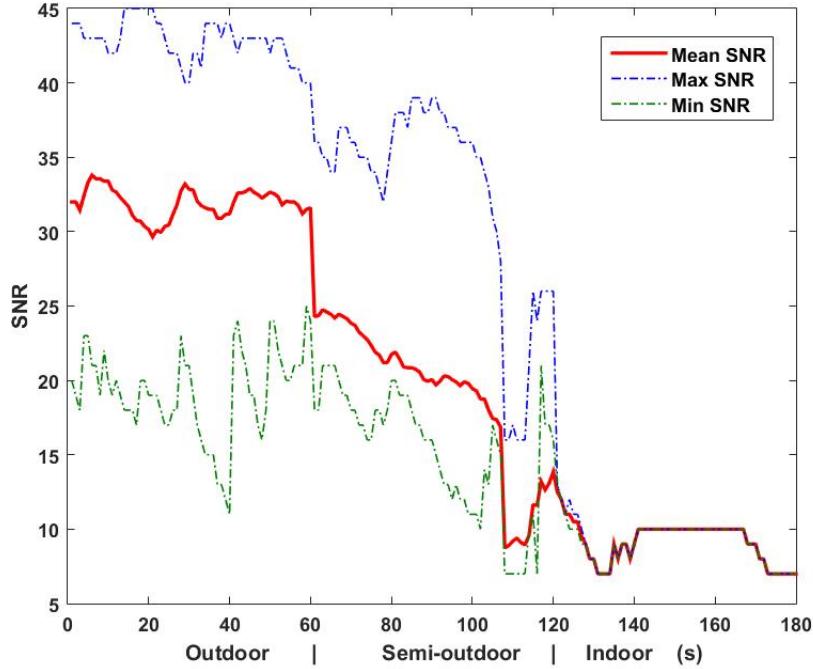


Figure 8.2: SNR of GPS signals in different environments.

BlueDetect for positioning. Otherwise, Bluetooth will be turned off, and GPS will still be used for localization.

Conversely, when the RSSs of BLE beacons installed at the boundary of the semi-outdoor environment are detected ($B \in \mathbf{B}_{boundary}$), the intention of going outdoors can be inferred, and then, the GPS module will be turned on for outdoor LBS in advance for a seamless transition. If no BLE beacon signal is detected for a period of longer than a threshold σ_t , which is four consecutive samples (sampling rate: 645 ms/sample) and a relatively stable GPS signal is maintained, the Bluetooth will be switched off.

8.1.3 Seamless transition between semi-outdoors and indoors

Algorithm 8.2 shows the transition methodologies between semi-outdoors and indoors of BlueDetect. In order to achieve seamless transition between semi-outdoors and indoors, one pair of BLE beacons is deployed at two sides of each door. For

each pair, the one placed outside the door is employed as a landmark of the entrances/exits, and the other one is installed 3–5 m away from the door inside the indoor environment. When the client’s mobile device detects the signals of these two beacons, the intention of going into or out of indoor environments can be determined by comparing the RSSs of the two beacons. To be specific, if $B_{indoor} > B_{semi-outdoor}$ maintains for σ_t , which is four consecutive samples (sampling rate: 645 ms/sample), it indicates that the user is entering into an indoor environment with high confidence. Therefore, BlueDetect will turn on the WiFi module of the mobile device to perform WiFi AP scanning for an Internet connection through the WiFi network and leverage WiFi based IPS for indoor positioning. Conversely, when $B_{indoor} < B_{semi-outdoor}$ for σ_t , we confirm that the client is going out of the indoor to the semi-outdoor environment, and the WiFi module will be turned off, which the iBeacon mode of BlueDetect is activated for localization.

Algorithm 8.2: Bluedetect IO detection and localization algorithm (semi-outdoor \rightleftharpoons indoor)

```

input:  $B$  - Bluetooth signal (iBeacon),  $W$  - WiFi signal,  $\sigma_t$ - duration threshold
output: Location of the mobile device
if  $B_{indoor} < B_{semi-outdoor}$  for  $\sigma_t$  then
    Turn off WiFi;
    Utilize  $B$  for localization;
     $Environment\ Type \leftarrow Semi-outdoor$ 
else
    Utilize  $W$  for localization;
     $Environment\ Type \leftarrow Indoor$ 
end if
```

8.1.4 Semi-outdoors (iBeacon)

Reliable LBS in a semi-outdoor environment is not readily available since neither GPS nor IPS can perform satisfactorily in this scenario. We employ the emerging iBeacon technology to fill in this gap. iBeacons make use of BLE proximity sensing to broadcast their unique identifiers to nearby portable mobile devices and trigger a location-based action on these devices. Since the iBeacon protocol uses very

short duration messages and does not need a paired connection with mobile devices (broadcast only), it is much more power efficient than classical Bluetooth protocols and less power hungry on the user-side than GPS and WiFi [59]. With such a merit, a BLE beacon can run on a coin cell battery for months or even for years. According to a recent study on the battery life of 16 major iBeacon hardware devices [62], by setting the advertising interval as 645 ms, an iBeacon with a CR2450 620-mAh coin cell battery is able to provide 11.2 months of life, which increases to two years as the advertising interval is increased to 900 ms. Nowadays, the iBeacon protocol is becoming a built-in standard for mobile devices, and a high density deployment of BLE beacons in buildings for multiple purposes will be expected in the near future.

For the iBeacon mode of BlueDetect, only a few of the portable, low-cost and battery-powered BLE beacons are deployed as landmarks in semi-outdoor environments for context detection, as well as localization and navigation. From the energy saving perspective, both GPS and WiFi modules of mobile devices are turned off when the semi-outdoor status is confirmed, since they supply no valuable information for IO detection and LBS under this circumstance. The common geographical structure of a semi-outdoor environment such as a corridor or a sky bridge is elongated; thus, a sparse deployment of BLE beacons is adequate to cover the entire area. According to the iBeacon protocol, the unique identifying information of each beacon is proximity universally unique identifier (UUID), major value and minor value. These parameters can be used to identify the building, the floor and the exact location of each BLE beacon. With the RSSs and locations of these beacons, we leverage weighted path loss (WPL) [17], a log-distance path loss model-based localization algorithm, to estimate the real-time location of a client's mobile device in semi-outdoor environments. The methodology of WPL is described as follows: Suppose a client's mobile device receives RSS from n BLE beacons simultaneously. The RSS of i -th iBeacon B_i can be expressed as:

$$B_i = P_0 - 10\alpha \log(d_i) + X_\sigma, 1 \leq i \leq n \quad (8.1)$$

where the reference path loss coefficient P_0 and the path loss exponent α need to be calibrated, and X_σ represents a zero Gaussian random noise with standard deviation σ . Then, based on Equation 8.1, the distance d_i between the mobile device and the i -th iBeacon is calculated by:

$$d_i = 10^{\frac{P_0 - B_i + X_\sigma}{10\alpha}} \quad (8.2)$$

The real-time estimated location of the mobile device, (x, y) , is computed by:

$$(x, y) = \frac{1}{c} \sum_{i=1}^n \frac{1}{d_i} (x_i, y_i) \quad (8.3)$$

where $c = \sum_{i=1}^n \frac{1}{d_i}$ is the normalization constant and (x_i, y_i) indicates the physical coordinates of the i -th iBeacon.

In an effort to make the log-distance path loss model robust in a semi-outdoor environment, we first conducted experiments to analyze the effects of Non-line-of-sight (NLOS) and the orientation of mobile device on the RSS from BLE beacons.

We performed an experiment in a covered corridor (typical semi-outdoor environment) to analyze the effects of NLOS firstly. The signal strengths were measured at several different distances away from a beacon, which was attached at the ceiling of the corridor. At each location, one user carried a mobile device (Nexus 6) facing toward the beacon, and 100 RSS samples were collected under the line-of-sight (LOS) condition. In addition, another 100 RSS samples were collected when another occupant was standing between the user and the beacon to block the LOS as an NLOS condition. Table 8.1 compares the mean RSS values under both LOS and NLOS conditions at 1–9 meters away from the beacon. As shown in Table 8.1, the NLOS RSS value is 4–5 dBm smaller than the LOS RSS value at each location, because the obstacle (occupant) attenuated the signal strengths. Therefore, we further consider the NLOS effects in the process of log-distance path loss modeling for localization in a semi-outdoor environment.

Table 8.1: Analysis of BLE beacon's RSS variations under LOS and NLOS conditions.

Distance from the Beacon (m)	1	2	3	4	5	6	7	8	9
Line of Sight (LOS) (dBm)	-59	-62	-64	-66	-68	-70	-72	-75	-79
None Line of Sight (NLOS) (dBm)	-63	-66	-68	-69	-72	-75	-77	-80	-83

In addition, we conducted an experiment in the covered corridor to evaluate the influence of different orientations of mobile devices on the RSS emitted from Beacons. Similar to previous experiment, we recorded the signal strengths at several different distances away from a beacon placed at the ceiling of corridor. At each location, one user carried a mobile device (Nexus 6) facing four different orientations (0° , 90° , 180° , 270°) to measure the RSS values. 100 RSS samples were recorded at each orientation. Table 8.2 compares the mean RSS values of four orientations at nine distinct locations from a beacon. It can be observed from Table 8.2 that the RSS values of the 0° holding orientation at all locations are largest among the four directions because the LOS condition is satisfied. On the other hand, the RSS values of 180° are smallest, since the user blocked the LOS. The RSS values recorded at the orientations of 90° and 270° are usually similar to each other because their signal transmission conditions are similar. In summary, the average RSS variation caused by different holding orientations of the mobile device is 1.929 dBm, which should not be ignored for log-distance path loss modeling.

Therefore, we include the effects of NLOS and different holding orientations of the mobile device to precisely estimate the parameters P_0 and α in the log-distance path loss model for semi-outdoor localization. To be specific, we measured RSSs at 14 different distances away from a BLE beacon placed at the ceiling of the corridor. At each location, one user carried a mobile device facing four different orientations with distinct LOS conditions (e.g. 0° , total LOS; 180° , total NLOS; 90° and 270° , partial LOS) to measure the RSS values. 100 RSS samples were recorded at each orientation. P_0 is determined as the mean RSS value at a 0.5-m distance, and α

Table 8.2: Analysis of BLE beacon's RSS variations with the influence of different holding orientations of the mobile device.

