[UCI]I804:24011-200000267387

August 2019

Ph.D. Dissertation

Practical Indoor Localization System Using Bluetooth Low Energy Beacons

Graduate School of Chosun University

Department of Information and Communication Engineering

Santosh Subedi

Practical Indoor Localization System Using Bluetooth Low Energy Beacons

저전력 블루투스 비콘을 활용한 실용성 있는
실내 측위 시스템

August 23, 2019

Graduate School of Chosun University

Department of Information and Communication Engineering

Santosh Subedi

Practical Indoor Localization System Using Bluetooth Low Energy Beacons

Advisor: Prof. Jae-Young Pyun

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

April 2019

Graduate School of Chosun University

Department of Information and Communication Engineering

Santosh Subedi

수베디 산토쉬의 박사학위논문을 인준함

위원장 조선대학교 교수 권구락 (인)

위 원 조선대학교 교수 변재영 (인)

위 원 조선대학교 교수 이범식 (인)

위 원 순천대학교 교수 강의성 (인)

위 원 순천대학교 교수 박상현 (인)

2019 년 06 월

조 선 대 학 교 대 학 원

초 록

저전력 블루투스 비콘을 활용한 실용성 있는 실내 측위 시스템

수베디 산토쉬
지도교수: 변재영
정보통신공학과, 대학원, 조선대학교

최근 들어 스마트 기기·기술의 향상으로 위치기반 서비스(LBS)에 대한 사회적·상업적 관심이 크게 높아지고 있다. 이를 위해 GNSS(Global Navigation Satelite System)이 실외 환경에서 정확하고 신뢰할 수 있는 위치 정보를 탐색하고 결정하기 위해 오랫동안 사용되어 왔다. 그러나 GNSS 신호는 너무 약해서 건물을 통과할 수 없고 신뢰할 수 있는 실내 LBS를 제공할 수 없으므로 실내 측위 시스템(IPS)에 대해서 광범위한 연구 개발을 시작하게 되었다.

PS 개발을 위해 다양한 기술과 기법이 연구되어 왔다. 무선 주파수 기반 무선기술, 특히 Wi-Fi와 BLE(Bluetooth Low Energy)는 실내 LBS에 널리 사용된다. IPS를 구현하기 위한 기술로는 근접성, 삼각측량, 삼변측량, 핑거프린팅 등이 있다. 핑거프린팅 기법은 정확도가 뛰어나서 IPS 개발에 널리 적용되어지고 있다. 위치 정보를 얻어내기 위해 AOA(Angle of arrival), TOA(Time of arrival), TDOA(Time difference of arrival), RSS(Received signal strength)와 같은 신호 측정 방법이 활용되고 있으며, 무선기술, 위치 결정 및 신호 측정 방식에 따라 측위 정확도는 달라질 수 있다.

본 논문에서는, 기존의 핑거프린팅 기반 측위 기술의 실질적인 문제를 해결하기 위한 측위 기술들을 제시한다. 첫째, IPS 문제, 실내 위치 측위 기초, AP 신호 특성 및 여과 등을 설명한다. 또한, 위치 측위의 다양한 기법과 신호 측정 원리로 실용 IPS의 실현 이면의 장애물에 대해 논의한다. 둘째, 기본적인 무선 주파수(RF) 기반 IPS 기술과 그 신호 측정 원리를 설명한다. 또한, IPS를 위한 다양한 위치 결정 알고리즘과 성

능 지표를 간략히 논한다. 셋째, Wi-Fi와 BLE 접속점(AP)의 RSS 특성을 제시하고, 테스트베드 환경에서 두 무선기술의 신호 전파 매개변수를 추정한다. 그리고, 시간 변동과 변동하는 RSS를 원활하게 하도록 몇 가지 신호 필터링 방법을 제시한다.

실내 위치 확인 시스템(IPS) 설계에 핑거프린팅 기법이 널리 채택되고 있지만 상당한 문제를 가지고 있다. 계산 복잡성은 측위 효율에 부정적인 영향을 미칠 수 있다. 본 논문에서 일반적인 Wk-NN 핑거프린팅과 WC(Weighted Centroid)를 결합하여 허용 가능한 측위 정확도를 위해 RP(Reference Point)의 수를 줄인다. RP 수가 줄어들기 때문에 핑거프린팅 데이터에 필요한 시간을 단축한다. 마찬가지로, 두 번째 접근방식은 RP의 핑거프린팅을 나타내기 위해 여러 가지 특징을 채택하고, 계산 복잡성을 줄이기 위해 클러스터링을 이용한다. 이 접근방식은 AP의 등급 및 신호 강도인 WC를 핑거프린팅으로 사용한다. IPS의 계산 시간과 측위 추정 오류를 줄이는 클러스터링 구성요소로 APC(Affinity Propagation Clustering)를 사용한다.

비록 핑거프린팅 확률론적 접근방식이 Wk-NN 핑거프린팅 보다 더 정확한 결과를 산출하지만, 측위 효율성은 떨어진다. 시간이 오래 걸리고 노동 집약적인 오프라인 작업량은 실용적 IPS를 실현하기 위해 핑거프린팅 기반 기계 학습 접근법을 제안한다. 사전 측정 없이 핑거프린팅 정보를 예측하고 GPR(Gaussian Process Regression) 및 AP본 논문에서는 IPS를 위해 스마트폰과 BLE 비콘의 수신 신호 핑거프린팅을 이용한다. 제안된 방법이 다양한 LBS에서 유익할 것으로 기대한다.

ABSTRACT

Practical Indoor Localization System Using Bluetooth Low Energy Beacons

Santosh Subedi
Advisor: Prof. Jae-Young Pyun
Department of Information Communication Engineering
Graduate School of Chosun University

In recent time, social and commercial interest on location-based service (LBS) is significantly increased owing to the rise in the number of smart devices and technologies. The global navigation satellite systems (GNSS) has long been employed for LBS to navigate and determine accurate and reliable location information at outdoor environments. However, the GNSS signals are too weak to penetrate buildings and unable to provide reliable indoor LBS. Hence, the incompetence of GNSS at indoor environment invites extensive research and development of an indoor positioning system (IPS).

Various technologies and techniques have been studied for IPS development. The radio frequency based wireless technologies, particularly Wi-Fi and Bluetooth low energy (BLE) are widely used for indoor LBS. The localization techniques on IPS are proximity, trilateration, triangulation, and fingerprinting localization. The fingerprinting localization method is largely accepted for IPS development owing to its good localization accuracy. The signal measurement principles like the angle of arrival (AOA), time of arrival (TOA), time difference of arrival (TDOA), and received signal strength (RSS) are adopted for realizing the localization

techniques. The wireless technology, the positioning technique, and the signal measurement principle can vary according to the required performance metrics of the IPS.

In this dissertation, a practical IPS is presented to solve the real issues of conventional fingerprinting localization. The first half of the dissertation is dedicated to IPS problems, indoor localization fundamentals, and the AP signal characteristics and filtration. First, the hurdles behind the realization of practical IPS is discussed with different localization techniques and signal measurement principles. Second, the basic radio frequency (RF) based IPS technologies and their signal measurement principles are elucidated. In addition, a brief discussion of different positioning algorithms and the performance metrics of the IPS are presented. Third, the dissertation presents the RSS characteristics of Wi-Fi and BLE access points (APs) and estimates the signal propagation parameters of both the wireless technologies in the testbed environment. Moreover, some signal filtering methods are elaborated to smooth the time-varying and fluctuating RSS.

Although fingerprinting is widely adopted for designing indoor positioning system (IPS), it holds significant problems. The labor-intensive and time-consuming offline phase is the major issue in the realization of fingerprinting localization. In addition, computational complexity can possess an adverse effect on localization efficiency. At first, the conventional weighted k-nearest neighbor (Wk-NN) fingerprinting localization is illustrated. Two different improvisation techniques over the typical Wk-NN fingerprinting localization are proposed in the dissertation. The first suggested technique combines the Wk-NN fingerprinting with the weighted centroid (WC) localization to reduce the number of reference points (RPs) for acceptable localization accuracy. Since the number of RPs are reduced, it reduces the time required for acquiring fingerprinting data. Similarly, the second approach employs multiple features to represent a fingerprinting of an RP and utilizes clustering to reduce the computational complexity. This approach uses the WC, a rank and signal strength of APs as

the fingerprint information. It uses affinity propagation clustering (APC) as a clustering component that reduces the computational time and the localization estimation error of the IPS.

While the probabilistic approach of fingerprinting yields more accurate localization result than the Wk-NN fingerprinting, it is computationally expensive. Moreover, the time-consuming and labor-intensive offline workload intensify the practical limits to realize a practical IPS. Hence, keeping this fact on the mind, a machine learning approach of fingerprinting localization is suggested along with the APC, which helps to reduce both the offline workload and online computational cost. The proposed approach predicts the fingerprint information at locations with no prior measurements and clusters the predicted RSS information using Gaussian process regression (GPR) and APC, respectively.

The IPS approaches on this dissertation rely on the received signal fingerprint of BLE beacons that is easy to acquire with the modern smartphones. It is expected that the proposed method will be beneficial in various LBSs.

Index Terms: Affinity propagation clustering, Bluetooth low energy, Fingerprinting localization, Gaussian process regression, Indoor positioning system, Location based services, Weighted centroid localization

Acronyms

AOA	Angle of Arrival
AP	Access Point
APC	Affinity Propagation Clustering
BLE	Bluetooth Low Energy
CDF	Cumulative Distribution Function
EA	Exponential Averaging
GNSS	Global Navigation Satellite Systems
GPR	Gaussian Process Regression
GPS	Global Positioning System
IPS	Indoor Positioning System
ISM	Industrial, Scientific, and Medical
LBS	Location Based Service
MA	Moving Average
MSE	Mean Square Error
NFC	Near Filed Communication
RF	Radio Frequency
RFID	Radio Frequency Identification
RP	Reference Point
RSS	Received Signal Strength
RSSI	Received Signal Strength Indication
TOA	Time of Arrival
TDOA	Time Difference of Arrival
UWB	Ultra-Wide Band
WMA	Weighted Moving Average
WC	Weighted Centroid
Wi-Fi	Wireless Fidelity
Wk-NN	Weighted k-Nearest Neighbor

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Location-based Services

Localization is the discovery of a location of a user, which is a basic need for pervasive applications such as behavior recognition, smart medication, and smart building that require accurate position information of the users to yield accurate and timely services. The demand for location-based service (LBS) has gradually increased at present owing to the rapid development and popularization of smart devices and technologies. The extensively used technology for LBS is global navigation satellite systems (GNSS). GNSS based LBS are employed in consumer products such as vehicle navigation, navigation services on the smartphone and geotagging and in scientific observation systems such as variations in the earth's rotation and monitoring the tectonic plates. The global positioning system (GPS) [1], the Russian GLONASS [2], the European GALILEO [3], and the Chinese BeiDou Satellite Navigation System [4] are some emerging GNSS. Although GNSS-based LBS are widely used, their performance is limited to the outdoor environment only. In addition, indoor environments are often complex owing to the presence of obstacles and environment changes, resulting in signal fluctuation or noise. Hence, it invites extensive research on indoor LBS with alternative wireless technology.

The indoor LBS termed as an indoor positioning system (IPS) is realized with the different signal source or access points (APs). There are two choices of signal source: either already deployed APs like Wi-Fi and geomagnetic fields, or deploy a new signal source like Bluetooth low energy (BLE) beacons and ultra-wideband (UWB) radio signal tags. The signal free solutions in IPS is a dead-reckoning technique that uses off-the-shelf mobile sensors to detect position changes. Some of the wireless signal measuring principles in IPS are received signal strength (RSS), time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA). RSS has been widely used for designing IPS owing to its non-requirement of extra hardware and easy implementation. Indoor LBS is applicable at asset management, people tracking, trade fairs and events, etc.

1.2 The IPS Problem

The GNSS requires line of sight (LOS) signal from at least four satellites to locate a receiver accurately [5]. This requirement of GNSS is not favorable at the indoor environment, which triggers a search for new techniques of IPS. As the indoor environment is complex, multipath propagation and shadowing effect on a radio signal is common. Hence, the received signal can contain LOS and NLOS signal components. It results in less accurate time synchronization and propagation

time measurement, which poses a problem on IPS that relies on signal measurement principles like TOA, TDOA, and AOA. Moreover, the RSS is also unstable owing to superimposition of multipath signals of varying phases. Meanwhile, the magnetic signal has a very limited discernibility in addition to the requirement of proper calibration of the magnetometer in the smartphone.

Since the indoor environment is much less regular compared to outdoor environments, it is difficult to model the indoor radio signal propagation. The indoor signal propagation model is usually based on a propagation path and the known obstacles where minor indoor changes can render the signal propagation model invalid. The localization methods like trilateration [6] and weighted centroid (WC) localization [7] rely on the signal propagation model to estimate the distance from the RSS. Moreover, these methods require precise calibration of path loss exponent for every indoor environment. Fingerprinting localization is the most widely employed IPS that rely on radio fingerprint to produce localization result. However, the training phase of this method is labor-intensive and time-consuming and the time complexity of execution phase grows with the size of the localization area. Moreover, the instability of the RSS at the indoor environment enforces frequent update of the radio map database.