Distance from the Beacon (m)	Orientation 0° (dBm)	Orientation 90° (dBm)	Orientation 180° (dBm)	Orientation 270° (dBm)
1	-59	-60	-63	-59
2	-62	-63	-66	-61
3	-64	-66	-68	-63
4	-66	-68	-69	-67
5	-68	-71	-72	-70
6	-70	-72	-75	-72
7	-72	-73	-77	-74
8	-75	-77	-80	-76
9	-79	-81	-83	-80

is estimated by the least-squares method. Based on our experimental results, the estimated values of P_0 is -56.75 dBm and α is 1.577. The raw RSS measurements and mean RSS values at each reference point are demonstrated in Figure 8.3 with the curve fitting by the least-squares method. As shown in Figure 8.3, the RSS value decreases to -70 dBm and remains at this level after 6.5 m. Therefore, we define the minimum effective RSS broadcast from each iBeacon to be $B_{min} = -70$ dBm. Accordingly, BLE beacons only need to be placed with 7–10-m intervals for localization in a semi-outdoor environment.

8.2 Evaluation

We implemented a prototype of BlueDetect on the Android platform and tested its performance using two different mobile devices (Nexus 6 and Samsung Galaxy S4). These devices support the BLE protocol and operate with Android 5.0 OS. The experiments were conducted in the campus of Nanyang Technological University. As shown in Figure 8.4, the test walking route includes all three types of environment scenes: indoors (green section), semi-outdoors (red section), outdoors (blue section), with the total route length around 450 meters. The test walking route was specifically designed to include all transition scenarios. Figure 8.5 demonstrates the

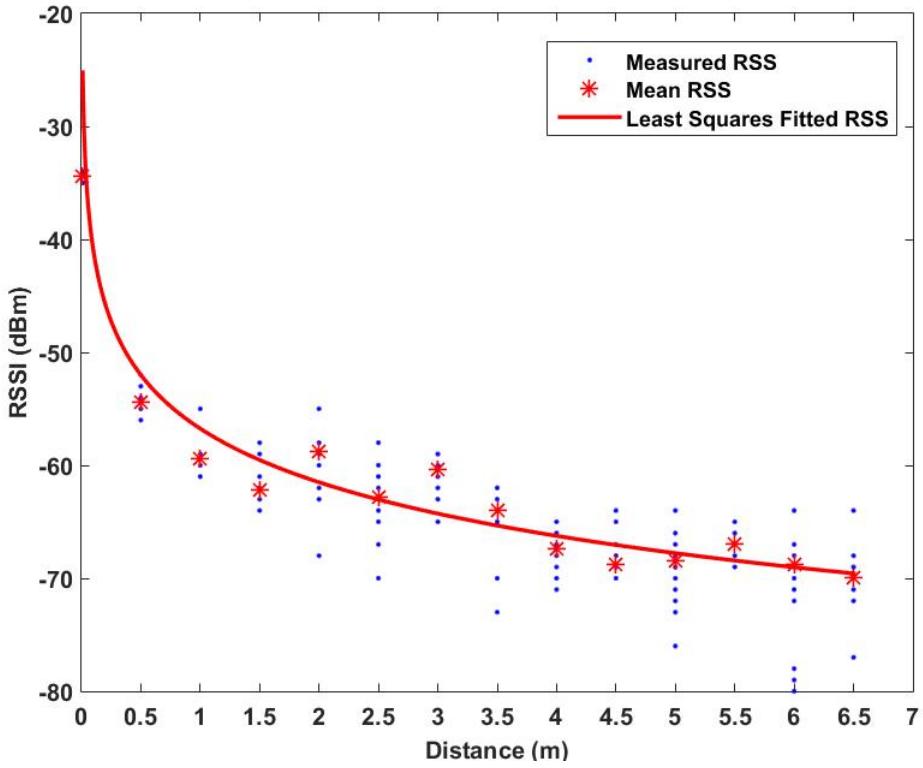


Figure 8.3: Relationship between RSSI of a BLE beacon and distance.

screenshots of BlueDetect in different environment scenes and the corresponding localization techniques adopted.

In the semi-outdoor environment, only 12 low-cost and power-efficient Estimote BLE beacons [157] were placed to cover the total 350 m^2 semi-outdoor area (including covered corridors and connections between buildings). The Estimote BLE beacon is equipped with Nordic Semiconductor's NF51822 chipset, by setting these beacons to a transmission power of -12 dBmW and an advertising interval of 645 ms [157]. Each beacon can provide 21.4 months of life with a CR2477 1000 mAh coin cell battery. As shown in Figure 8.6, the small-sized and battery-powered Estimote BLE beacons can be placed at the ceiling of corridors or attached to walls without cumbersome deployment process. As mentioned in Section 8.1, two pairs of beacons were installed at both sides of the two entrances for detecting seamless transition between semi-outdoor and indoor environments. Our WiFi based IPS [14, 20] is adopted to provide indoor LBS when the indoor environment is con-

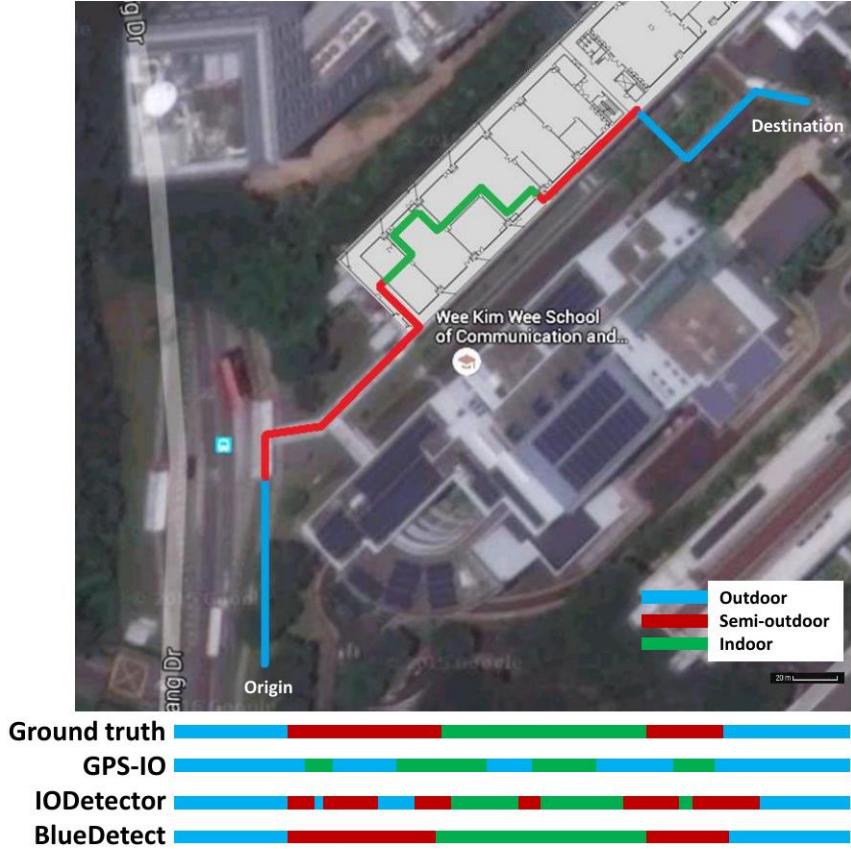


Figure 8.4: The test walking route in the university campus and indoor-outdoor (IO) detection accuracy comparison.

firmed by BlueDetect. During the experiments, the user was walking along the route 10 times at different periods of the day under different weather conditions for a comprehensive evaluation of our system. The following sections will present the performance evaluation of BlueDetect in terms of IO detection accuracy, localization accuracy in a semi-outdoor environment and power consumption, respectively.

8.2.1 IO detection accuracy of BlueDetect

To validate the contextual detection accuracy of BlueDetect, we used a camera to record the entire walking process. In this way, we obtained the real contextual information (outdoors, semi-outdoors or indoors) as the ground truth with timestamps. The IO detection error is introduced when the environmental type identified by the system is different from the ground truth. We defined the IO detection accuracy

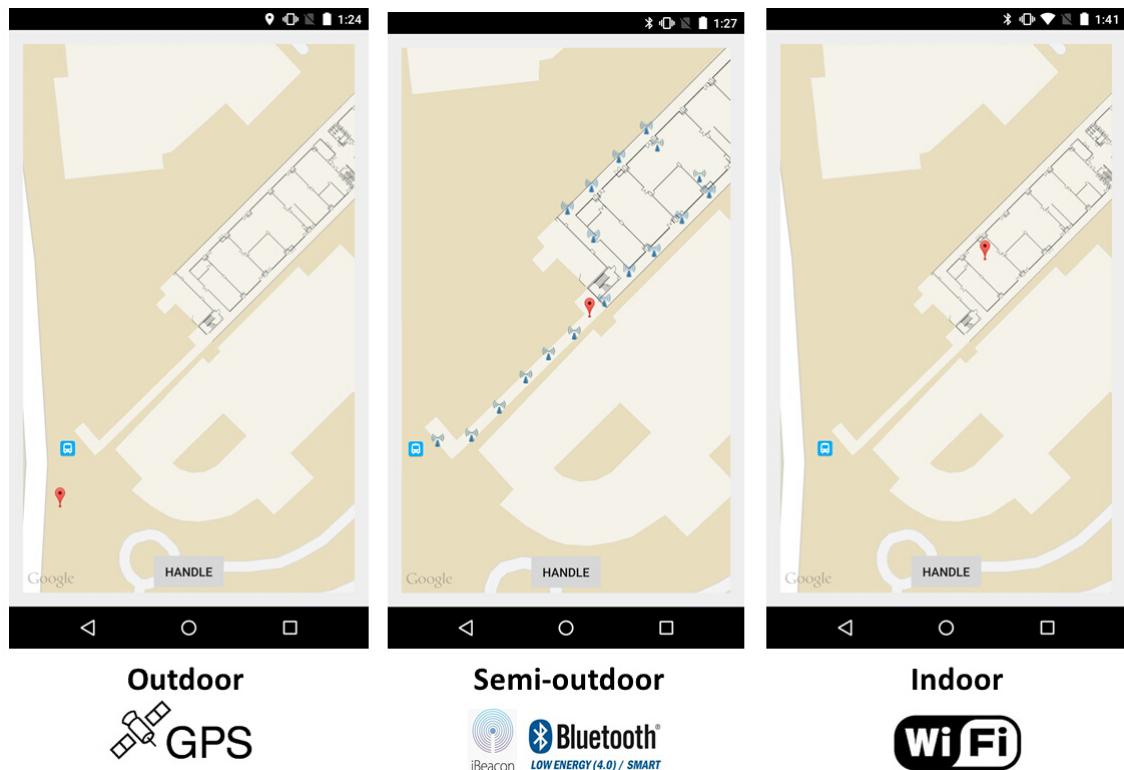


Figure 8.5: Screenshots of BlueDetect in the three environment types.



Figure 8.6: Estimote BLE Beacon.

as the percentile when the system identifies the environmental type to be matched with the ground truth. The IO detection accuracy of BlueDetect is compared to GPS-IO [121] and IODetector [124]. The detection results of the three approaches are illustrated at the bottom of Figure 8.4 compared to the ground truth of the test walking route. It can be observed from Figure 8.4 that the IO detection accuracy of BlueDetect is the best among the three approaches. According to our experimental data, the overall IO detection accuracy of BlueDetect is 96.2%, which is much higher than GPS-IO (56.1%) and IODetector (78.8%). It is clear that BlueDetect can capture each transition between different environments precisely and effectively from outdoors to semi-outdoors, semi-outdoors to indoors, indoors to semi-outdoors or semi-outdoors to outdoors.

In comparison, IODetector suffers from the misdetection of certain semi-outdoor and indoor contexts. The performance of GPS-IO is worst among the three approaches because the GPS signals are still high in some rooms with large windows; these indoor environments are easily misclassified as outdoor environments when the availability of GPS signals is the sole indicator for IO detection. Moreover, the performance of GPS-IO further degrades on rainy or cloudy days, because the LOS paths between the mobile device and satellites are blocked by thick cloud. It is noted that the influence of weather conditions on RSS of BLE beacons is negligible based on our experimental statistics.

8.2.2 Localization accuracy of BlueDetect in semi-outdoor environments

In addition to the IO detection accuracy analysis of BlueDetect, we also evaluate the localization accuracy of BlueDetect in semi-outdoor environments. Existing research works usually assume that GPS or WiFi based IPS can deliver LBS in semi-outdoor environments [76, 122, 123]. However, this assumption hardly holds in reality. In semi-outdoor environments, the localization accuracy of GPS degrades severely since some LOS paths to satellites are blocked by buildings and WiFi based

IPS has a bottleneck due to the sparse deployment of WiFi APs in this scenario. Thus, the localization accuracy of sole GPS, WiFi or the integration of both is not precise and robust in semi-outdoor environments. On the contrary, by only employing a sparse density of BLE beacons, BlueDetect is able to provide accurate and reliable positioning and navigation service in semi-outdoor areas. We recorded the estimated locations of the mobile device by BlueDetect and GPS, and compared these to the measured ground truth. The ground truth was obtained as follows: we first marked the ground with a 1-m grid along the test walking route and measured the physical coordinates of all grid points. Then, we used a camera to record the entire walking process. With the timestamps of grid points, the estimated locations of each grid point by BlueDetect and GPS are compared to its physical location. We define the location estimation error e to be the distance between the real location coordinates (x_0, y_0) and the system estimated location coordinates (x, y) , *i.e.*, $e = \sqrt{(x - x_0)^2 + (y - y_0)^2}$.

According to our experimental results, BlueDetect provides 2.18 m on average in semi-outdoor areas with an enhancement of accuracy around 89.87%, compared to GPS (21.53 m). Figure 8.7 depicts the cumulative distribution of the localization error of these two approaches. The distance error distribution of BlueDetect is mainly within 10 m with the 90th percentile of 7.94 m. On the contrary, the error distribution of GPS is much more scattered, where the 90th percentile is up to 37.33 m.

8.2.3 Power consumption of BlueDetect

Nowadays, the battery consumption of an app is a critical metric to evaluate its feasibility. In this section, we analyze the energy consumption of BlueDetect and compared it to other approaches. BlueDetect performs IO detection by employing the Bluetooth module on mobile device; GPS-IO leverages the GPS module; and IODetector [124] utilizes the on-board light sensor and inertial measurement unit (IMU) sensors (including the accelerator, magnetometer and proximity sensors). In

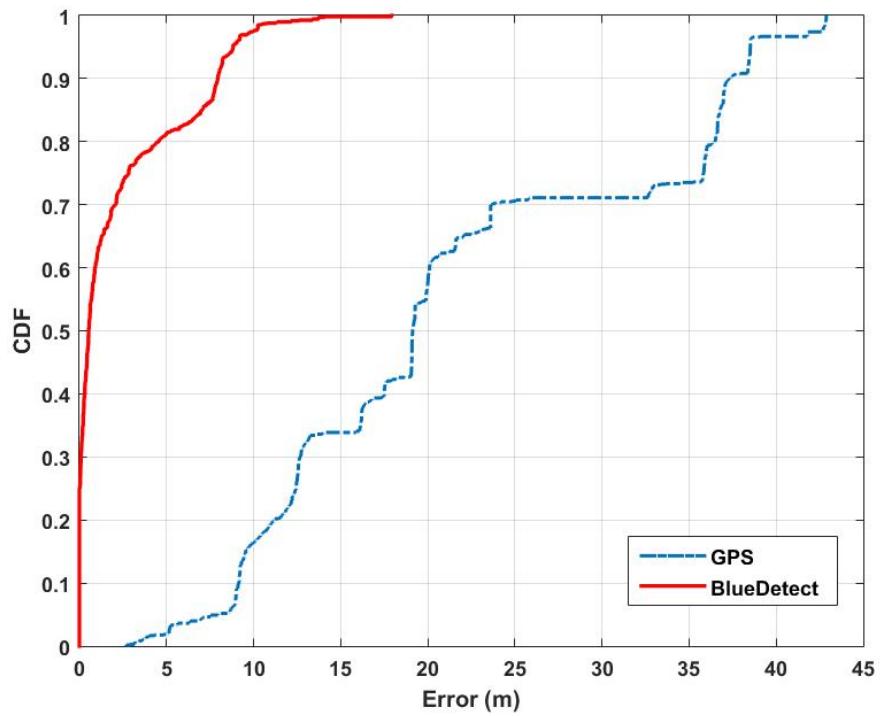


Figure 8.7: Comparison of the cumulative distribution of the location error between GPS and BlueDetect in semi-outdoor environments.

order to calculate how much power the mobile device needs for each approach, we develop a power-monitoring app that allows us to measure the battery consumption of certain on-board sensors precisely. The methodology is that after we select the sensors to be used for IO detection, the power-monitoring app records the battery readings when we start and finish the test walking route. The difference of these two readings is the power consumption of the IO detection scheme. The screen brightness is set at the minimum level during this process. This power consumption evaluation is conducted on a Nexus 6. Figure 8.8 is a screenshot of the power-monitoring app. The sensors adopted for each approach (GPS-IO, IODetector and BlueDetect) were turned on for 30 min and repeated five times. The average experimental data are adopted for evaluation. Figure 8.9 demonstrates how much power is consumed for the three IO detection methods.

As shown in Figure 8.9, the power consumption of BlueDetect is only 119 mAh in 30 min, which is the minimum among the three methods. On the contrary, the GPS module is the most power-hungry sensor, which consumed 213 mAh during

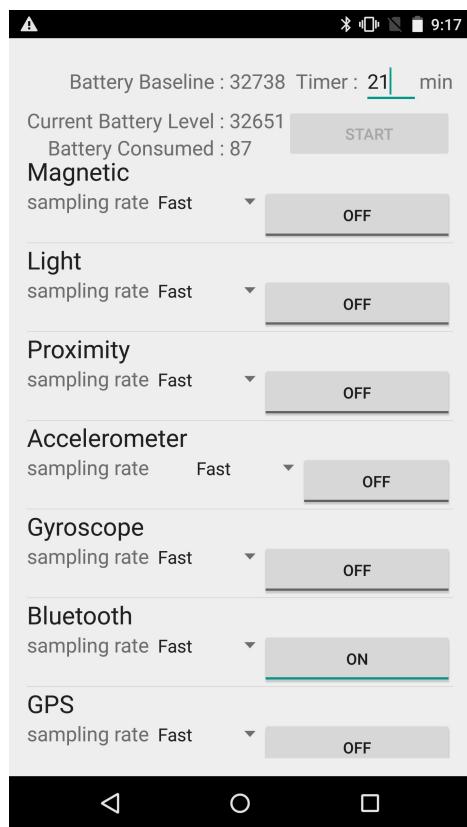


Figure 8.8: Screenshot of the power-monitoring app.

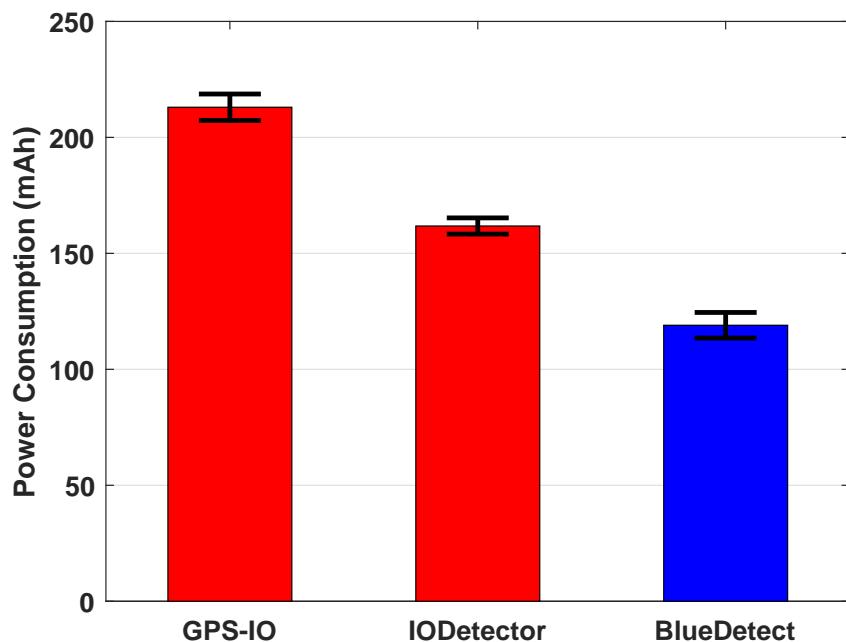


Figure 8.9: Power consumption of various sensors on a mobile device (Nexus 6) for different IO detection methods.

one test walking route. GPS should be turned off in both semi-outdoor and indoor environments since it provides useless information while drawing a huge amount of energy at the same time. The total energy consumption of IODetector is 162 mAh, which is distributed on light, cellular, magnetic field and accelerator and proximity sensors. Although its power consumption is less than GPS-IO, it consumes 26.54% more power than BlueDetect. Moreover, its detection accuracy is restricted by many factors. For instance, the light detector is useless if the mobile device is inside one's pocket while BlueDetect is immune to such a case.