The signal free localization method relies on mobile sensors such as accelerometers, gyroscope, magnetometer, and barometer and can track the users by continuously estimating their displacement from a known starting point. The necessity of known starting point and integrated sensor readings to measure a position resulting in an unacceptable accumulated error are the problems in dead reckoning based IPS.

1.3 Research Objective

The objective of this carried research is to develop the practical indoor localization system using radio signals that can be employed with a smartphone, which can overcome the problems adhered to the conventional fingerprinting localization.

1.4 Contributions of Dissertation

This dissertation aims at contributing to the field of indoor LBS. In particular, this study focuses on the enhancement of training and execution phases of the conventional fingerprinting localization to improve the performance of IPS. In this dissertation, some innovative approaches are presented to reduce the training workload and increase positioning accuracy with reduced execution time complexity.

The first contribution of this dissertation is dedicated to improvising the conventional fingerprinting localization by combining it with WC. The traditional fingerprinting localization is time-consuming and labor work demanding to construct the radio map database. A novel method

is presented to reduce the required number of reference points (RPs) in conventional Weighted K-nearest neighbor (WK-NN) such that it reduces the time required for reading radio frequency signals. In particular, the suggested method employes WC in a pipeline where first WC operation is performed with location and distance information of APs and second WC operation is performed with location and distance information of k RPs estimated by fingerprinting localization operated with lightly populated RPs. The experimental results show that the proposed localization technique can reduce the required fingerprint RPs by more than 40% compared to normal fingerprinting localization method with a similar localization estimation error.

The second contribution of this dissertation is dedicated to improvising the execution phase of the conventional WK-NN fingerprinting localization. The time complexity of the conventional fingerprinting or flat WK-NN fingerprinting method grows with the size of testbed or number of RPs. A novel approach is put forward that uses multiple features to represent a radio fingerprint and uses a clustering technique to reduce the computational time for localization. In particular, the suggested method uses a rank vector, the WC, and the RSS of nearby APs to represent a fingerprint and employs affinity propagation clustering (APC) to reduce the time complexity. The execution phase contains two steps: coarse localization and fine localization. The stored WC is compared with newly estimated WC to determine the cluster head RP in the coarse localization. Later, fine localization is computed from the cluster member RPs using the rank vector, the RSS, and the WC. Experimental results show that the proposed method significantly improves the performance of the positioning system in positioning accuracy, radio-map database size, and computational time.

The performance of probability based fingerprinting localization is better than Wk-NN fingerprinting localization. However, the demand of time and effort to construct a radio map and high time complexity of execution phase still persist. The third contribution of this dissertation is dedicated to minimizing both the training workload and execution time complexity of probability based fingerprinting localization. In particular, the proposed method uses regression and clustering to construct the radio map and to minimize the searching space of RPs, respectively. This study uses Gaussian process regression (GPR) to predict the RSS at locations with no prior measurements and the constructed radio map is further preprocessed to obtain the RSS clustering information using APC. Experimental results show that the suggested probability based fingerprinting approach can reduce the offline workload, increase localization accuracy with less computational cost, and outperform the existing methods.

1.5 Thesis Overview

In this chapter, after an initial introduction to the LBS, the IPS problems are discussed, wherein challenges for both the signal measurement principles and realizing IPS methods are explained.

Then research objective is briefly summarized followed by the contributions of the dissertation.

Chapter 2 discusses the fundamentals of indoor localization including the signal measurement principles upon which localization relies, the available technologies for indoor localization, as well as common localization estimation algorithms. This chapter is concluded by summarizing some common performance metrics that constitute a criterion for IPS evaluation.

Chapter 3 presents the characteristics of Wi-Fi and BLE signals that are the basis of many radio signal based IPS. Particularly, the Wi-Fi and BLE signals are compared in terms of distance estimation and RSS resolution and estimate their signal propagation parameter in the testbed. Furthermore, some RSS filtration methods and the obtained results are illustrated.

In Chapter 4, the conventional WK-NN fingerprinting localization algorithm and practical issues on realizing the method are discussed. The proposed fingerprinting methods to address the issues in training and execution phase of the WK-NN fingerprinting localization are elaborated, implemented, and analyzed. Additionally, APC algorithm is explained.

In Chapter 5, operating algorithm and implementation issues of probability based practical fingerprinting localization are explained. Particularly, the regression approach using GPR is elaborated. The suggested IPS approach based on GPR is investigated and analyzed.

Chapter 6 concludes the thesis, outlining the lessons drawn from the study.

Chapter 2

Indoor Localization Fundamentals

2.1 Introduction

The use of the localization system in our daily life has a long history. Various physical entities such as cairn (human-made pile of stones), the position of stars in the sky, and the geomagnetic field have long been used to estimate the location. With the development of techniques and technologies, the localization system has evolved a lot. For example, the ancient Greek philosopher Thales discovered the underlying theory of triangulation [8]. Later, GPS using RF technology achieved great success for outdoor positioning. The precise inertial measurement devices allow missile and aircraft to localize themselves and navigate accurately.

The indoor LBS has been gaining more attention at present owing to the development of modern smartphones and wireless technologies. Various localization estimation techniques are available depending on the kind of technology used and measurement made. The major IPS technologies are Wi-Fi, BLE, UWB, PDR, visible light, etc. Some of the IPS algorithms are fingerprinting, trilateration, WC localization, and proximity. These technologies and techniques can be employed to realize and IPS using the measurement quantities like time of arrival (TOA), the angle of arrival (AOA), RSS, sensor readings, etc. Different techniques have their own advantages and disadvantages.

In this chapter, the technologies for IPS are first discussed, followed by a discussion on different signal measurement principles or the measurement quantities. Later, some of the common IPS algorithms are illustrated. The discussion on different performance metrics of IPS concludes this chapter.

2.2 The Technologies for Indoor Localization

Many representative technologies have been employed to develop IPS. The technologies can be categorized into a radio frequency (RF), light wave, acoustic wave, and mechanical as shown in Figure 2.1.

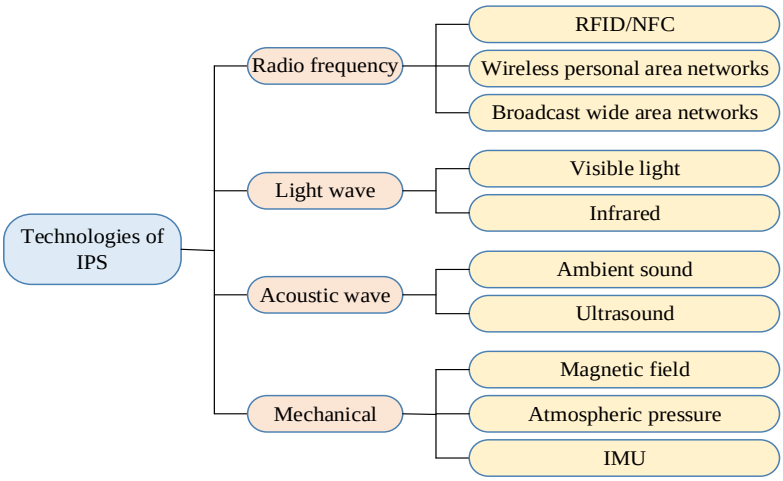


Figure 2.1: Categorization of major technologies used in IPS development.

RF technology such as Wi-Fi and BLE are the most used technologies in IPS development. Similarity, magnetic field, IMU, and atmospheric pressure can be employed for IPS development. Apart from them, light wave such as visible light and infrared, acoustic waves such as ambient sound and ultrasound are also utilized in indoor localization. Since the thesis aims to study IPS approaches based on BLE, an overview of the localization solutions based on some selected RF technologies is emphasized.

The RF technology can be further categorized into broadcast wide area network (WAN), wireless personal area network (WPAN), and RFID/NFC. The cellular network, TV/FM radio signals, and GPS repeaters belong to WAN family, whereas Wi-Fi, BLE, WSN, and UWB belong to WPAN family.

2.2.1 RFID/NFC

The radio frequency identification (RFID) systems rely on two main components namely RFID tag and RFID reader to fulfill their objective. The RFID reader wirelessly acquires the electronically stored information of RFID tags. The reader contains a transceiver to transmit RF signals and read the data emitted from the tags. The tags can be categorized as passive and active. The passive tags get energy from incoming radio signal whereas active tags are powered by a battery. RFID systems operate in four frequency bands: low frequency (125 kHz), high frequency (13.56 MHz), ultra-high frequency (433, 868-915 MHz), and microwave frequency (2.45 GHz, 5.8 GHz). The property of RFID to detect and recognize the nearby tag makes it enable to use for IPS.

Some of the RFID localization system using passive tags are [9, 10, 11, 12] where the tags are deployed on the floor at fixed distance forming a grid and estimate localization results by detecting multiple tags. The work in [13, 14] are based on active tags where RSS is used to estimate the location of the user. RFID has also been combined with other technologies for IPS. For example, [15] combines it with the ultrasonic sensor, [14, 16] combine with image sensors to detect the location of objects.

Near Field Communication (NFC) is the short-range (5 centimeter or less) wireless communication technology. In particular, it is a specialized branch within the family of RFID technologies (high-frequency band of RFID). Localization can be realized with NFC by deploying a number of tags at places of interest, where a location is estimated simply by touching the tag with the NFC equipped device [17, 18].

2.2.2 UWB

Ultra-wideband (UWB) uses very low energy for short-range and high-bandwidth communications over a large portion of the radio spectrum. In general, an emitted radio wave is considered UWB if its bandwidth exceeds 500 MHz or 20% of the carrier frequency. The properties of UWB such as very less power consumption, effective penetration through dense materials, and less sensitive to the multipath effect owing to a very short duration of UWB pulses makes the UWB suitable for IPS development.

IPS based on UWB can estimate location accurately owing to the possibility of precise time measurements of the propagation time of UWB pulses. Yanjia et al. proposed a robust method to mitigate the path overlapping effects that induce TOA and AOA based positioning inaccuracy [19]. Their method is based on the spectral observation of beamforming and yields the least squares estimation of joint TOA and AOA with low computational cost. As the performance of UWB based IPS is deteriorated in NLOS channel, [20] proposed a method to identify NLOS by measuring signal strengths in the first path and multi-path. RSS based UWB IPS system has also been put forward to have accuracy between 0.1 to 0.2 meters [21].

2.2.3 WSN

Wireless sensor networks (WSN) are the group of spatially dispersed and dedicated sensor nodes for monitoring and recording the physical conditions of the environment and organizing the collected data at a central location [22]. The nodes of WSN are equipped with a processor, storage, a power supply, a transceiver, and one or many sensors, with an actuator. WSN operates at an unlicensed band of 2.4 GHz and it can use a number of off the shelf wireless technologies like Bluetooth, UWB, and ZigBee where most applications use IEEE 802.15.4 and ZigBee [23].

Some of the sensor nodes in WSN called anchor node are aware of their own position information. Therefore the localization problem in WSN based IPS is to determine the location

of other nodes based on location information obtained from the anchor nodes. IPS using WSN normally consists of distance or angle estimation between nodes or their combination to produce localization result.

2.2.4 Wi-Fi

Wi-Fi is a technology for radio wireless local area networking (WLAN) of devices based on 802.11 IEEE network standard, operating in the 2.4 and 5 GHz ISM radio bands. The devices that can use Wi-Fi include PCs, smartphones/tablets, smart TVs, video game consoles, digital audio players, cars, and printers. Wi-Fi is the most popular means of communicating data wirelessly and it is increasingly deployed everywhere including home and public indoor environments. Wi-Fi-based IPS are being intensively studied owing to the widespread deployment of Wi-Fi hot spots.

The localization methods like proximity, fingerprinting, trilateration, and triangulation can be realized using Wi-Fi. A basic Wi-Fi signal acquisition procedure is illustrated in Figure 2.2.

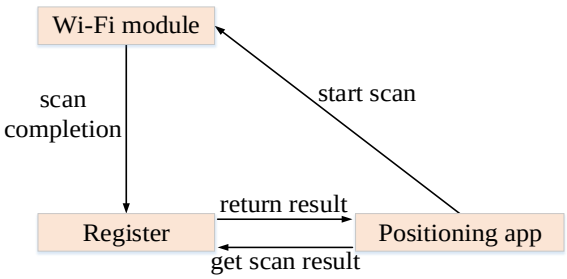


Figure 2.2: Wi-Fi signal acquisition procedure in an Android Smartphone.

Here, the positioning application sends the command to start the scanning process to the inbuilt Wi-Fi module. The Wi-Fi module starts scanning nearby Wi-Fi signals and when scanning is finished, it stores it in the cache. Moreover, it also sends the scan completion message to the operating system (OS). Later, the OS informs the positioning application about scan completion. The positioning application then sends a command to get the scanned results stored in the cache.