In summary, the high detection accuracy and low energy consumption of BlueDetect makes it affordable for users and upper-layer applications.

8.3 Conclusion

In this chapter, we proposed BlueDetect as an accurate, fast response and energy-efficient scheme for IO detection and seamless LBS running on the mobile device based on the emerging low-power iBeacon technology. By leveraging the portable BLE beacons and Bluetooth module on mobile devices, BlueDetect provides precise IO detection results to turn on/off on-board sensors (such as WiFi and GPS) smartly, improve their performances and reduce the power consumption of mobile devices simultaneously. Furthermore, seamless LBS such as positioning and navigation service can be realized by BlueDetect, especially in semi-outdoor environments, which cannot be achieved by GPS or IPS easily. We prototyped BlueDetect on multiple Android mobile devices and analyzed its performance comprehensively. It provides higher IO detection accuracy, higher localization accuracy in semi-outdoor environments and consumes less battery than existing schemes. It is a feasible solution for IO detection and can be extended for other services such as geo-fencing and floor identification, in the future. One potential issue for BlueDetect is the battery management of BLE beacons. The default broadcast message of iBeacon does not contain its real-time battery level. One possible solution is to add another wireless

communication module such as ZigBee with each beacon so its battery life can be sent out and received by the backend server.

Chapter 9

Conclusions and Future Works

9.1 Conclusions

The explosive proliferation of mobile devices and the popularity of social networks have spurred extensive demands on both outdoor and indoor LBSs in the past decades. Great efforts have been devoted to developing IPSs to enable reliable and precise indoor positioning and navigation, as well as indoor human activity sensing and inference in the past two decades [5–8]. In this thesis, we first analyzed and identified the most longstanding challenges of existing IPS, and then proposed 5 localization algorithms and designed 3 novel IPSs to overcome those challenges and improve the performance of the existing IPSs in terms of localization accuracy, reliability and efficiency. We have implemented and validated our IPSs in 11 distinct indoor environments including lab, office space, lecture theater and lobby, with total 8200 m² across four different regions of Singapore, U.S.A., Taiwan and Germany. Furthermore, various indoor LBSs such as indoor navigation on Google Glass and indoor geofencing for smart lighting control system, have been developed successfully with video demos available on our YouTube Channel [16]. The main contributions of the thesis can be summarized in the following aspects:

1. In Chapter 3, we proposed three localization algorithms: WPL, ELM and integrated WPL-ELM for RF based IPS. WPL is a centralized model-based

approach which does not require any offline site survey procedure and provides accurate location estimation of the target effectively. ELM is a machine learning fingerprinting-based localization algorithm which can provide higher localization accuracy than other existing fingerprinting-based approaches. The integrated WPL-ELM approach combines the fast estimation of WPL and the high localization accuracy of ELM.

2. In Chapter 4, we proposed STI, a new type of fingerprints which embodies more reliable and robust location signatures compared with traditional location fingerprints in the presence of heterogeneous devices and changing indoor environments. We also proposed a novel weighting scheme by taking into consideration the relative importance of each RSS sample according to its corresponding STI value, for the WELM training process. On these grounds, we proposed the STI-WELM scheme which inherits the advantages of both STI and WELM. As can be inferred from experiment results, the STI-WELM scheme enhances the precision of indoor positioning by 39.89% over RSS-ELM, 33.53% over SSD-ELM and 11.46% over STI-ELM, respectively, which confirms the superiority of the STI approach to the traditional RSS fingerprints as well as the recently developed SSD approach.
3. In Chapter 5, we proposed an indoor localization algorithm based on OS-ELM to address the two challenging problems of the existing WiFi-based IPS: the intensive cost of manpower and time for offline site survey, as well as the inflexibility to environmental dynamics. Both simulation analysis and experimental studies have shown that OS-ELM can tackle the problems satisfactorily. The fast learning speed of OS-ELM obviously reduces the time consumption and manpower costs for the site survey during the offline calibration phase. In addition, the online sequential learning ability of OS-ELM makes it possible to reflect and adapt to the environmental changes in a timely manner. Furthermore, it allows a more flexible collection and update of WiFi RSS fingerprints since OS-ELM can learn data with a varying chunk size. Experiments under

specific environmental dynamics such as variations of the occupancy distribution and events of opening or closing doors were also conducted. In summary, OS-ELM can provide high localization accuracy with a fast online sequential learning speed under various environmental changes and achieve superior performance to the existing approaches.

4. In order to reduce the computational burden of the WiFi-based IPS, we have proposed a novel AP selection strategy, OnlineMI, that is able to select the optimal subset of APs to improve the indoor localization accuracy as well in Chapter 6. In comparison with traditional methods, OnlineMI measures the collective discriminate ability of different groups of APs by the mutual information within the group. Furthermore, since its AP selection process is conducted online associated with the real-time location of the mobile device, OnlineMI can select the subset of APs that provide non-redundant information for indoor positioning consistently. The simulation and experimental results have demonstrated that the proposed OnlineMI-WKNN approach is able to relieve the computational burden by selecting a subset of APs online that contains the most critical information for localization, and in the meanwhile preserve or even enhance the localization accuracy of the entire IPS. The overall performance in terms of mean error and variance of error outperforms traditional AP selection methods.
5. In Chapter 7, we proposed, WinIPS, a WiFi-based non-intrusive indoor positioning system that enables automatic online radio map construction and adaptation for calibration-free indoor localization to overcome the major drawbacks of WiFi fingerprinting-based IPS. For RSS data acquisition, we developed WinSMS, a novel intelligent wireless system that can capture data packets transmitted in the existing WiFi traffic and extract the RSS and MAC addresses of both APs and mobile devices in a non-intrusive manner without introducing any extra hardware infrastructure. With the real-time RSS measurements of APs, the APs are becoming natural online reference points for

online radio map construction and adaptation, and hence the tedious offline site survey process can be completely discarded. Furthermore, in order to construct a more fine-grained radio map, we further proposed PSFM-GPR, a reliable regression technique dedicated to predict RSS on VRPs which can precisely capture the irregular RSS distribution over complex indoor environment. The online radio map is more adaptive and robust to environmental dynamics than traditional offline calibrated RSS database since it keeps updated with new measurements. We further proposed STI-WKNN, which determines the similarity between the RSS readings of mobile device and fingerprints stored in AP generated database according to the shapes of RSS vectors rather than raw RSS values to overcome the device heterogeneity issue. Extensive experiments have been carried out to validate the effectiveness of WinIPS in a real-world multi-functional office. The experimental results show that PSFM-GPR achieves 4.8 dBm average RSS estimation error. By leveraging the online radio map generated from PSFM-GPR, the localization accuracy of WinIPS is 1.718m on average. Furthermore, STI-WKNN improves the localization accuracy by 23.95% over the original KNN across heterogeneous mobile devices. In summary, WinIPS overcomes the bottlenecks of current WiFi-based IPS, making it promising for large-scale practical implementation.

6. In Chapter 8, we proposed BlueDetect as an accurate, fast response and energy-efficient scheme for indoor-outdoor (IO) detection and seamless LBS running on the mobile device based on the emerging low-power iBeacon technology. By leveraging the portable BLE beacons and Bluetooth module on mobile devices, BlueDetect provides a precise IO detection to turn on/off on-board sensors (such as WiFi and GPS) smartly, and hence improves their performances and reduces the power consumption of mobile devices simultaneously. Furthermore, seamless LBS such as positioning and navigation service can be realized by BlueDetect especially in semi-outdoor environments, which cannot be achieved by GPS or IPS easily. We prototyped BlueDetect on mul-

multiple Android mobile devices and analyzed its performance through various experiments. As revealed from the experiment result, it provides a higher IO detection accuracy, higher localization accuracy in semi-outdoor environments while consumes less battery than existing schemes. Besides serving as a feasible solution for IO detection, it can also be extended for other services such as geo-fencing and floor identification, in the near future.

9.2 Future works

In this thesis, the longstanding challenges of existing IPSs are identified firstly. Novel IPSs and localization algorithms are designed and proposed to address those challenges and they outperform existing approach in terms of localization accuracy, reliability and efficiency. Based on the research work conducted in the thesis, there are still some possible directions for future studies:

- The performance of our proposed localization algorithms are currently evaluated based on the Root Mean Squared Error (RMSE). However, the performance limitation of these algorithms can be further analyzed theoretically. For instance, the Cramer-Rao Lower Bound (CRLB) sets a lower limit for the covariance matrix of any unbiased estimate of unknown parameters [158]. The study of CRLB on localization using TOA [159], TDOA [160], AOA [161] and RSS fingerprint [162, 163] techniques have been investigated in the literature before. Based on the CRLB estimate, more efficient localization algorithms can be designed by tuning some specific parameters. For example, we may find the optimal parameters such as the number of hidden nodes in hidden layer and the chunk size for OS-ELM algorithms proposed in Chapter 5.
- We have designed, developed and implemented our RFID-based IPS (chapter 3), WiFi-based IPS (Chapters 4–7) and iBeacon-based IPS (chapter 8) successfully, which are operated according to single type of RF signal. Therefore, a potential research topic is to integrate the positioning information from

multiple IPSs in order to realize more accurate and robust indoor LBS. Each IPS has its unique advantages. For instance, WiFi-based IPS performs well in complex indoor environment by adopting fingerprinting based localization algorithms; iBeacon-based IPS is suitable for semi-outdoor environment since it is extremely easy to deploy; PDR-based IPS is able to track the location of mobile devices, as well as their users precisely in short time by leveraging various equipped on-board sensors. Thus, an appropriate fusion framework should be designed to integrate the advantages of these IPSs while compensating their drawbacks simultaneously.

- In Chapters 4–7, our proposed WiFi-based IPS leveraged RSSI values for fingerprinting based indoor localization. The experimental results have demonstrated that our IPS is able to provide around 2 m localization accuracy on average. However, this performance may not satisfy the requirements of some indoor LBSs which require higher localization accuracy. Notice that the localization error may be mainly attributed to the use of RSS measurements, which has two major drawbacks. Firstly, RSS readings suffer from a large variation due to the multipath effects including the diffusion, reflection and diffraction in complex indoor environment. Secondly, RSS is a coarse measurement of the receiver power, which is unable to precisely describe the signal characteristics and introduces undesirable localization error. In comparison, Channel State Information (CSI), which is becoming available on commercial WiFi infrastructures, has a great potential to serve as an alternative to RSS. It is able to provide the RF channel properties (both amplitude and phase) over each Orthogonal Frequency Division Multiplexing (OFDM) subcarrier across the physical layer [164]. Unlike the RSS values, CSI values are more stable and fine-grained. Signal propagation model based algorithms such as TOA, TDOA and AOA can be utilized with CSI measurements due to its ability to extract the LOS path for localization purpose [165]. Therefore, the CSI readings or its combination with RSS and other sensor measurements can be used together

to realize indoor LBS with a much higher localization accuracy.

- Our WiFi-based IPSs proposed in Chapters 4–7 utilized 2.4 GHz and 5 GHz according to IEEE 802.11n standards. The total available bandwidth is limited by 40 MHz, which is a natural limitation on the performance of existing IPSs. Fortunately, the latest released IEEE 802.11ac standards is able to support 160 MHz channel bandwidth, which is four times the bandwidth of 802.11 n. Therefore, the exploration of this wider bandwidth for a more precise indoor localization is an attractive research topic. Furthermore, by leveraging the high frequency band (60 GHz) of IEEE 802.11ad and wide bandwidth of IEEE 802.11ac, WiFi imaging will become possible to capture small human body movement or gesture. In this way, the location and activity of each occupant will be obtained accurately.
- The IPSs developed in this thesis are able to provide accurate 2D location information of the occupants. We will focus on the 3D indoor localization in the future since 2D location information are not sufficient for indoor environments such as shopping malls, hospitals and museums. Other sensors such as barometers and iBeacons will be fused with our IPSs for floor identification and 3D indoor tracking.

Author's Publications

Journal papers:

1. **H. Zou**, B. Huang, X. Lu, H. Jiang, and L. Xie, “A Robust Indoor Positioning System based on the Procrustes Analysis and Weighted Extreme Learning Machine”, *IEEE Transactions on Wireless Communication*, vol. 15, no. 2, pp. 1252-1266, 2016.
2. **H. Zou**, H. Jiang, Y. Luo, J. Zhu, X. Lu, and L. Xie, “BlueDetect: An iBeacon Enabled Scheme for Accurate and Energy Efficient Indoor-Outdoor Detection and Seamless Location-based Service”, *Sensors*, vol. 16, no. 2, pp. 268-286, 2016.
3. **H. Zou**, X. Lu, H. Jiang, and L. Xie, “A Fast and Precise Indoor Localization Algorithm based on Online Sequential Extreme Learning Machine”, *Sensors*, vol. 15, pp. 1804-1824, 2015.
4. **H. Zou**, L. Xie, Q.-S. Jia, and H. Wang, “Platform and Algorithm Development for a RFID-Based Indoor Positioning System”, *Unmanned Systems*, vol. 2, no. 3, pp. 279-291, 2014.
5. Z. Chen, **H. Zou**, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, “Fusion of WiFi, Smartphone Sensors and Landmarks Using the Kalman Filter for Indoor Localization”, *Sensors*, vol. 15, pp. 715-732, 2015.

6. R. Jia, M. Jin, **H. Zou**, Y. Yesilata, L. Xie, and C. Spanos, “MapSentinel: Map-Aided Non-intrusive Indoor Tracking in Sensor-Rich Environments”, *Sensors*, vol. 16. no. 4, pp. 472-491, 2016.
7. **H. Zou**, M. Jin, H. Jiang, L. Xie, and C. Spanos, “WinIPS: WiFi-based Non-intrusive Indoor Positioning System Enabling Online Radio Map Construction”, submitted to *IEEE Transactions on Industrial Informatics*, 2016.
8. **H. Zou**, H. Jiang, J. Yang, L. Xie, and C. Spanos, “Non-intrusive Occupancy Sensing in Commercial Buildings”, submitted to *Energy and Buildings*, 2016.
9. **H. Zou**, S. Chien, H. Jiang, L. Xie, and C. Spanos, “WinLight: An Occupancy based Adaptive Lighting Control System using WiFi-based Non-intrusive Sensing”, to be submitted to *Energy and Buildings*, 2016.

Conference papers:

1. **H. Zou**, Z. Qiu, L. Xie, and Y. Hong, “Consensus-based Parallel Extreme Learning Machine for Indoor Localization”, in *IEEE Global Communications Conference (GLOBECOM)*, 2016.
2. **H. Zou**, B. Huang, X. Lu, H. Jiang, and L. Xie, “Standardizing Location Fingerprints Across Heterogeneous Mobile Devices for Indoor Localization”, in *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 503-508, 2016.
3. **H. Zou**, Y. Luo, X. Lu, H. Jiang, and L. Xie, “A mutual information based online access point selection strategy for WiFi indoor localization”, in *IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 180-185, 2015.
4. **H. Zou**, H. Jiang, X. Lu, and L. Xie, “An Online Sequential Extreme Learning Machine Approach to WiFi Based Indoor Positioning”, in *IEEE World Forum on the Internet of Things (WF-IoT)*, pp. 111-116, 2014.

5. **H. Zou**, L. Xie, Q.-S. Jia, and H. Wang, “An Integrative Weighted Path Loss and Extreme Learning Machine Approach to RFID based Indoor Positioning”, in *IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1-6, 2013.
6. **H. Zou**, H. Wang, L. Xie, and Q.-S. Jia, “An RFID indoor positioning system by using weighted path loss and extreme learning machine”, in *IEEE International Conference on Cyber-Physical Systems, Networks, and Applications (CPSNA)*, pp. 66-71, 2013.
7. M. Jin, **H. Zou**, K. Weekly, R. Jia, A. Bayen, and C. Spanos, “Environmental sensing by wearable device for indoor activity and location estimation”, in *Annual Conference of the IEEE Industrial Electronics Society (IECON)*, pp. 5369-5375, 2014.
8. K. Weekly, **H. Zou**, L. Xie, Q.-S. Jia, and A. Bayen, “Indoor occupant positioning system using active RFID deployment and particle filters”, in *IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 35-42, 2014.
9. X. Lu, C. Yu, **H. Zou**, H. Jiang, L. Xie, “Extreme Learning Machine with Dead Zone and its Application to WiFi Based Indoor Positioning”, in *IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 625-630. 2014.
10. X. Lu, Y. Long, **H. Zou**, C. Yu, L. Xie, “Robust Extreme Learning Machine for Regression Problems with Noisy Measurements in WiFi based Indoor Positioning System”, in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1-6, 2014.
11. K. Weekly, M. Jin, **H. Zou**, C. Hsu, C. Spanos, and A. Bayen, “Building-in-Briefcase (BiB): Sensor Technology Towards Energy-Efficient Buildings”, available online at <http://arxiv.org/abs/1409.1660>, 2015.

12. X. Lu, H. Wen, **H. Zou**, H. Jiang, L. Xie, and N. Trigoni, “Robust Occupancy Inference with Commodity WiFi”, in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016.
13. B. Huang, G. Qi, X Yang, L Zhao, and **H. Zou**, “Exploiting cyclic features of walking for pedestrian dead reckoning with unconstrained smartphones”, in *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pp. 374-385, 2016.
14. **H. Zou**, Y. Zhou, H. Jiang, B. Huang, L. Xie, and C. Spanos, “Adaptive Localization in Dynamic Indoor Environments by Transfer Kernel Learning”, submitted to *IEEE Wireless Communications and Networking Conference (WCNC)*, 2017.
15. **H. Zou**, Z. Chen, H. Jiang, L. Xie, and C. Spanos, “Accurate Indoor Localization and Tracking Using Mobile Phone Inertial Sensor, WiFi and iBeacon”, submitted to *IEEE international Symposium on Inertial Sensors and Systems (INERTIAL)*, 2017.

Posters:

1. **H. Zou**, Y. Zhou, H. Jiang, B. Huang, L. Xie, and C. Spanos, “A Transfer Kernel Learning based Strategy for Adaptive Localization in Dynamic Indoor Environments”, in *ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 462-464, 2016.
2. **H. Zou**, M. Jin, H. Jiang, L. Xie, and C. Spanos, “WinIPS: WiFi-based Non-intrusive Indoor Positioning System Enabling Online Radio Map Construction”, in *IEEE Conference on Computer Communications (INFOCOM)*, pp. 699-700, 2016.

3. R. Jia, M. Jin, **H. Zou**, Y. Yesilata, L. Xie, and C. Spanos, “MapSentinel: Map-Aided Non-intrusive Indoor Tracking in Sensor-Rich Environments”, in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (BuildSys)*, pp. 109-110, 2015.

Awards:

1. **H. Zou**, H. Jiang, and L. Xie, “WiFiGenius: An Accurate and Reliable WiFi-based Indoor Localization and Navigation System”, awarded with the **3rd Place Award** in *Microsoft Indoor Localization Competition - IPSN 2014 (Infrastructure-Free Category)*, Berlin, Germany, 2014.
2. **H. Zou**, H. Jiang, X. Lu, Z. Chen, J. Chen, J. Zhu, Y. Luo, L. Xie, Y. Soh, M. Jin, and C. Spanos, “WiFiGenius: An Accurate and Reliable WiFi-based Indoor Localization and Navigation System”, awarded with the **4th Place Award** in *Microsoft Indoor Localization Competition - IPSN 2015 (Infrastructure-Free Category)*, Seattle, USA, 2015.