2.2.5 BLE

Bluetooth low energy (BLE) was released as Bluetooth version 4.0 in June 2010. The BLE is designed for devices that do not require large amounts of data transfer and is intended for short-range wireless transmission with low power consumption and cost [24, 25]. It is reported that the power draw of the smartphone is lower for BLE than for Wi-Fi [26]. Similar to Wi-Fi, BLE

operates at an ISM band of 2.4 GHz. The frequency band is divided into 40 channels spaced at 2 MHz apart, among which the three channels are used for broadcasting advertisement packets. It is noteworthy that the three advertising channels are strategically placed to avoid interference with coexisting technologies such as IEEE 802.11 and ZigBee [27].

Similar to Wi-Fi, the localization methods like proximity, trilateration, and fingerprinting can be realized using BLE. The tag device can estimate the RSS from a nearby BLE beacon by intercepting the advertisement packets transmitted by the beacons. A brief illustration of the BLE signal scanning process is given in Figure 2.3.

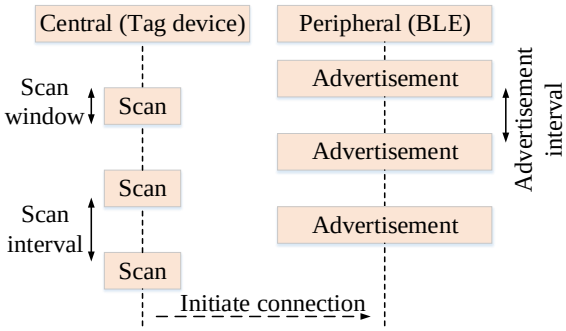


Figure 2.3: BLE signal scanning process..

As shown in Figure 2.3 , the tag device scans for advertisement packets transmitted by the BLE beacons. The advertisement interval can range from (100 to 2000) milliseconds. The typical advertisement interval of BLE beacons used in IPS is 300 milliseconds by considering the normal walking speed (1.3 m/s). Moreover, the scan interval also can be set in the positioning application. A typical value of scan interval is 1000 millisecond (1 sec) to produce positioning result at every second.

2.3 Signal Measurement Principles

2.3.1 RSS

The received signal strength (RSS) is a measurement of power present in a received radio signal. The RSS value is measured in decibel-milliwatt (dBm) and has a typical negative value ranging between nearly 0 dBm (excellent signal) to less than -100 dBm (poor signal). As the distance between the transmitter and receiver increases, the RSS gets attenuated owing to many factors including the antenna of transmitting and receiving devices, number of walls and floors, the number of people and furniture, etc. Note that the RSS does not decrease linearly as the distance increases

[28].

RSS modeling is usually done by the combined effects of large-scale fading and small-scale fading [29]. The large-scale fading component depicts the signal attenuation as the signal travels over a distance and is absorbed by objects such as walls and floors along the way to the smartphone. This fading component predicts the mean of the RSS and usually has a log-normal distribution [30]. Similarly, the small-scale fading describes the fluctuation of signal due to multipath fading. For the NLOS component, the small-scale fading is modeled with a Rayleigh distribution whereas, for LOS component, it is modeled by Rician distribution. In IPS, the fluctuating RSS are filtered using many approaches such as Gaussian filter [31], moving average filter, and exponential averaging [32].

2.3.2 TOA

The time of arrival (TOA) is the travel time or time of flight of a radio signal from a transmitter to a receiver. As the signal travels with a known velocity, the distance can be directly calculated from the TOA. Figure 2.4 illustrates a TOA measurement based localization system.

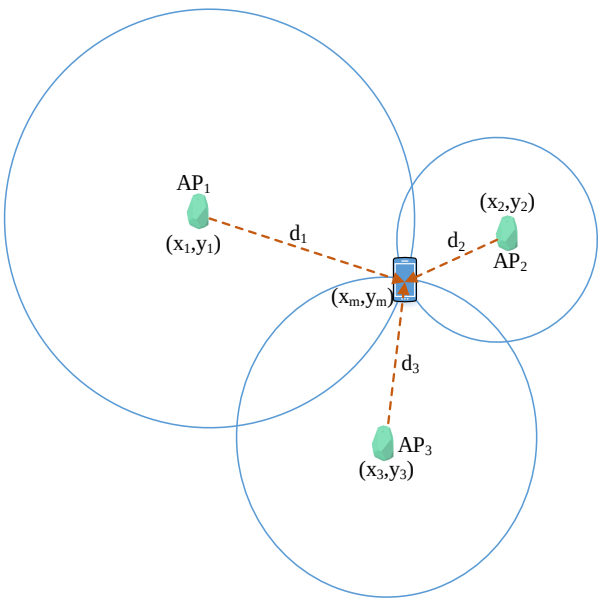


Figure 2.4: Localization based on TOA measurement.

Let c be the speed of light, then the distance between i^{th} AP and the tag device can be estimated by the following relation:

$$d_i = (t_i - t_0) \times c \tag{2.1}$$

where t_0 and t_i are the time instant of signal transmission and signal reception respectively, and $c = 3 \times 10^8\ m/s$. The TOA technique require preciese time synchronization for transmitters and receivers. The estimated distance can be utilized for trilateration algorithm to estimate user location. TOA has been used with various wireless technologies like UWB [33] and Wi-Fi [34].

2.3.3 TDOA

For time difference of arrival (TDOA) measurement, the difference in arrival time from multiple APs is employed. In TDOA based localization, the distance difference between the tag device and APs is calculated based on time difference measurements as shown in the following Figure 2.5.

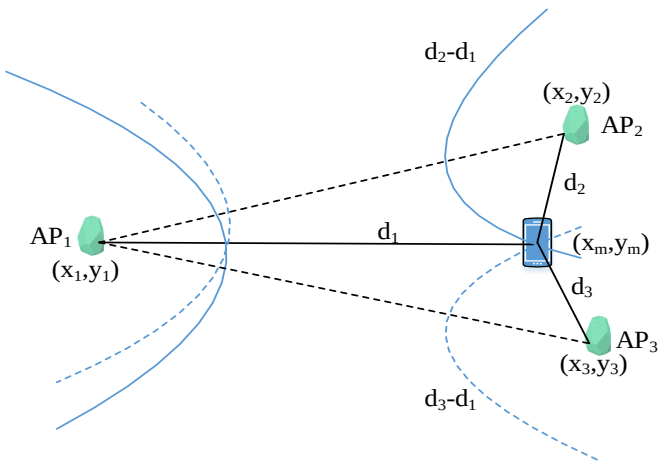


Figure 2.5: Localization based on TDOA measurement.

Here, the difference of distance to APs and to the AP where the signal first arrives is:

$$d_{ij} = (t_i - t_j)c = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2} - \sqrt{(x_j - x_m)^2 + (y_j - y_m)^2} \tag{2.2}$$

where t_i and t_j are the time instant of signal reception from AP i and j, respectively. Geometrically, with a given TDOA measurement, the tag device must lie on a hyperboloid with a constant range difference between the two APs. Apart from TOA, TDOA needs time synchronization of APs only.

2.3.4 AOA

The angle of arrival (AOA) measurement determines the direction of propagation of a radio wave incident on an antenna array. AOA determines the direction by measuring the TDOA at individual elements of the antenna array. AOA-based localization system estimates the location of the tag device as the intersection point of pairs of hypothetical signal paths particular angles as shown in the following Figure 2.6

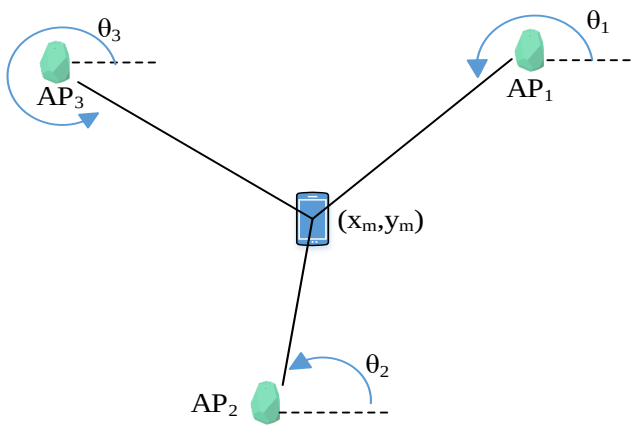


Figure 2.6: Localization based on AOA measurement.

At 2D plane, the AOA approach requires only two APs. In AOA-based IPS, time synchronization between the APs and the tag device is not required. However, it may require relatively complex hardware to obtain angle measurement.

2.4 Algorithms for IPS

Localization estimation at indoor is determined by employing different signal measurement principles to measure the physical quantities that change accordingly to changes in the position of the tag device. Different localization algorithms can be implemented depending on different localization technologies and signal measurement principles. The main localization algorithm for IPS is categorized in Figure 2.7.

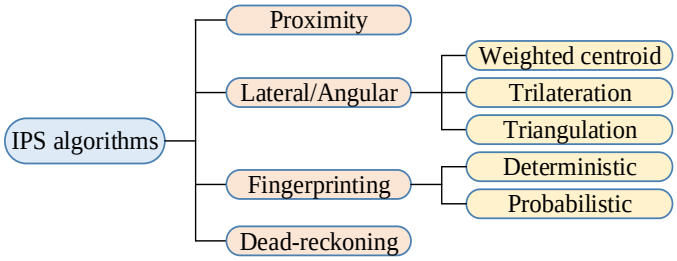


Figure 2.7: Categorization of the main algorithms used in IPS development.

All the above-categorized algorithms have their own advantages and disadvantages, among them fingerprinting has been widely used due to its high localization accuracy.

2.4.1 Proximity

The word proximity is defined as nearness in space, time or relationship. As the definition suggests, proximity in IPS provides symbolic location information if an object is present within the vicinity of an AP where the vicinity is determined by the received signal strength. A proximity-based IPS is illustrated in Figure 2.8.

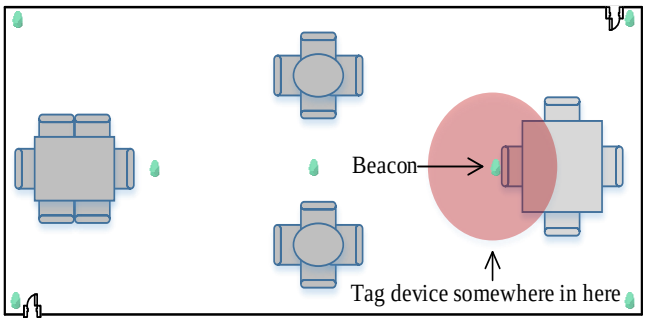


Figure 2.8: Variation on number of iterations with respect to the damping factor.

In proximity based IPS, when an AP is detected by a tag device, the tag's position is associated with the AP's location. In this scenario, when the tag device detects more than one APs in its vicinity, the tag device's location can be referred to the real location of the AP having the strongest signal. The proximity-based IPS is the simplest among all the algorithms and it is very easy to

implement. However, for better localization accuracy or high resolution, a dense deployment of APs is mandatory. Generally, the IPS using wireless technologies like WSN, BLE [35, 36], RFID [37], and NFC [17] is employed for proximity-based IPS development.

2.4.2 Lateral/Angular

The lateral technique estimates the position of a tag device by measuring the distances from multiple APs. The distance can be obtained from signal measuring principles like RSS, TOA, and TDOA. Similarly, the angulation technique estimates the location of the tag device by computing angles relative to multiple APs using AOA. WC localization and trilateration are the lateral methods whereas triangulation is the angulation method.

WC Localization

In proximity based IPS, when the tag device detects multiple APs in its vicinity, the location can be estimated as a centroid of the real location of the detected APs. Furthermore, certain weight can be assigned to each detected AP based on their signal strength to estimate a weighted centroid. The simplest WC localization equation is defined by following set of equations:

$$\begin{aligned} x_w &= \frac{\sum_{j=1}^{u} x_j \times w_j}{\sum_{j=1}^{u} w_j} \\ y_w &= \frac{\sum_{j=1}^{u} y_j \times w_j}{\sum_{j=1}^{u} w_j} \\ w_j &= \frac{1}{d_j^g} \end{aligned} \tag{2.3}$$

where (x_w, y_w) is the estimated WC, (x_j, y_j) is the previously known AP coordinate, d_j is the estimated distance between the tag device and j^{th} AP, g is the degree of weight, and u is the total number of APs considered for WC localization.

Figure 2.9 illustrates the WC localization procedure. The WC localization has the following characteristics:

1. The estimated location is confined inside the APs real location only.
2. The estimated location is dragged towards the nearest AP (BLE beacon) owing to its largest weight.

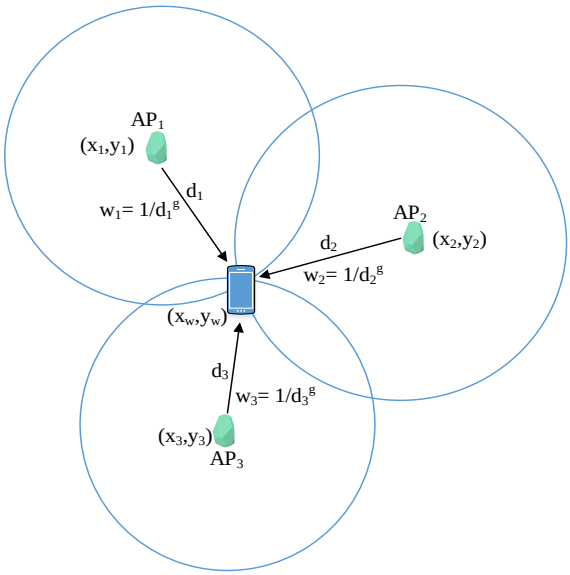


Figure 2.9: Procedure for estimating WC localization at a tag device.

The degree of weight (g) can be adjusted as per the distance between the deployed APs. A large value of g drives the WC location very close to the real location of the AP with the strongest signal, whereas a very low value (close to zero) yields a geometrical centroid among the u APs.

Trilateration

Trilateration is based on measured distances between a tag device and a number of APs with their known real location coordinates. Given the distance to an AP, it is known that the tag device must be along the circumference of a circle centered at the AP and radius equal to the tag-AP distance. For a 2D localization, at least three non-collinear APs are needed whereas, for 3D localization, at least four non-coplanar APs are required to perform trilateration operation. The AP (beacon) deployment scenario for 2D trilateration localization is shown in Figure 2.10.

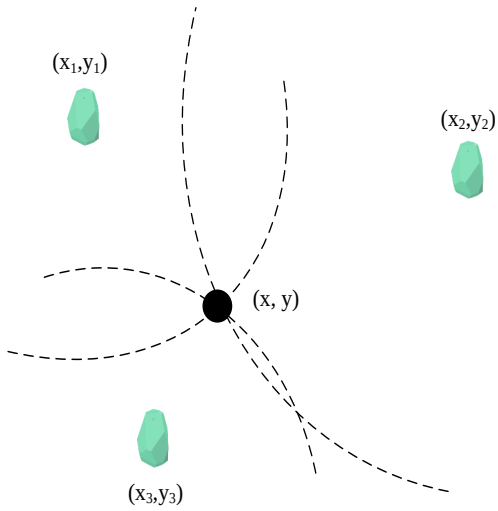


Figure 2.10: AP deployment for trilateration localization where the black dot represents a tag device.

Let us consider B APs with their real location coordinate x_i = $(x_i, y_i\ )$ $(i = 1, 2, \ldots .B)$ and unknown location of the tag device be x = (x, y). The distances between the tag device and the APs is d_i $(i = 1, 2, \ldots, B)$. The relationship between APs/tag positions and their distances in 2D can be written as:

$$\begin{bmatrix} (x_1 - x)^2 + (y_1 - y)^2 \\ (x_2 - x)^2 + (y_2 - y)^2 \\ \vdots \\ (x_B - x)^2 + (y_B - y)^2 \end{bmatrix} = \begin{bmatrix} d_1^2 \\ d_2^2 \\ \vdots \\ d_B^2 \end{bmatrix} \tag{2.4}$$

Equation (2.4) can be represented as $Ax = b$ where A and b are defined as:

$$A = \begin{bmatrix} 2(x_B - x_1) & 2(y_B - y_1) \\ 2(x_B - x_2) & 2(y_B - y_2) \\ \vdots & \vdots \\ 2(x_B - x_{B-1}) & 2(y_B - y_{B-1}) \end{bmatrix} \tag{2.5}$$

$$b = \begin{bmatrix} d_1^2 - d_B^2 - x_1^2 - y_1^2 + x_B^2 + y_B^2 \\ d_2^2 - d_B^2 - x_2^2 - y_2^2 + x_B^2 + y_B^2 \\ \vdots \\ d_{B-1}^2 - d_B^2 - x_{B-1}^2 - y_{B-1}^2 + x_B^2 + y_B^2 \end{bmatrix} \tag{2.6}$$

The location of the tag device can be estimated based on the least squares system using x $= (A^T A)^{-1} A^T b$.

Triangulation

In contrast to trilateration, triangulation uses angle measurements in addition to distance measurements to estimate the position of the tag device. Two angles and one length are required for a 2D localization. Particularly, triangulation utilizes the geometric properties of triangles to estimate the tag location. A simple illustration of triangulation is shown in Figure 2.11.

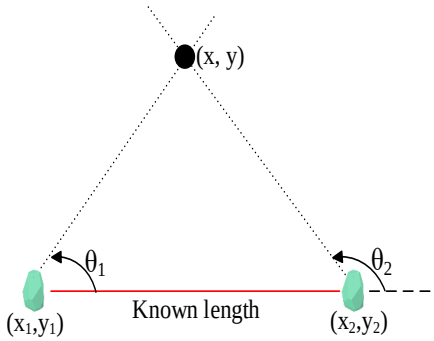


Figure 2.11: AP deployment for triangulation localization where the black dot represents a tag device.

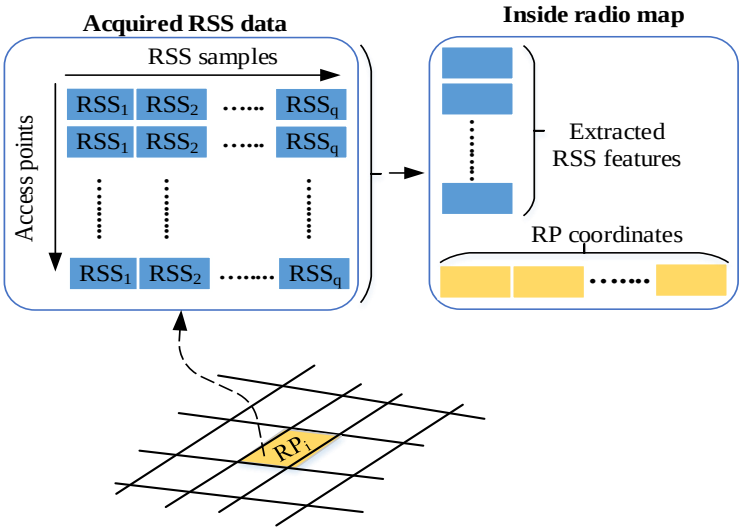
Given the known length between the APs (known location coordinates of the APs) and after estimating the AOA, the location of the tag device can be estimated as follows:

$$\begin{aligned} x &= x_2 + cos(\theta_2)\frac{y_2 - y_1 - tan(\theta_1)(x_2 - x_1)}{cos(\theta_2)tan(\theta_1) - sin(\theta_2)} \\ y &= y_2 + sin(\theta_2)\frac{y_2 - y_1 - tan(\theta_1)(x_2 - x_1)}{cos(\theta_2)tan(\theta_1) - sin(\theta_2)} \end{aligned} \tag{2.7}$$

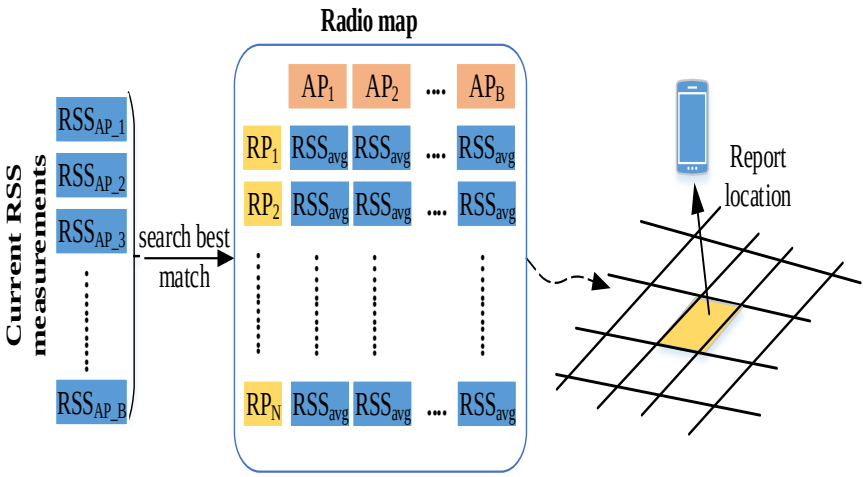
2.4.3 Fingerprinting

Fingerprinting is also called scene analysis where signal strength at RPs are measured and stored in the database along with the location of the coordinate of the RPs. For localization, new signal strength is measured and compared with the saved ones to estimate a location. Hence, a fingerprinting localization has two phases of operations as illustrated by Figure 2.12.

Fingerprinting is the most widely used indoor localization method due to its good localization accuracy and non-requirement of LOS measurements of APs. The technologies like Wi-Fi, BLE, and geomagnetic field can be used to realize the fingerprinting localization. Although fingerprinting localization has good localization accuracy, it comes with time-consuming and labor-intensive



(a)



(b)

Figure 2.12: The working procedure of fingerprinting localization. (a) Offline phase. (b) Online phase.

offline phase. Fingerprinting localization can be categorized as deterministic and probabilistic. The former approach implements fingerprinting data comparison algorithms to find the estimated position, whereas the latter approach yields localization information by estimating a probability distribution over the RPs.

2.4.4 Dead-reckoning

In dead reckoning positioning, localization is estimated based on the last estimated position that is updated by a displacement determined from velocity and heading direction over an elapsed time step. The speed is computed based on accelerometer readings and orientation is determined by the magnetic field and gyroscope measurements.

An illustration of basic dead-reckoning based IPS is shown in Figure 2.13. The time integral of the accelerometer reading yields a continuous estimate of the instantaneous speed of the tag device if the initial speed is known. The displacement with respect to a starting point can be obtained with a second integration as [38]:

$$\vec{p}(t) = \int_0^t \vec{v}(\tau)d(\tau) = \int_0^t \int_0^\tau \vec{a}(x)dxd\tau, \tag{2.8}$$

where $\vec{p}(t)$ is the displacement vector, $\vec{v}(\tau)$ is the instantaneous speed, and $\vec{a}(x)$ is the acceleration from the sensor reading. Similarly, the heading can be obtained by integrating the gyroscope sensor reading as:

$$\theta(t) = \int_0^t w(\tau)d\tau, \tag{2.9}$$

where $\theta(t)$ is the angle of rotation with respect to the previous azimuth angle and $w(\tau)$ is the angular velocity obtained from the sensor reading.

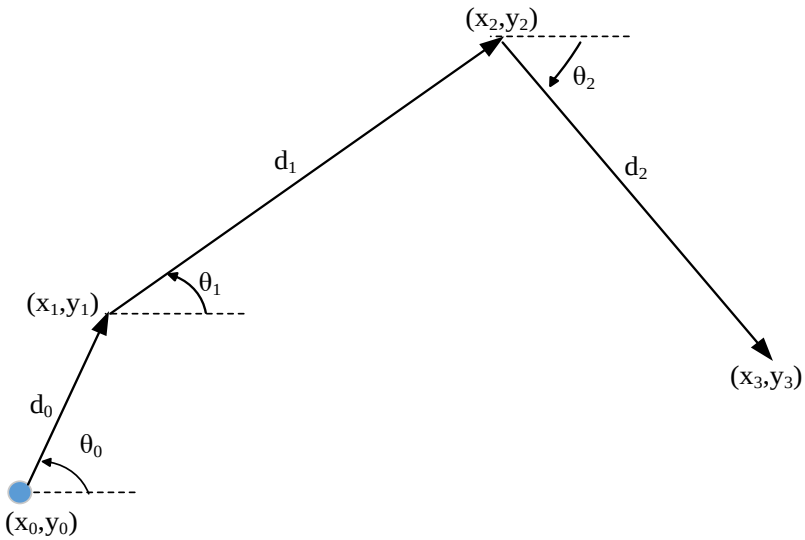


Figure 2.13: Location estimation based on sensor dead reckoning where the blue dot represents the initial point.

In addition, the magnetic field gives an absolute angle value between the tag device heading and the geomagnetic north pole. The gyroscope readings are very sensitive and suffer from severe error accumulation owing to integrating angular velocity over a long time. Since the magnetic field orientation is robust over time but susceptible to interference, the magnetic and gyroscope readings are combined using complementary filters to generate a more accurate azimuth angle estimation.

2.5 The Performance Metrics

Since there are different technologies and methods to realize an IPS, the most important performance metric is the localization accuracy. In addition to localization accuracy, other performance indicators of an IPS are complexity, scalability/ robustness, and cost.

2.5.1 Accuracy and Precision

The localization accuracy can be defined as a difference between an estimated location and the true location of the tag device. Similarly, the precision indicates the degree to which repeated location estimates produce identical results under the unchanged conditions.

Usually, the mean squared error (MSE) is used as the accuracy indicator. Indoor environments are often complex due to the presence of different obstacles and environmental changes, resulting in signal fluctuation. In the meantime, high localization accuracy (meter level) is often expected

for adequate location-based service. The localization accuracy of the IPS depends on the used technology and techniques.

The precision yields information like how convergent the localization result can be over many trials or how consistently the system works. The cumulative distribution function (CDF) is used as a precision indicator. The CDF represents the distribution of the distance error between the estimated location and the true location of the tag device. The MSE as well CDF should be exploited while comparing two or more localization algorithms.

2.5.2 Complexity

For an IPS, its complexity can be categorized in terms of hardware and software. The adopted technology and the signal measurement principle for the IPS accounts for hardware complexity. For example, technologies like Wi-Fi, BLE, and geomagnetic filed is supported by most of the present-day smartphones. However, standard mobile devices do not support technologies like UWB and ultrasound and the IPS using such technologies should use a dedicated system that requires proprietary equipment. Moreover, geomagnetic based IPS can produce localization result without deploying any hardware however, Wi-Fi and BLE need to be deployed.