Bibliography

- [1] LAUNCH, “Google Maps Takes a Stab at Bing with Indoor Maps Feature for Airports and Malls,” *Google*, 2015, <http://blog.launch.co/blog/google-maps-takes-a-stab-at-bing-with-indoor-maps-feature-fo.html>.
- [2] R. Mautz, “Indoor positioning technologies,” Ph.D. dissertation, Habil. ETH Zürich, 2012, 2012.
- [3] B. Buchli, F. Sutton, and J. Beutel, *GPS-equipped wireless sensor network node for high-accuracy positioning applications*. Springer, 2012, pp. 179–195.
- [4] MarketsandMarkets, “Indoor location market in 2019,” *MarketsandMarkets*, 2016, <http://www.marketsandmarkets.com/PressReleases/indoor-location.asp>.
- [5] H. Liu, H. Darabi, P. Banerjee, and J. Liu, “Survey of wireless indoor positioning techniques and systems,” *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1067–1080, 2007.
- [6] Y. Gu, A. Lo, and I. Niemegeers, “A survey of indoor positioning systems for wireless personal networks,” *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 1, pp. 13–32, 2009.
- [7] G. Deak, K. Curran, and J. Condell, “A survey of active and passive indoor localisation systems,” *Computer Communications*, vol. 35, no. 16, pp. 1939–1954, 2012.

- [8] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, “A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned,” in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 2015, pp. 178–189.
- [9] Department of Energy (DoE), “Energy efficiency trends in residential and commercial building,” *DOE*, 2015, http://apps1.eere.energy.gov/buildings/publications/pdfs/corporate/bt_stateindustry.pdf.
- [10] National Environment Agency (NEA), “Energy Efficiency in Singapore,” *NEA*, 2015, [http://www.nea.gov.sg/cms/ccird/E2%20Singapore%20\(for%20upload\).pdf](http://www.nea.gov.sg/cms/ccird/E2%20Singapore%20(for%20upload).pdf).
- [11] V. L. Erickson, M. Á. Carreira-Perpiñán, and A. E. Cerpa, “Occupancy modeling and prediction for building energy management,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 10, no. 3, p. 42, 2014.
- [12] Y. Agarwal, B. Balaji, S. Dutta, R. K. Gupta, and T. Weng, “Duty-cycling buildings aggressively: The next frontier in HVAC control,” in *10th International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2011, Conference Proceedings, pp. 246–257.
- [13] C. Wu, Z. Yang, Y. Liu, and W. Xi, “Will: Wireless indoor localization without site survey,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 839–848, 2013.
- [14] H. Zou, X. Lu, H. Jiang, and L. Xie, “A fast and precise indoor localization algorithm based on an online sequential extreme learning machine,” *Sensors*, vol. 15, no. 1, pp. 1804–1824, 2015.
- [15] A. Mahtab Hossain, Y. Jin, W.-S. Soh, and H. N. Van, “Ssd: A robust rf location fingerprint addressing mobile devices’ heterogeneity,” *Mobile Computing, IEEE Transactions on*, vol. 12, no. 1, pp. 65–77, 2013.

- [16] H. Zou, “Location based services,” <http://www.youtube.com/user/Parfaiteful>.
- [17] H. Zou, L. Xie, Q.-S. Jia, and H. Wang, “Platform and algorithm development for a rfid-based indoor positioning system,” *Unmanned Systems*, vol. 2, no. 03, pp. 279–291, 2014.
- [18] H. Zou, H. Wang, L. Xie, and Q.-S. Jia, “An rfid indoor positioning system by using weighted path loss and extreme learning machine,” in *Cyber-Physical Systems, Networks, and Applications (CPSNA), 2013 IEEE 1st International Conference on*. IEEE, 2013, pp. 66–71.
- [19] H. Zou, L. Xie, Q.-S. Jia, and H. Wang, “An integrative weighted path loss and extreme learning machine approach to rfid based indoor positioning,” in *4th IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013)*. IEEE, 2013, Conference Proceedings.
- [20] H. Zou, B. Huang, X. Lu, H. Jiang, and L. Xie, “A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine,” *Wireless Communications, IEEE Transactions on*, vol. 15, no. 2, pp. 1252–1266, 2016.
- [21] ——, “Standardizing location fingerprints across heterogeneous mobile devices for indoor localization,” in *Wireless Communications and Networking Conference (WCNC), 2016 IEEE*. IEEE, 2016, pp. 1–6.
- [22] H. Zou, H. Jiang, X. Lu, and L. Xie, “An online sequential extreme learning machine approach to wifi based indoor positioning,” in *Internet of Things (WF-IoT), 2014 IEEE World Forum on*. IEEE, 2014, pp. 111–116.
- [23] H. Zou, Y. Luo, X. Lu, H. Jiang, and L. Xie, “A mutual information based online access point selection strategy for wifi indoor localization,” in *Automation Science and Engineering (CASE), 2015 IEEE International Conference on*. IEEE, 2015, pp. 180–185.

- [24] H. Zou, M. Jin, H. Jiang, L. Xie, and C. Spanos, “Winips: Wifi-based non-intrusive ips for online radio map construction,” in *Computer Communications Workshops (INFOCOM WKSHPS), 2016 IEEE Conference on*. IEEE, 2016, pp. 1081–1082.
- [25] H. Zou, H. Jiang, Y. Luo, J. Zhu, X. Lu, and L. Xie, “Bluedetect: An ibeacon-enabled scheme for accurate and energy-efficient indoor-outdoor detection and seamless location-based service,” *Sensors*, vol. 16, no. 2, p. 268, 2016.
- [26] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, “Epsilon: A visible light based positioning system,” in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 331–343.
- [27] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, “Luxapose: Indoor positioning with mobile phones and visible light,” in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 447–458.
- [28] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos, “How feasible is the use of magnetic field alone for indoor positioning?” in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012, pp. 1–9.
- [29] J. Blankenbach, A. Norrdine, and H. Hellmers, “Adaptive signal processing for a magnetic indoor positioning system,” *Indoor positioning and indoor navigation*, 2011.
- [30] T. E. Abrudan, Z. Xiao, A. Markham, and N. Trigoni, “Distortion rejecting magneto-inductive 3-d localization (magloc),” *IEEE Journal on Selected Areas in Communications*, no. 99, 2015.
- [31] C. Zhang, K. P. Subbu, J. Luo, and J. Wu, “Groping: Geomagnetism and crowdsensing powered indoor navigation,” *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 387–400, 2015.

- [32] V. Y. Skvortzov, H.-K. Lee, S. Bang, and Y. Lee, “Application of electronic compass for mobile robot in an indoor environment,” in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 2963–2970.
- [33] J. Blankenbach and A. Norrdine, “Position estimation using artificial generated magnetic fields,” in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–5.
- [34] R. Want, A. Hopper, V. Falcao, and J. Gibbons, “The active badge location system,” *ACM Transactions on Information Systems (TOIS)*, vol. 10, no. 1, pp. 91–102, 1992.
- [35] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM, 2000, pp. 32–43.
- [36] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper, “Implementing a sentient computing system,” *Computer*, vol. 34, no. 8, pp. 50–56, 2001.
- [37] K. Liu, X. Liu, and X. Li, “Guoguo: Enabling fine-grained indoor localization via smartphone,” in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 235–248.
- [38] A. Ens, F. Höflinger, J. Wendeberg, J. Hoppe, R. Zhang, A. Bannoura, L. M. Reindl, and C. Schindelhauer, “Acoustic self-calibrating system for indoor smart phone tracking,” *International Journal of Navigation and Observation*, vol. 2015, 2015.
- [39] Z. Chen, Q. Zhu, and Y. C. Soh, “Smartphone inertial sensor based indoor localization and tracking with ibeacon corrections,” 2016.
- [40] H. Ying, C. Silex, A. Schnitzer, S. Leonhardt, and M. Schiek, “Automatic step detection in the accelerometer signal,” in *4th International Workshop*

- on Wearable and Implantable Body Sensor Networks (BSN 2007).* Springer, 2007, pp. 80–85.
- [41] J. Kappi, J. Syrjarinne, and J. Saarinen, “Mems-imu based pedestrian navigator for handheld devices,” in *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2001)*, 2001, pp. 1369–1373.
 - [42] Y. Makihara, N. T. Trung, H. Nagahara, R. Sagawa, Y. Mukaigawa, and Y. Yagi, “Phase registration of a single quasi-periodic signal using self dynamic time warping,” in *Asian Conference on Computer Vision.* Springer, 2010, pp. 667–678.
 - [43] P. Goyal, V. J. Ribeiro, H. Saran, and A. Kumar, “Strap-down pedestrian dead-reckoning system,” *measurements*, vol. 2, p. 3, 2011.
 - [44] P. Barralon, N. Vuillerme, and N. Noury, “Walk detection with a kinematic sensor: Frequency and wavelet comparison,” in *Engineering in Medicine and Biology Society, 2006. EMBS’06. 28th Annual International Conference of the IEEE.* IEEE, 2006, pp. 1711–1714.
 - [45] B. Choe, J.-K. Min, and S.-B. Cho, “Online gesture recognition for user interface on accelerometer built-in mobile phones,” in *International Conference on Neural Information Processing.* Springer, 2010, pp. 650–657.
 - [46] S. Shin, C. Park, J. Kim, H. Hong, and J. Lee, “Adaptive step length estimation algorithm using low-cost mems inertial sensors,” 2007.
 - [47] H. Weinberg, “Using the adxl202 in pedometer and personal navigation applications,” *Analog Devices AN-602 application note*, vol. 2, no. 2, pp. 1–6, 2002.
 - [48] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, “Robust pedestrian dead reckoning (r-pdr) for arbitrary mobile device placement,” in *Indoor Positioning*

- and Indoor Navigation (IPIN), 2014 International Conference on.* IEEE, 2014, pp. 187–196.
- [49] R. Harle, “A survey of indoor inertial positioning systems for pedestrians,” *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 3, pp. 1281–1293, 2013.
 - [50] G. Girard, S. Côté, S. Zlatanova, Y. Barette, J. St-Pierre, and P. Van Oosterom, “Indoor pedestrian navigation using foot-mounted imu and portable ultrasound range sensors,” *Sensors*, vol. 11, no. 8, pp. 7606–7624, 2011.
 - [51] A. R. J. Ruiz, F. S. Granja, J. C. P. Honorato, and J. I. G. Rosas, “Accurate pedestrian indoor navigation by tightly coupling foot-mounted imu and rfid measurements,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 178–189, 2012.
 - [52] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, “Lightweight map matching for indoor localisation using conditional random fields,” in *Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on.* IEEE, 2014, pp. 131–142.
 - [53] J. Hightower, R. Want, and G. Borriello, “Spoton: An indoor 3d location sensing technology based on rf signal strength,” *UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA*, vol. 1, 2000.
 - [54] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, “Landmarc: indoor location sensing using active rfid,” *Wireless networks*, vol. 10, no. 6, pp. 701–710, 2004.
 - [55] D. Lieckfeldt, J. You, and D. Timmermann, “Exploiting rf-scatter: Human localization with bistatic passive uhf rfid-systems,” in *Wireless and Mobile Computing, Networking and Communications, 2009. WIMOB 2009. IEEE International Conference on.* IEEE, 2009, pp. 179–184.

- [56] B. Wagner, N. Patwari, and D. Timmermann, “Passive rfid tomographic imaging for device-free user localization,” in *Positioning Navigation and Communication (WPNC), 2012 9th Workshop on*. IEEE, 2012, pp. 120–125.
- [57] K. W. Kolodziej and J. Hjelm, *Local positioning systems: LBS applications and services*. CRC press, 2006.
- [58] X. Jiang, Y. Liu, and X. Wang, “An enhanced approach of indoor location sensing using active rfid,” in *Information Engineering, 2009. ICIE’09. WASE International Conference on*, vol. 1. IEEE, 2009, pp. 169–172.
- [59] R. Heydon, *Bluetooth low energy: the developer’s handbook*. Prentice Hall, 2012.
- [60] L. Pei, R. Chen, J. Liu, T. Tenhunen, H. Kuusniemi, and Y. Chen, “Inquiry-based bluetooth indoor positioning via rssI probability distributions,” in *Advances in Satellite and Space Communications (SPACOMM), 2010 Second International Conference on*. IEEE, 2010, pp. 151–156.
- [61] ibeacon for developers, <https://developer.apple.com/ibeacon/>.
- [62] aislelabs, “The hitchhikers guide to ibeacon hardware: A comprehensive report by aislelabs (2015),” <http://www.aislelabs.com/reports/beacon-guide/>.
- [63] H. Mehmood, N. K. Tripathi, and T. Tipdecho, “Indoor positioning system using artificial neural network,” *Journal of Computer science*, vol. 6, no. 10, p. 1219, 2010.
- [64] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, “Intelligent dynamic radio tracking in indoor wireless local area networks,” *Mobile Computing, IEEE Transactions on*, vol. 9, no. 3, pp. 405–419, 2010.
- [65] P. Bahl and V. N. Padmanabhan, “Radar: An in-building rf-based user location and tracking system,” in *INFOCOM 2000. Nineteenth Annual Joint*

- Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2. Ieee, 2000, Conference Proceedings, pp. 775–784.
- [66] M. Youssef and A. Agrawala, “The horus wlan location determination system,” in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, 2005, Conference Proceedings, pp. 205–218.
 - [67] S.-H. Fang, T.-N. Lin, and K.-C. Lee, “A novel algorithm for multipath fingerprinting in indoor wlan environments,” *Wireless Communications, IEEE Transactions on*, vol. 7, no. 9, pp. 3579–3588, 2008.
 - [68] S.-H. Fang, Y.-T. Hsu, and W.-H. Kuo, “Dynamic fingerprinting combination for improved mobile localization,” *Wireless Communications, IEEE Transactions on*, vol. 10, no. 12, pp. 4018–4022, 2011.
 - [69] N. Alsindi, Z. Chaloupka, N. AlKhanbashi, and J. Aweya, “An empirical evaluation of a probabilistic rf signature for wlan location fingerprinting,” *Wireless Communications, IEEE Transactions on*, vol. 13, no. 6, pp. 3257–3268, 2014.
 - [70] S. He and S.-H. G. Chan, “Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.
 - [71] K. Yedavalli and B. Krishnamachari, “Sequence-based localization in wireless sensor networks,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 81–94, 2008.
 - [72] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, “Walkie-markie: Indoor pathway mapping made easy,” in *Proceedings of the 10th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2013, pp. 85–98.
 - [73] Y. Kim, H. Shin, and H. Cha, “Smartphone-based wi-fi pedestrian-tracking system tolerating the rss variance problem,” in *Pervasive Computing and*

- Communications (PerCom), 2012 IEEE International Conference on.* IEEE, 2012, pp. 11–19.
- [74] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, “No need to war-drive: unsupervised indoor localization,” in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 2012, pp. 197–210.
- [75] Y. Jiang, Y. Xiang, X. Pan, K. Li, Q. Lv, R. P. Dick, L. Shang, and M. Hannigan, “Hallway based automatic indoor floorplan construction using room fingerprints,” in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 2013, pp. 315–324.
- [76] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, “Indoor localization without the pain,” in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 2010, pp. 173–184.
- [77] H. Lim, L.-C. Kung, J. C. Hou, and H. Luo, “Zero-configuration indoor localization over ieee 802.11 wireless infrastructure,” *Wireless Networks*, vol. 16, no. 2, pp. 405–420, 2010.
- [78] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, *WLAN positioning systems: principles and applications in location-based services*. Cambridge University Press, 2012.
- [79] M. Youssef, A. Youssef, C. Rieger, U. Shankar, and A. Agrawala, “Pinpoint: An asynchronous time-based location determination system,” in *Proceedings of the 4th international conference on Mobile systems, applications and services*. ACM, 2006, pp. 165–176.
- [80] D. Niculescu and B. Nath, “Ad hoc positioning system (aps) using aoa,” in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. Ieee, 2003, pp. 1734–1743.

- [81] D. Turner, S. Savage, and A. C. Snoeren, “On the empirical performance of self-calibrating wifi location systems,” in *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*. IEEE, 2011, pp. 76–84.
- [82] A. Matic, A. Papliatseyeu, V. Osmani, and O. Mayora-Ibarra, “Tuning to your position: Fm radio based indoor localization with spontaneous recalibration,” in *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*. IEEE, 2010, pp. 153–161.
- [83] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai, and M. Wiseman, “Indoor location sensing using geo-magnetism,” in *Proceedings of the 9th international conference on Mobile systems, applications, and services*. ACM, 2011, pp. 141–154.
- [84] A. Varshavsky, E. de Lara, J. Hightower, A. LaMarca, and V. Otsason, “Gsm indoor localization,” *Pervasive and Mobile Computing*, vol. 3, no. 6, pp. 698–720, 2007.
- [85] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte, and J. Huerta, “Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems,” *Expert Systems with Applications*, vol. 42, no. 23, pp. 9263–9278, 2015.
- [86] S.-H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” *City*, vol. 1, no. 2, p. 1, 2007.
- [87] B. Ferris, D. Fox, and N. Lawrence, “WiFi-SLAM Using Gaussian Process Latent Variable Models,” in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007, pp. 2480–2485.
- [88] X. Chai and Q. Yang, “Reducing the calibration effort for probabilistic indoor location estimation,” *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 649–662, 2007.

- [89] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, “A probabilistic approach to wlan user location estimation,” *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.
- [90] E. A. Martnez, R. Cruz, and J. Favela, *Estimating user location in a WLAN using backpropagation neural networks*. Springer, 2004, pp. 737–746.
- [91] Z.-l. Wu, C.-h. Li, J.-Y. Ng, and K. Leung, “Location estimation via support vector regression,” *Mobile Computing, IEEE Transactions on*, vol. 6, no. 3, pp. 311–321, 2007.
- [92] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, “Received-signal-strength-based indoor positioning using compressive sensing,” *Mobile Computing, IEEE Transactions on*, vol. 11, no. 12, pp. 1983–1993, 2012.
- [93] C.-t. Huang, C.-h. Wu, Y.-N. Yao-Nan Lee, and J.-T. Chen, “A novel indoor rss-based position location algorithm using factor graphs,” *Wireless Communications, IEEE Transactions on*, vol. 8, no. 6, pp. 3050–3058, 2009.
- [94] J.-g. Park, D. Curtis, S. Teller, and J. Ledlie, “Implications of device diversity for organic localization,” in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011, Conference Proceedings, pp. 3182–3190.
- [95] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, “Kernel-based positioning in wireless local area networks,” *Mobile Computing, IEEE Transactions on*, vol. 6, no. 6, pp. 689–705, 2007.
- [96] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [97] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 2, pp. 513–529, 2012.

- [98] S.-H. Fang, T.-N. Lin, and K.-C. Lee, “A novel algorithm for multipath fingerprinting in indoor wlan environments,” *Wireless Communications, IEEE Transactions on*, vol. 7, no. 9, pp. 3579–3588, 2008.
- [99] S.-H. Fang and T.-N. Lin, “Projection-based location system via multiple discriminant analysis in wireless local area networks,” *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 9, pp. 5009–5019, 2009.
- [100] S.-H. Fang, T.-N. Lin, and P. Lin, “Location fingerprinting in a decorrelated space,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, no. 5, pp. 685–691, 2008.
- [101] M. A. Youssef, A. Agrawala, and A. Udaya Shankar, “Wlan location determination via clustering and probability distributions,” in *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on.* IEEE, 2003, pp. 143–150.
- [102] Y. Chen, Q. Yang, J. Yin, and X. Chai, “Power-efficient access-point selection for indoor location estimation,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 18, no. 7, pp. 877–888, 2006.
- [103] A. W. Tsui, Y.-H. Chuang, and H.-H. Chu, “Unsupervised learning for solving rss hardware variance problem in wifi localization,” *Mobile Networks and Applications*, vol. 14, no. 5, pp. 677–691, 2009.
- [104] H. Cheng, F. Wang, R. Tao, H. Luo, and F. Zhao, “Clustering algorithms research for device-clustering localization,” in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on.* IEEE, 2012, Conference Proceedings, pp. 1–7.
- [105] S.-H. Fang, C.-H. Wang, S.-M. Chiou, and P. Lin, “Calibration-free approaches for robust wi-fi positioning against device diversity: a performance comparison,” in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th.* IEEE, 2012, Conference Proceedings, pp. 1–5.