As for the chosen signal measurement principle, obtaining RSS from Wi-Fi and BLE are relatively easy with standard mobile phone owing to the fact that such devices typically need to scan RSS for their routine functioning. However, it is not easy to obtain accurate time and angle measurements that increase the complexity of the IPS.

Regarding the software, complexity depends on the computation load represented by the calculations required to perform localization. In a server-based IPS, the execution of the localization algorithm is carried out on a centralized server where the positioning could be calculated quickly owing to its powerful processing capability and abundant power supply. However, if the positioning algorithm is executed in the tag device, it may increase the complexity. Moreover, the complexity concerned with the software also depends on the technique used for IPS development. For example, the fingerprinting localization has larger complexity and it grows as the localization environment increases. Here, the complexity can be minimized using clustering.

2.5.3 Scalability and Robustness

Scalability in IPS refers to the ability of the localization system to perform well even when any change on the area of interest for localization and/or on signal source occurs. The changes can be an extension of the localization area and/or extension of signal coverage. If an IPS need not be taken down in such a scenario, it is considered as good scalability. For example, when any IPS is constructed, it provides services in a limited area of interest and an increase in the localization area might be needed in some future time. In addition, in some cases, the transmitting power and

signal-broadcasting rate can be increased for good signal coverage. In such situations, the positioning techniques like proximity, WCL, and trilateration are easy to expand by simply adding the identical signal sources and updating the system with the location coordinate information of the added hardware. However, the fingerprinting based system needs an offline site survey for every change in the localization area or signal source. When the localization area is expanded, the radio map for the extended area needs to be freshly constructed. Moreover, when the transmitting power at the APs or signal broadcasting rate (e.g. advertisement packet broadcasting interval in BLE) is changed, a new site survey for the whole localization area is required. Hence, fingerprinting localization has relatively low scalability.

Robustness is also an important factor in IPS that allows the system to function normally without human intervention when the localization environment changes. For example, some signal sources could be out of service occasionally or testbed layout changes could cause some signal to no longer support LOS propagation. In this scenario, the IPS has to provide localization service with incomplete or noisy information (RSS fluctuation). The robustness can be gained by introducing redundant information into a localization estimation. For example, rather than using the only three APs for trilateration, an IPS can include many supplementary APs to make the system more robust. Moreover, for fingerprinting based IPS can adopt a larger set of RSS samples to increase robustness.

2.5.4 Cost

The cost factor in IPS represents not only the infrastructure cost but also the time and effort for system installation and maintenance procedures. Particularly, the cost depends on factors like the size of the localization area, the required accuracy, the used technology, power consumption, etc. The IPS system such as PDR, geomagnetic-based, and barometer-based can be realized with the smartphone only and do not need any additional infrastructure. Some signal sources like Wi-Fi and BLE require APs deployment however; Wi-Fi is already deployed for other purposes. Hence, if the IPS is based on without any additional infrastructure or on existing infrastructure, the cost can be substantially saved.

Moreover, if an IPS is robust and scalable, saving in time and labor is possible, which helps to reduce the cost of the system. In addition, an increase in localization space requires the deployment of more APs in localization technique like proximity. Power consumption at both the AP and the tag device is also a critical cost issue. For example, BLE consumes less power compared to Wi-Fi where BLE operates with a coin-shaped battery but Wi-Fi needs to be plugged into mains. Furthermore, when the localization operation is carried out on the server side, the power consumption of the tag device can be reduced. Lastly, the location update rates, signal broadcasting rate, and the desired system accuracy can also affect power consumption.

2.6 Concluding Remarks

This chapter reviews the fundamental aspects of IPS development. Particularly, the radio frequency based wireless technologies are emphasized and the signal measurement principles are elaborated. The basic algorithms in IPS are discussed along with the performance metrics of the IPS techniques.

Various technologies have been thoroughly investigated to develop IPS. The RF technologies mainly Wi-Fi and BLE are thriving well in IPS study due to their characteristics like signal penetration, power consumption, localization accuracy, convenient deployment of APs, and compatible with modern smartphones.

The RSS is widely adopted while developing IPS owing to ease of its measurement and non-requirement of extra hardware. RSS information can be used as radio signature in fingerprinting localization whereas it can be converted to distance for lateral-based localizations. Fingerprinting is the most extensively used localization technique owing to its good localization estimation accuracy. However, it has drawbacks like tedious and time-consuming radio-map construction phase.

Chapter 3

AP Signal Characteristics and Filtration

3.1 Introduction

In this chapter, characteristics of received signal strength (RSS) values of BLE beacons are investigated that are reported by the smartphone tag device. As Wi-Fi is also used widely for IPS development, we compare the characteristics of RSS from Wi-Fi and BLE. At first, the properties of RSS values from Wi-Fi and BLE APs are analyzed. This analysis of the RSS values is required to understand the underlying features of RSS patterns. Wi-Fi and BLE are compared in terms of IPS development. Second, different RSS filtration or smoothing techniques used for IPS development are presented. RSS smoothing helps to minimize the localization error induced by RSS fluctuations.

3.2 RSS characteristics

Since RSS is used as a signal measurement principle in this thesis, the signal characteristics of BLE are estimated through real field deployments. Moreover, Wi-Fi has been the competitor wireless technology of BLE in IPS development. Hence, the RSS from BLE and Wi-Fi are compared in an identical environment and the obtained results are analyzed.

A BLE beacon was deployed together along with the Wi-Fi AP where a Wi-Fi extender was used in AP mode as shown in Figure 3.1. The advertising interval of the BLE beacon was set to 300 ms and the transmitting power as $+4$ dBm. At first, the streaming RSS data are recorded from BLE and Wi-Fi APs in the corridor (2.6 meters wide). Note that a Wi-Fi module takes $3-4$ seconds to process the start scan command from the positioning application to acquire new scan results [39]. The acquired RSS data from the APs by placing the tag device at a fixed position is presented in Figure 3.2.

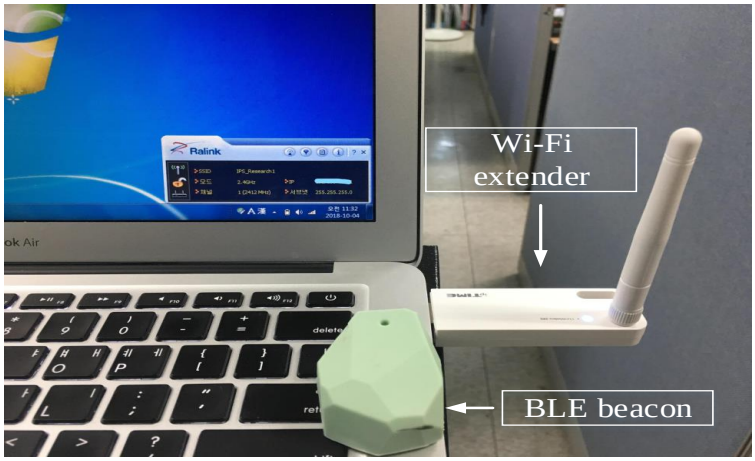


Figure 3.1: Pictorial representation of the APs (Wi-Fi and BLE).

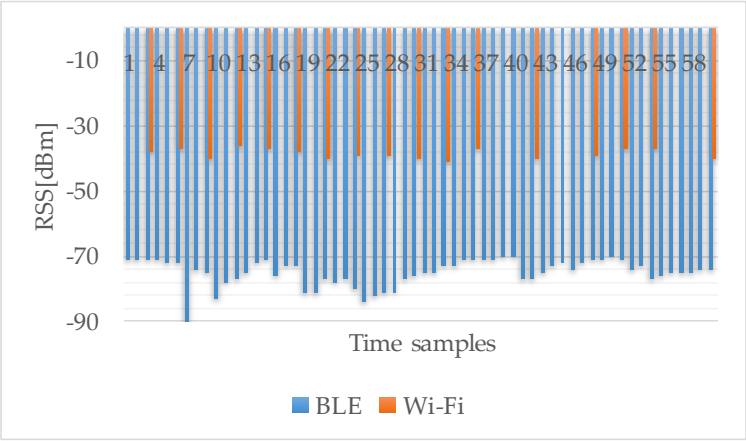
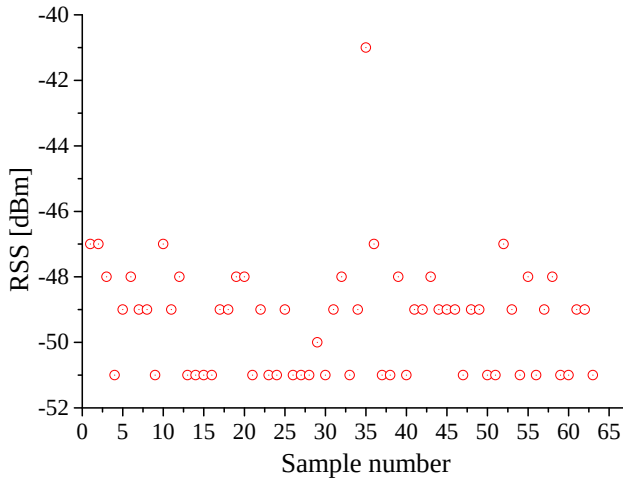


Figure 3.2: Detected RSS sequences from BLE and Wi-Fi at a specific sampling point on the testbed.

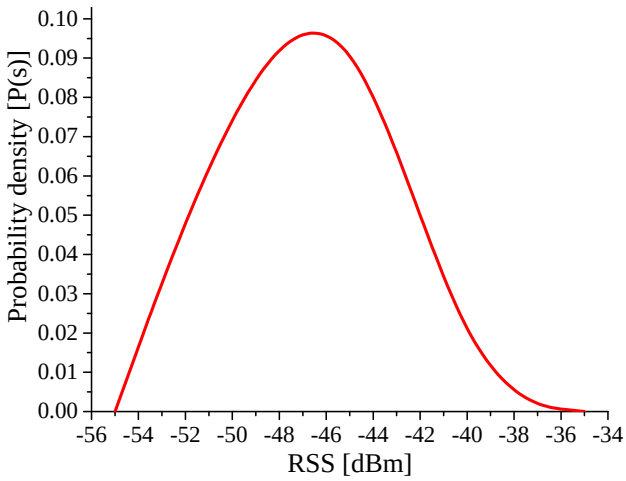
As seen in Figure 3.2, sixty RSS data were acquired in one minute (one sample per second) from the BLE beacon, whereas only 17 RSS data were acquired from the Wi-Fi AP for same time duration. In practice, repeated RSS data are acquired until new scan results are produced to perform sampling per second in Wi-Fi AP. In such a case, subsequent filtering must be performed on the obtained Wi-Fi data to remove the erroneous statistical analysis. Introduction of filtering also helps to remove the spikes of BLE data to produce smooth RSS. Different RSS filtering

approaches will be discussed in the next section of this chapter.

As it has been mentioned earlier that RSS fluctuates with time owing to the complex indoor environment, more RSS samples are further acquired from the BLE and Wi-Fi APs to analyze the distribution behavior of RSS. The plot of RSS samples and the resultant probability density plot are shown in Figure 3.3 and 3.4 respectively for BLE and Wi-Fi.

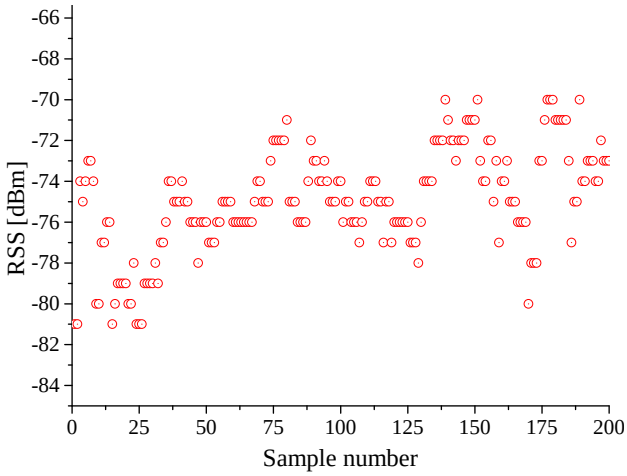


(a)

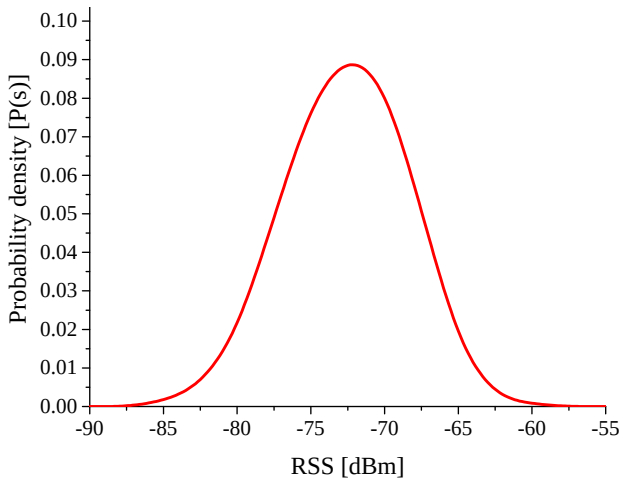


(b)

Figure 3.3: Distribution of Wi-Fi RSS at a certain point. (a) Fluctuation of RSS values. (b) Probability density function.



(a)



(b)

Figure 3.4: Distribution of BLE RSS at a certain point. (a) Fluctuation of RSS values. (b) Probability density function.

From Figure 3.3 and 3.4, it is seen that RSS fluctuates a lot even at a fixed point in the indoor environment. Moreover, it is also seen that the statistical distribution of RSS at a particular point can be thought of as a Gaussian distribution [31, 40]. Hence, the Gaussian filter can be employed for estimating the value of RSS in IPS.

Furthermore, the fingerprinting localization needs a spatially diverse radio signature data. In other words, the RSS from an AP should change notably with a change in distance or should have a good resolution. To examine this property of RSS, the average RSS from BLE and Wi-Fi were acquired at different distances from the fixed APs. The obtained result is presented in Figure 3.5.

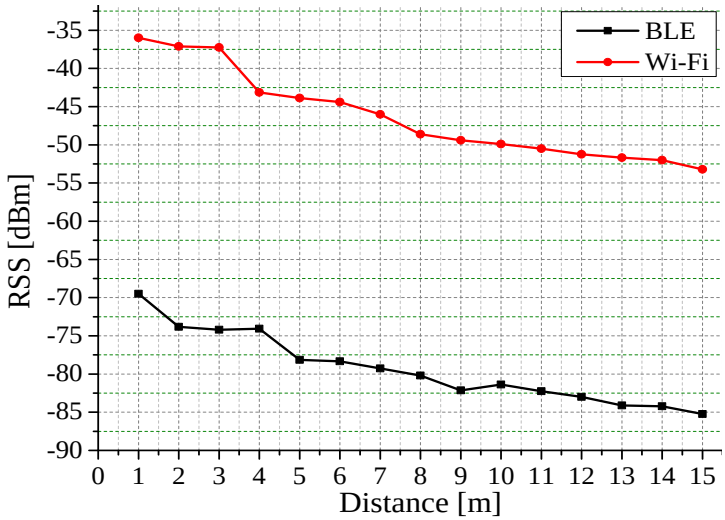


Figure 3.5: Estimated average RSS data from Wi-Fi and BLE at various distances.

Here, the range of average RSS data from Wi-Fi is -36.0 dBm to -53.2 dBm for one to fifteen meters distance. Similarly, from BLE the average RSS data ranges from -69.51 dBm to -85.24 dBm for the same distance. It shows that the resolution of BLE is better than Wi-Fi for indoor environment however, the transmission range of Wi-Fi is a lot more compared to BLE.

Finally, the distance from received RSS at the tag device is estimated. Note that accurate distance estimation is crucial in lateral based localization systems such as trilateration. A well-known log-distance path loss model for indoor radio signal propagation is used to convert the received RSS data into the distance given as [41, 42, 43]:

$$P_r(d) = A - 10 \times n \times log_{10}\left(\frac{d}{d_0}\right) + \chi_\sigma, \tag{3.1}$$

where $P_r(d)$ is the received RSS in dBm at distance d, A is the received RSS at a standard distance d_0 (defined as 1 meter), n is the path loss exponent, and χ_σ is zero mean Gaussian-distribution random variable with variance σ^2. From (3.1), we can derive:

$$d = d_0 \times 10^{\frac{A - P_r(d)}{10 \times n}} \tag{3.2}$$

In (3.2), A and n need to be calibrated for every signal attenuation environment before converting the RSS to distance. The APs (Wi-Fi and BLE as shown in Figure 3.1) were placed at the center of testbed and average RSS data (35 time samples at each data-sampling place) were recorded for each AP at a distance (APs to the tag device) of one meter in eight directions ($45°$ apart). The observed data is presented in Table 3.1.

Table 3.1: Wi-Fi and BLE RSS estimation at unit distance.

Directions	Wi-Fi [-dBm]	BLE [-dBm]
0°	-36.92	-73.61
45°	-44.1	-69.60
90°	-34.64	-63.90
135°	-31.64	-68.93
180°	-36.09	-64.10
225°	-35.3	-67.36
270°	-38.7	-67.47
315°	-41.0	-74.93

Similarly, n was estimated at various distances (2 to 25 meters) from the APs on the testbed as shown in Table 3.2. The final value of A and n were calibrated using linear regression [44].

Table 3.2: Path loss exponent estimation.

Distance [m]	Wi-Fi	BLE
2	1.32	1.77
4	1.51	2.20
7	1.23	1.53
9	1.29	1.28
12	1.12	1.18
14	1.71	1.23
15	1.20	1.04
17	1.40	1.38
20	1.36	1.36
27	1.43	1.35

The final calibrated values of A and n were $(-36.02, 1.32)$ and $(-66.42, 1.785)$ for Wi-Fi and BLE, respectively. Note that, n was estimated using (3.1) and calibrated value of A.

After calibrating A and n on the testbed, the distance was estimated employing RSS from Wi-Fi and BLE APs. In IPS, a reliable distance estimation is performed by utilizing the latest set of RSS reception using filtration (generally averaging or low pass filter). In addition, the filtration can be done in two ways: RSS domain and distance domain. The latest set of acquired RSS data is averaged and converted to distance in RSS domain whereas each received RSS sample is converted to corresponding distance and the set of estimated distance samples are averaged in the distance domain. Note that due to the non-linear relation of distance with RSS, averaging the RSS

and converting it to distance can add some error to the distance estimation.

The estimated distances (distance domain) along with the respective real-distance is presented in Figure 3.6. Here, a set of 35 RSS samples was recorded at every measurement place. The average error of distance estimation for Wi-Fi and BLE are 1.765m and 1.7005m, respectively.

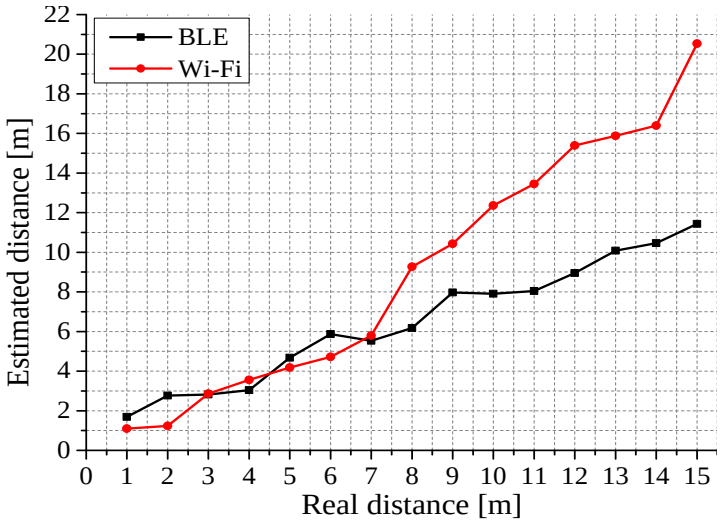


Figure 3.6: Estimated distance from acquired RSS data from Wi-Fi and BLE at various distances in the testbed.

Table 3.3 summarizes the Wi-Fi and BLE technologies in terms of IPS development.

Table 3.3: Comparison of Wi-Fi and BLE technology in terms of IPS.

Parameters	Wi-Fi	BLE
Deployment cost	Low	High
AP reliability	Not-dedicated to IPS	Dedicated to IPS
Hardware efficiency	Requires >=3 seconds to scan new RSS data	RSS sample acquired every second
AP differentiating parameters	SSID, BSSID (MAC)	UUID, MAC
Transmission range	High (∼50 meters)	Low (∼30 meters)
Power consumption	High	Low
Channel availability	Three independent channels at most (2.4 GHz band) [39]	Three advertisement channels [45]
Proximity detection	Normally final location estimation is available	Immediate, Near, and Far proximity available

3.3 RSS Filtration

It is seen from Figure 3.2, 3.3, and 3.4 that RSS fluctuates over time even in a fixed indoor environment. This instability of RSS results in increased localization estimation error. The reasons behind RSS exhibiting such high variability in space and time are various noise factors, interferences, and attenuation. Hence, irrespective to the technique used for localization, RSS filtration or smoothing is required to minimize the localization estimation error. Some of the RSS filtering techniques are averaging/ moving average filter, Kalman filter [46, 47, 48], Gaussian filter, and exponential averaging.

3.3.1 Averaging/Moving Average Filter

The time average of RSS data is generally considered for the stability of randomly fluctuating RSS data in IPS development. By intuition, the averaging operation can also be seen as a low pass

filter where it tries to minimize the rapid changes or fluctuation on RSS values. The averaging filter is mostly used in training phase of the fingerprinting localization where at a certain RP, the time average of many RSS samples from an AP is considered the radio signature of that RP. For example, if a testbed is divided into N uniform girds, the RSS data of B APs are acquired at r^{th} RP ($r \in \{1, 2, \ldots, N\}$) q times, i.e., $AP_r = \{ap_{r,1}, ap_{r,2}, ..., ap_{r,B}\}$, where $ap_{r,i} = Avg(ap^1_{r,i}, ap^2_{r,i}, ..., ap^q_{r,i})$.

In moving average (MA) filter, a set of last RSS samples are averaged to get a smooth RSS value. Here, the set is determined by a window size that moves along with the new sample of RSS data acquisition. The moving average filter is generally used in the testing phase of fingerprinting localization and lateral-based localization methods. The moving average filter can be further upgraded to a weighted moving average (WMA) filter where the elements of the sets are assigned a certain weight. The performance of MA and WMA in the testbed environment is compared. The weighting factors and the windows sizes are listed in Table 3.4.

Table 3.4: Weighting factors and window sizes for the experiment.

Window size	**Set of weighting factors**
4	[0.4, 0.3, 0.2, 0.1]
5	[0.3, 0.2, 0.2, 0.2, 0.1]

MA and WMA were used to measure 100 samples of smoothed values at the hallways and the computer lab. For the experiment, the tag device was held at a fixed distance to the AP. The standard deviation (SD) of filtered RSS samples were calculated as a measure of smoothness of the filters used. The observed values are presented in Table 3.5 where (W)MA-4 and (W)MA-5 denote the (W)MA filters with windows sizes of 4 and 5, respectively.

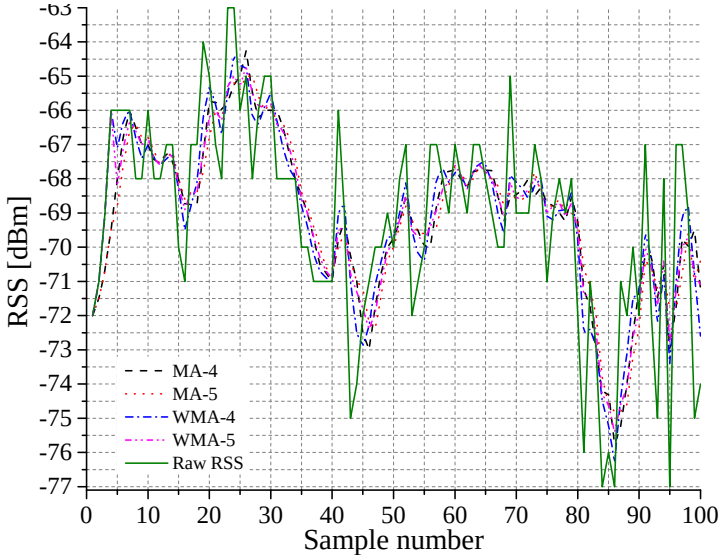
Table 3.5: Standard deviations of RSS samples from various filters in different testbeds.

Filters	**Computer lab**	**Corridor**
Raw RSS	3.025214	1.710898
MA-4	2.335847	1.307184
MA-5	2.245309	1.248511
WMA-4	2.428521	1.397725
WMA-5	2.2891	1.314257

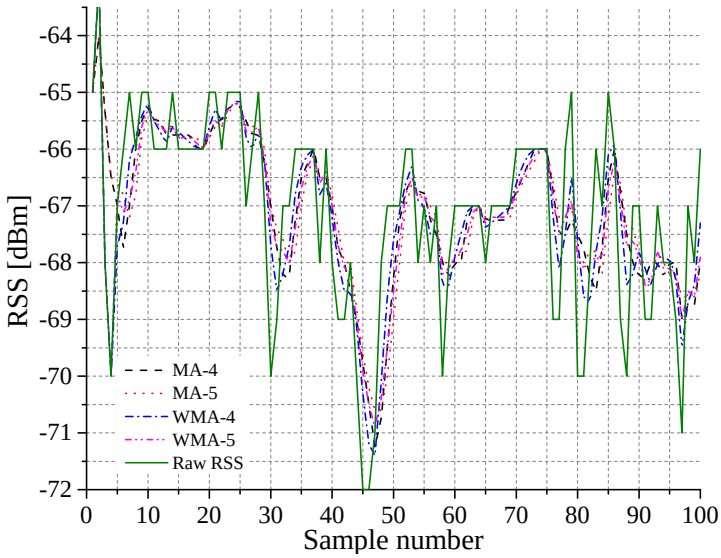
As seen in Table 3.5, a smooth result is delivered by the (W)MA filter with a window size of 5. The smoothness can further increase with an increase in the size of window size, however, it will be less responsive towards the recent RSS observation.