- [106] C. Figuera, J. L. Rojo-lvarez, I. Mora-Jimnez, A. Guerrero-Curieses, M. Wilby, and J. Ramos-Lpez, “Time-space sampling and mobile device calibration for wifi indoor location systems,” *Mobile Computing, IEEE Transactions on*, vol. 10, no. 7, pp. 913–926, 2011.
- [107] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, “Practical robust localization over large-scale 802.11 wireless networks,” in *Proceedings of the 10th annual international conference on Mobile computing and networking*. ACM, 2004, Conference Proceedings, pp. 70–84.
- [108] M. B. Kjrgaard, “Indoor location fingerprinting with heterogeneous clients,” *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 31–43, 2011.
- [109] F. Della Rosa, H. Leppakoski, S. Biancullo, and J. Nurmi, “Ad-hoc networks aiding indoor calibrations of heterogeneous devices for fingerprinting applications,” in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, Conference Proceedings, pp. 1–6.
- [110] F. Dong, Y. Chen, J. Liu, Q. Ning, and S. Piao, *A calibration-free localization solution for handling signal strength variance*. Springer, 2009, pp. 79–90.
- [111] M. B. Kjrgaard and C. V. Munk, “Hyperbolic location fingerprinting: A calibration-free solution for handling differences in signal strength (concise contribution),” in *Pervasive Computing and Communications, 2008. PerCom 2008. Sixth Annual IEEE International Conference on*. IEEE, 2008, Conference Proceedings, pp. 110–116.
- [112] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Gam, “A system for lease: Location estimation assisted by stationary emitters for indoor rf wireless networks,” in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2. IEEE, 2004, pp. 1001–1011.

- [113] Y. Du, D. Yang, and C. Xiu, “A novel method for constructing a wifi positioning system with efficient manpower,” *Sensors*, vol. 15, no. 4, pp. 8358–8381, 2015.
- [114] M. M. Atia, A. Noureldin, and M. J. Korenberg, “Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks,” *Mobile Computing, IEEE Transactions on*, vol. 12, no. 9, pp. 1774–1787, 2013.
- [115] J. Yin, Q. Yang, and L. M. Ni, “Learning adaptive temporal radio maps for signal-strength-based location estimation,” *Mobile Computing, IEEE Transactions on*, vol. 7, no. 7, pp. 869–883, 2008.
- [116] S. Sorour, Y. Lostanlen, S. Valaee, and K. Majeed, “Joint indoor localization and radio map construction with limited deployment load,” *Mobile Computing, IEEE Transactions on*, vol. 14, no. 5, pp. 1031–1043, 2015.
- [117] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, “Zee: zero-effort crowdsourcing for indoor localization,” in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, Conference Proceedings, pp. 293–304.
- [118] S. J. Pan, J. T. Kwok, Q. Yang, and J. J. Pan, “Adaptive localization in a dynamic wifi environment through multi-view learning.” in *Proceedings of the national conference on artificial Intelligence*, vol. 22, no. 2. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007, p. 1108.
- [119] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, “Lightweight map matching for indoor localization using conditional random fields,” in *The International Conference on Information Processing in Sensor Networks (IPSN’14)*. ACM, 2014.
- [120] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [121] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, “No need to war-drive: unsupervised indoor localization,” in *Proceedings of the*

- 10th international conference on Mobile systems, applications, and services.*
ACM, 2012, Conference Proceedings, pp. 197–210.
- [122] L. Ravindranath, C. Newport, H. Balakrishnan, and S. Madden, “Improving wireless network performance using sensor hints,” in *Proceedings of the 8th USENIX conference on Networked systems design and implementation*. USENIX Association, 2011, pp. 21–21.
- [123] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, “Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones,” in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2009, pp. 85–98.
- [124] P. Zhou, Y. Zheng, Z. Li, M. Li, and G. Shen, “Iodetector: A generic service for indoor outdoor detection,” in *Proceedings of the 10th acm conference on embedded network sensor systems*. ACM, 2012, pp. 113–126.
- [125] V. Radu, P. Katsikouli, R. Sarkar, and M. K. Marina, “A semi-supervised learning approach for robust indoor-outdoor detection with smartphones,” in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. ACM, 2014, pp. 280–294.
- [126] T. Chrysikos, G. Georgopoulos, and S. Kotsopoulos, “Site-specific validation of itu indoor path loss model at 2.4 ghz,” in *World of Wireless, Mobile and Multimedia Networks & Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*. IEEE, 2009, pp. 1–6.
- [127] G.-B. Huang, D. H. Wang, and Y. Lan, “Extreme learning machines: a survey,” *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.
- [128] W. Xiao, P. Liu, W.-S. Soh, and G.-B. Huang, “Large scale wireless indoor localization by clustering and extreme learning machine,” in *Information Fu-*

- sion (FUSION), 2012 15th International Conference on.* IEEE, 2012, pp. 1609–1614.
- [129] G.-B. Huang, H. Babri *et al.*, “Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions,” *Neural Networks, IEEE Transactions on*, vol. 9, no. 1, pp. 224–229, 1998.
- [130] G.-B. Huang, “An insight into extreme learning machines: random neurons, random features and kernels,” *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.
- [131] Y. Jiang, X. Pan, K. Li, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang, “Ariel: Automatic wi-fi based room fingerprinting for indoor localization,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, Conference Proceedings, pp. 441–450.
- [132] V. Radu and M. K. Marina, “Himloc: Indoor smartphone localization via activity aware pedestrian dead reckoning with selective crowdsourced wifi fingerprinting,” in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*. IEEE, 2013, Conference Proceedings, pp. 1–10.
- [133] C. Laoudias, D. Zeinalipour-Yazti, and C. G. Panayiotou, “Crowdsourced indoor localization for diverse devices through radiomap fusion,” in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*. IEEE, 2013, Conference Proceedings, pp. 1–7.
- [134] J. C. Gower, “Generalized procrustes analysis,” *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
- [135] X. Lu, H. Zou, H. Zhou, L. Xie, and G.-B. Huang, “Robust extreme learning machine with its application to indoor positioning,” *IEEE transactions on cybernetics*, vol. 46, no. 1, pp. 194–205, 2016.
- [136] W. Zong, G.-B. Huang, and Y. Chen, “Weighted extreme learning machine for imbalance learning,” *Neurocomputing*, vol. 101, pp. 229–242, 2013.

- [137] S. L. Jackson, *Research methods and statistics: A critical thinking approach*. Cengage Learning, 2015.
- [138] T. S. Rappaport, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.
- [139] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [140] K. Kaemarungsi and P. Krishnamurthy, “Analysis of wlans received signal strength indication for indoor location fingerprinting,” *Pervasive and mobile computing*, vol. 8, no. 2, pp. 292–316, 2012.
- [141] H. Zou, “Indoor navigation on google glass,” <http://www.youtube.com/watch?v=vJ4kJu4ivdE>, 2015.
- [142] Y.-C. Chen, J.-R. Chiang, H.-h. Chu, P. Huang, and A. W. Tsui, “Sensor-assisted wi-fi indoor location system for adapting to environmental dynamics,” in *Proceedings of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2005, pp. 118–125.
- [143] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A fast and accurate online sequential learning algorithm for feedforward networks,” *Neural Networks, IEEE Transactions on*, vol. 17, no. 6, pp. 1411–1423, 2006.
- [144] G. B. Orr and K.-R. Müller, *Neural networks: tricks of the trade*. Springer, 2003.
- [145] G.-B. Huang, P. Saratchandran, and N. Sundararajan, “An efficient sequential learning algorithm for growing and pruning rbf (gap-rbf) networks,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 6, pp. 2284–2292, 2004.
- [146] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.

- [147] J. M. Keller, M. R. Gray, and J. A. Givens, “A fuzzy k-nearest neighbor algorithm,” *Systems, Man and Cybernetics, IEEE Transactions on*, no. 4, pp. 580–585, 1985.
- [148] V. W. Zheng, E. W. Xiang, Q. Yang, and D. Shen, “Transferring localization models over time.” in *AAAI*, 2008, pp. 1421–1426.
- [149] Z. Sun, Y. Chen, J. Qi, and J. Liu, “Adaptive localization through transfer learning in indoor wi-fi environment,” in *Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on*. IEEE, 2008, pp. 331–336.
- [150] OpenWrt. <https://openwrt.org/>.
- [151] Libpcap. <http://www.tcpdump.org/>.
- [152] C. K. Williams and C. E. Rasmussen, “Gaussian processes for machine learning,” *the MIT Press*, 2006.
- [153] B. Ferris, D. Haehnel, and D. Fox, “Gaussian processes for signal strength-based location estimation,” in *In proc. of robotics science and systems*. Citeseer, 2006.
- [154] F. Seco, C. Plagemann, A. R. Jiménez, and W. Burgard, “Improving rfid-based indoor positioning accuracy using gaussian processes,” in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–8.
- [155] S. Kumar, R. M. Hegde, and N. Trigoni, “Gaussian process regression for fingerprinting based localization,” *Ad Hoc Networks*, 2016.
- [156] H. Zou, “Seamless indoor outdoor navigation,” <http://www.youtube.com/watch?v=HPOtwv15sHs>.
- [157] estimote, <https://www.estimote.com>.

- [158] H. L. Van Trees, *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons, 2004.
- [159] A. Dersan and Y. Tanik, “Passive radar localization by time difference of arrival,” in *MILCOM 2002. Proceedings*, vol. 2. IEEE, 2002, pp. 1251–1257.
- [160] C. Chang and A. Sahai, “Estimation bounds for localization,” in *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*. IEEE, 2004, pp. 415–424.
- [161] S. Venkatesh and R. M. Buehrer, “Multiple-access insights from bounds on sensor localization,” *Pervasive and Mobile Computing*, vol. 4, no. 1, pp. 33–61, 2008.
- [162] N. Patwari and A. O. Hero III, “Location estimation accuracy in wireless sensor networks,” in *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, vol. 2. IEEE, 2002, pp. 1523–1527.
- [163] K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan, “Ecolocation: a sequence based technique for rf localization in wireless sensor networks,” in *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 2005, p. 38.
- [164] Z. Yang, Z. Zhou, and Y. Liu, “From rssi to csi: Indoor localization via channel response,” *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, p. 25, 2013.
- [165] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, H. Cai, Y. Liu, and L. M. Ni, “Wifi-based indoor line-of-sight identification,” *Wireless Communications, IEEE Transactions on*, vol. 14, no. 11, pp. 6125–6136, 2015.