Figure 3.7 shows the graphical representation of the observed RSS with the (W)MA filters. It

can be observed in Figure 3.7 that the smoothing operation has reduced the spikes of raw RSS.



(a)



(b)

Figure 3.7: The plot of smoothed RSS from (W) MA filters in different testbeds. (a) Computer lab. (b) Corridor.

3.3.2 Kalman Filter

A Kalman filter uses a series of measurements observed over time that contains statistical noise to produce estimates of unknown variables that tend to more accurate than those based on single measurement. The regular Kalman filter has been employed to filter the noisy RSS data in the variously published literature [49, 50]. Since the Kalman filter is designed for a linear system, the tag device is assumed static to simplify the filter. A simplified Kalman filter equation is shown as:

- Time update equations:

$$\hat{x}_k^- = \hat{x}_{k-1}^- \tag{3.3}$$

$$P_k^- = P_{k-1} + Q \tag{3.4}$$

- Measurement update equations:

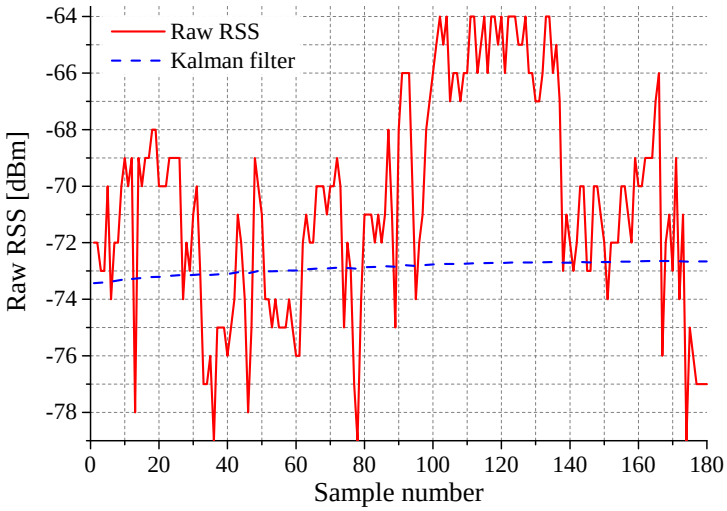
$$K_k = \frac{P_k^-}{P_k^- + R} \tag{3.5}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(Z_k - \hat{x}_k^-) \tag{3.6}$$

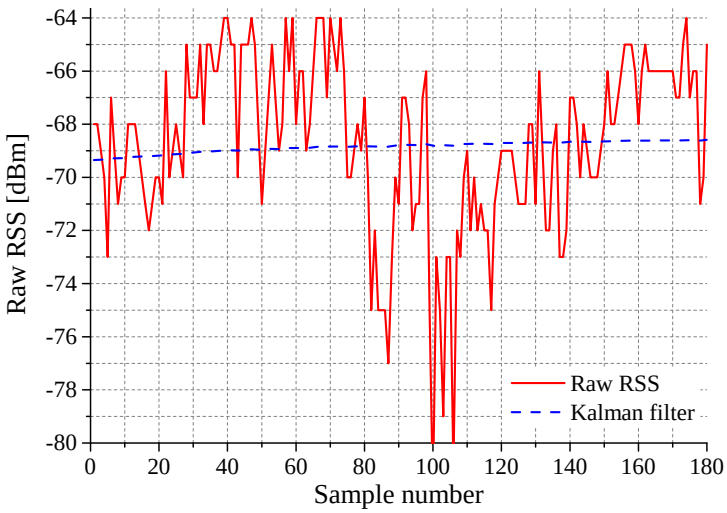
$$P_k = (1 - K_k)P_k^- \tag{3.7}$$

Here, $\hat{x}_k^-$ is priori estimate and $\hat{x}_{k-1}^-$ is posteriori state estimate. Similarly, Q is process variance, R is the measurement variance, and P_k^- and P_k are priori and posteriori error variances. K_k is Kalman gain at time instant k and Z_k is the observed raw RSS.

A very small value is set to Q as 0.000001 whereas R and P are set as 0.1 and 0.001, respectively. The observed filtered result is presented in Figure 3.8.



(a)



(b)

Figure 3.8: The plot of smoothed RSS from Kalman filter in different testbeds. (a) Computer lab. (b) Corridor.

3.3.3 Gaussian Filter

As seen in Figure 3.3 and 3.4, the statistical distribution of RSS can be thought of as a Gaussian distribution. Therefore, as the Gaussian distribution can represent the randomness of RSS in a real environment, Gaussian filter can be employed for estimating the value of RSS while constructing the radio map in fingerprinting localization.

The probability density function is formulated as:

$$f(RSS) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(RSS-\mu)^2}{\sigma^2}} \tag{3.8}$$

where μ and σ^2 are mean and variance of the Gaussian filter given as:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} RSS_n \tag{3.9}$$

$$\sigma^2 = \frac{1}{N-1}\sum_{n=1}^{N} (RSS_n - \mu)^2 \tag{3.10}$$

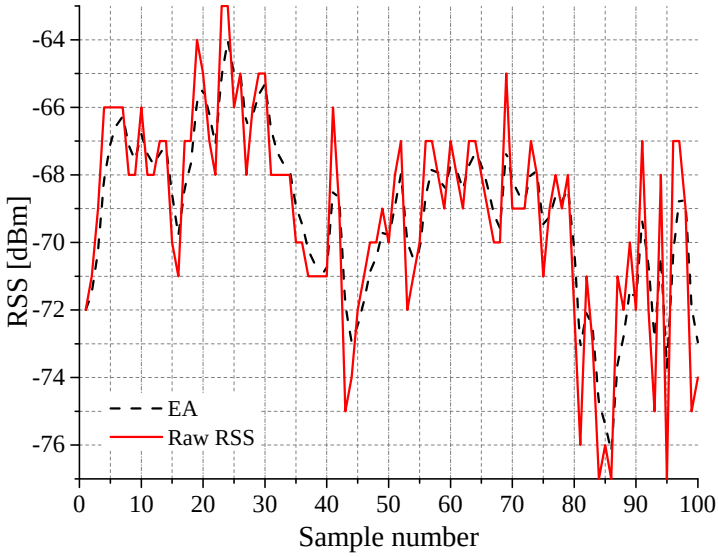
Particularly, the centralized RSS values that lie in the effective range of $\mu - \sigma$ and $\mu + \sigma$ are accepted as useful RSS data. The effective range consists of 68.20% of the total data, which is then averaged to get a estimated RSS value.

3.3.4 Exponential averaging (EA) filter

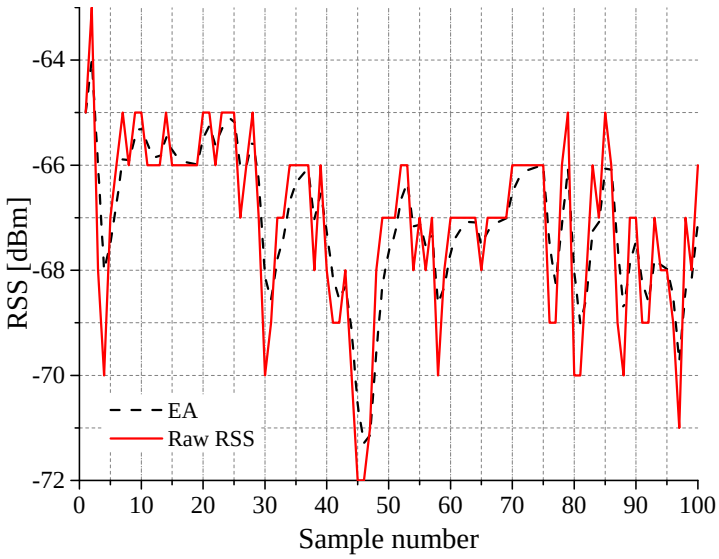
The exponential averaging can be termed as a formalization of the simple learning process as it yields the smoothed RSS as soon as two observations are available [32]. In general, EA filter can be used in the testing phase of fingerprinting localization or in lateral-based localization methods. The EA for an RSS data sample at an instant t (RSS_{EA_t}) is given by the following relation:

$$RSS_{EA_t} = \beta \times RSS_t + (1-\beta) \times RSS_{EA_{t-1}}, \tag{3.11}$$

where β is the smoothing factor whose value ranges from $0 < \beta < 1$. Here, the larger value of β reduces the level of smoothing, whereas its value close to zero has greater smoothing effect that is less responsive towards the recent RSS observations. Note that a moving average filter hold for window-sized RSS observation whereas EA produces smoothing result with only the first two RSS observations. Figure 3.9 illustrates the effectiveness of EA (with $\beta = 0.5$ to give equal preference to previous and current RSS samples) filter compared to the raw RSS.



(a)



(b)

Figure 3.9: The plot of smoothed RSS from EA filter in different testbeds. (a) Computer lab. (b) Corridor.

As seen in Figure 3.9, the use of EA filter helps to reduce the sudden changes of raw RSS value.

3.4 Concluding Remarks

This chapter focuses on RSS data acquisition and estimation. Particularly, Wi-Fi and BLE signals are used to understand the RSS characteristics. The statistical distribution of RSS data from Wi-Fi and BLE suggests that RSS follows a Gaussian distribution. The parameters of the log-distance path loss model for the indoor environment are calibrated for both the wireless technologies.

For RSS estimation, different RSS filtering algorithms are introduced where observed RSS estimation result of BLE signal is reported. The filtering techniques like an averaging filter and Gaussian filter can be employed in the training or offline phase of the fingerprinting localization. Similarly, Kalman filter, EA filter, and moving average filter can be utilized for both the test phase of fingerprinting localization and lateral-based localization techniques.

Chapter 4

Improvisation Over Conventional Fingerprinting Localization

In this chapter, two different novel fingerprinting approaches are proposed to improvise the conventional fingerprinting localization. Particularly, the suggested approaches address the issues in the offline and online phase of weighted k-nearest neighbor (Wk-NN) fingerprinting localization. The first method helps to minimize the offline workload of Wk-NN fingerprinting localization by combining it with WC localization. The next suggested method makes use of diverse fingerprinting features and clustering to increase localization estimation accuracy and to reduce computational cost. The experimental results obtained by real field deployment are presented to show the superiority of the proposed methods over the conventional counterpart.

4.1 Process Flow of Wk-NN Fingerprinting Localization

The Wk-NN fingerprinting is the deterministic-based fingerprinting localization method and has two phases of operation. The first is the training or offline phase where radio map is constructed whereas the second is the testing or the online localization phase. In Wk-NN, k RPs are selected based on the similarity of their fingerprints with the measured RSS [51]. The location is then estimated as a weighted sum of the real locations of the k RPs. A basic process flow of Wk-NN is illustrated in Figure 4.1.

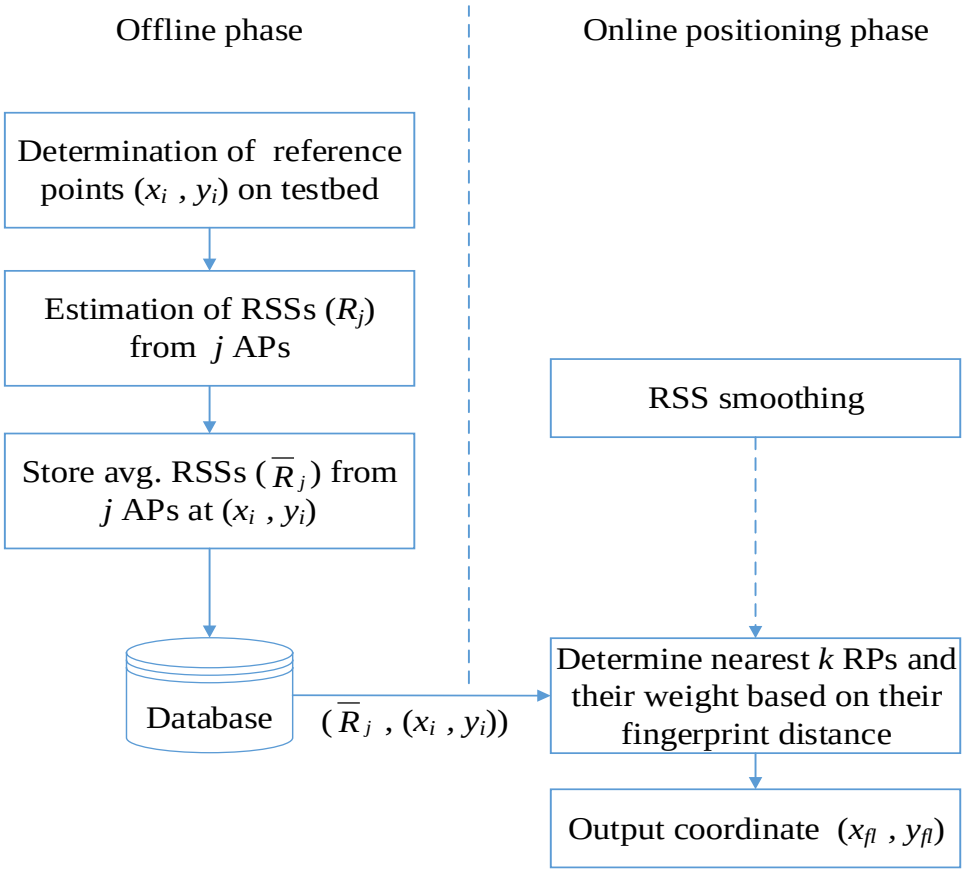


Figure 4.1: Process flow of Wk-NN fingerprinting localization.

4.1.1 Offline or Training Phase

The offline phase is designed for RSS acquisition at each RPs across the testbed. At this stage, RSS from all the APs is collected at each RP and stored in the database for future reference. The offline phase is illustrated on the left side of Figure 4.1. In particular, the database contains coordinates of the RPs and the average RSS from each AP that is measured individually at all RPs. At an RP, RSS can be collected in four directions ($0°, 90°, 180°, 270°$) for better localization accuracy as shown in Figure 4.2.

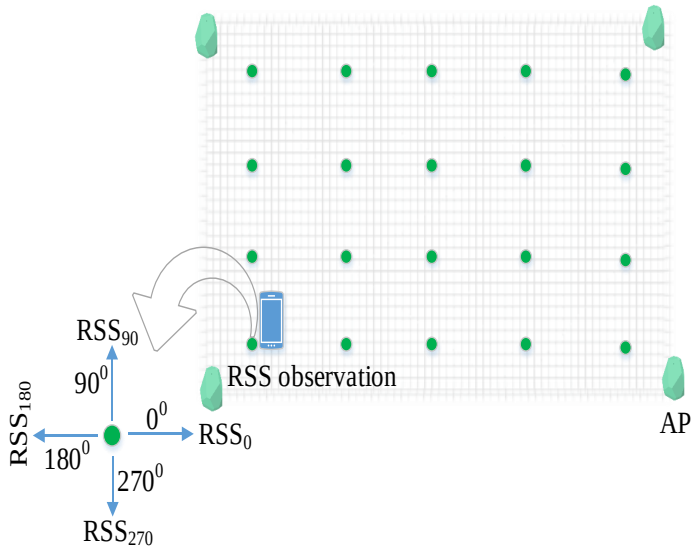


Figure 4.2: The offline phase of fingerprinting localization. The green dots are the reference points, RSS_0 represents RSS value at $0°$ to the base direction.

4.1.2 Online or Execution Phase

On online or execution phase of typical fingerprinting localization, the observed RSS is compared against the stored one in the database, and then the coordinate with the closest match is considered as the estimated tag device location. However, in Wk-NN, nearest k RPs are selected and the location is estimated as a weighted sum of their real location. The positioning distance (D_j) between the stored RSS and the online observed RSS at j^{th} RP is given by:

$$D_j = \sum_{i=1}^{B} \sqrt{(RSS_{i_{online}} - RSS_{i_{offline}})^2} \; j = 1, 2, ..., N, \tag{4.1}$$

where i is the number of APs ranging from 1 to B. The RPs are arranged with ascending order of D_j and the first k RPs with their known positions $J_z[x_z, y_z]$ are selected to estimate the final location (T_{Wk-NN}) using the following relation:

$$T_{Wk-NN} = \frac{\sum_{z=1}^{k} J_z \times W_z}{\sum_{z=1}^{k} W_z}, \; where \; W_z = \frac{1}{D_z} \tag{4.2}$$

4.2 Related Works

The development in the fields of modern smartphones and wireless communication technologies such as Wi-Fi, UWB, and BLE have made it possible to implement IPS with a few meters of accuracy. Mostly, fingerprinting is adopted for IPS development to achieve the required accuracy.

The nearest neighbor fingerprinting is the primitive kind of fingerprinting localization, which estimates the location of the tag device as the location coordinate of the RP that has the closest fingerprinting resemblance with the observed RSS vector. This approach of localization is improvised by considering the nearest k RPs where a certain weight is allocated to each selected RP to yield better localization estimation. Variously published literature have put forward different localization approaches to improvise the conventional fingerprinting localization.

4.2.1 Wk-NN Fingerprinting Approaches

Pavel et al. present an IPS research work based on Wk-NN positioning method using BLE beacons [52]. The k-nearest fingerprints are found in a radio map database by employing the Euclidean distance between the observed RSS and the referred one from the database. This work further compares the localization methods based on Wi-Fi and a combination of BLE and Wi-Fi. They recommend that the combination of wireless technologies help to increase the localization accuracy. Next work based on BLE beacons using fingerprinting technique is reported in [31] where a Gaussian filter is used to preprocess the received RSS. This work proposes a distance-weighted filter based on the triangle theorem of trilateral relations to filter out the wrong distance value caused by an abnormal RSS.

The traditional Wk-NN fingerprinting has also been realized with Wi-Fi signals. Reference [53] elaborates recent advances on Wi-Fi fingerprinting localization. They overview on advanced localization techniques and efficient system development utilizing Wi-Fi technology in their survey work. An improvisation over the conventional Wk-NN fingerprinting using Wi-Fi signals is put forward in [54] and [55]. The former approach uses average RSS and standard deviation of Wi-Fi signals at the RPs from the APs to construct a fingerprint radio map. Both the average RSS and the standard deviation is processed to estimate a Euclidean distance on the online phase. With the Euclidean distance, k RPs are selected to estimate a coarse location. Furthermore, a joint probability for each RP is calculated, based on which the k RPs are selected to estimate another coarse location. Later, both the coarse localization estimations are fused together employing a shortest Euclidean distance and largest joint probability to yield a final localization estimation. Meanwhile, the later approach proposes to use Manhattan distance instead of Euclidean distance to compare the closeness of acquired Wi-Fi signal strength with the stored database.

4.2.2 Hybrid Approaches

The RSS-based fingerprinting localization has been integrated with other techniques and technologies for better localization performance. Reference [56] integrates the fingerprinting with motion sensor-based positioning where the latest position of the tag device is estimated by adding the previous position estimate and the position displacement from sensor-based positioning with the help of Kalman filter. Similarly, [57] suggests an approach of utilizing a user movement pattern and feeding the information to RSS-based localization system. This work exploits the recurrent neural networks to process RSS information to predict the user movement pattern.

The deterministic-based localization system can be enhanced by utilizing the true position coordinate of the AP. Reference [58] uses pedestrian dead reckoning (PDR) together with BLE beacons to estimate the user location. This work employs the proximity information of beacons for correction of the estimated position of PDR. An approach in IPS utilizing the estimated distance between the AP and the tag device along with the true location of the AP is put forward in [41]. Here, the acquired RSS is converted to distance using a propagation model where the estimate distance and the AP's location is sent to the server through a cellular or Wi-Fi network. At the server, the received data is fed to an extended Kalman filter to estimate the final location.

4.2.3 Clustering Based Approaches

The performance of indoor fingerprinting positioning can be improved with RSS clustering [59]. An RSS clustering method chooses a set of cluster centers to reduce the sum of squared distances between the RSS value and their corresponding centers. For example, a K-means clustering [60] begins by choosing both the number of output clusters and the corresponding set of initial cluster heads, where the clustering algorithm iteratively refines the output clusters to decrease the sum of squared distances [61]. Hence, K-means clustering has a requirement of an arbitrary selection of initial cluster centers whereas APs starts by assigning each RP the same chance to become a cluster center. Reference [62] uses APC for clustering the testbed using Wi-Fi RSS data. Here, the cluster-head is determined on the coarse localization and Wk-NN is used for fine localization. In addition to APC and K-means, other clustering methods in IPS include fuzzy c-means and hierarchical clustering strategy (HCS) [63, 64, 65, 66].

4.3 Proposed Method-I

In this approach, an improvement over the conventional WK-NN fingerprinting localization is proposed by combining it with the WC localization. The suggested method uses the WC localization in a pipeline in two steps of operations. In the first step, WC estimation and traditional fingerprinting are realized individually where WC utilizes APs' real position information and

their estimated distance to the tag device and fingerprinting (FP_{light}) utilizes the lightly populated RS. The first step yields a WC estimation and k nearest RPs, which are used in the second step for the next WC estimation to estimate the final location of the tag device. The proposed localization method reduces the total number of RPs over the localization surface that minimizes the time required for reading radio frequency signals to construct the radio map. The radio map data formed by a site survey in a Wk-NN fingerprinting technique is as follows:

$$Radio\ map = \begin{bmatrix} RSS_1(RP_1) & RSS_2(RP_1) & \dots & RSS_B(RP_1) \\ RSS_1(RP_2) & RSS_2(RP_2) & \dots & RSS_B(RP_2) \\ \vdots & \vdots & \ddots & \vdots \\ RSS_1(RP_N) & RSS_2(RP_N) & \dots & RSS_B(RP_N) \end{bmatrix} \tag{4.3}$$

where $RSS_1(RP_1)$ is the RSS from an AP at RP_1 with B and N being the total number of deployed APs and total RPs on the testbed, respectively. Depending on the size of B and N, the size of the radio map dataset varies accordingly. The proposed method tries to minimize N, keeping the localization accuracy similar to the existing system.

4.3.1 Process Flow of the Proposed Positioning Approach

RSS Filtration

RSS exhibits high variability in space and time owing to several noise factors and attenuation. Therefore, it is important to estimate the correct RSS value that eventually reduces the localization estimation error. In this approach of IPS, the Gaussian filter is used to estimate RSS values in the training phase and the moving average filter is used for smoothing the real-time RSS in the online phase. The discussion on Gaussian and moving average filter are elaborated in Section 3.3 of Chapter 3.

Working Procedure

Similar to the conventional fingerprinting localization, the proposed approach has two phases of operation. On the offline phase, estimated RSS values from the deployed APs are collected at each RPs as shown in Figure 4.2. Note that the total number of RPs for the proposed method is lesser compare to RPs required for Wk-NN fingerprinting. The total number of RPs and the space between them in the proposed method is strategically chosen as illustrated in the experimental setup section of this chapter. The tag device was kept in the $Messaging/texting$ position in both the training and execution phase. The working procedure of the suggested method is illustrated in Figure 4.3.

At first, the acquired raw RSS is smoothed with moving average filter (the window size for this experiment is 10) to minimize the sudden changes in RSS value. The APs are then arranged

in descending order according to their estimated RSS value. Among these APs, first, m APs are selected for the first WC localization operation (WC_1). The distance between the tag device and the AP is estimated using (3.2) to estimate WC_1 (x_w, y_w) given by (2.3). Simultaneously, FP_{light} is realized to estimate $k - NN$ RPs (x_{fl}, y_{fl}), where $l \in 1, 2, \ldots, k$.

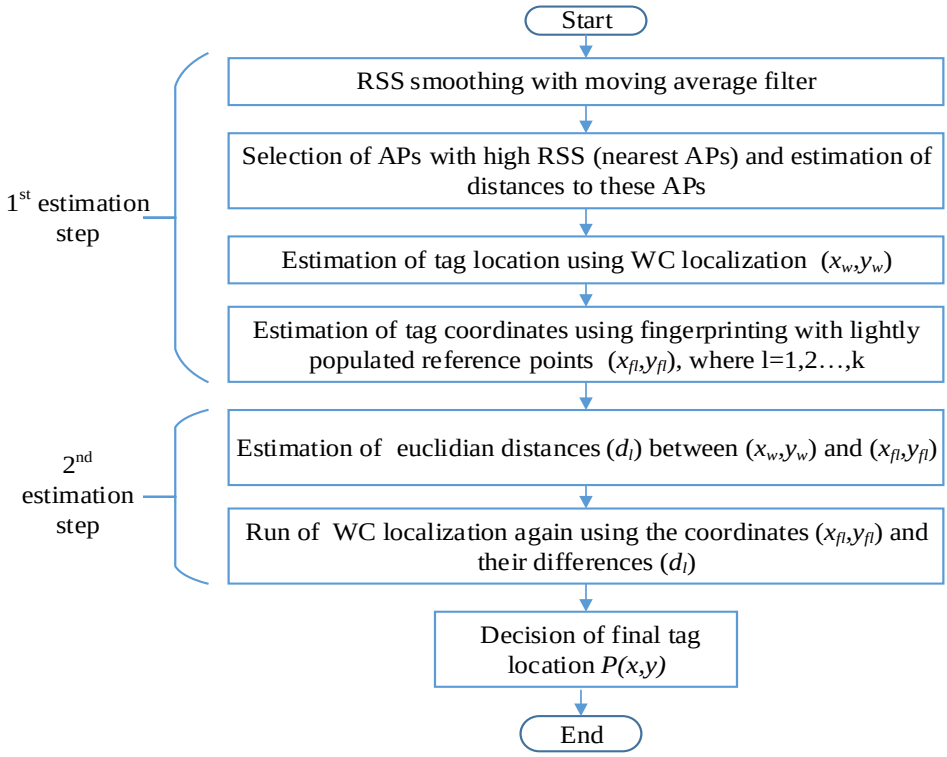


Figure 4.3: Working procedure for the proposed fingerprinting localization.

In the second estimation step, Euclidean distance (d_l) between the $k - NN$ RPs and (x_w, y_w) is estimated as below:

$$d_l = \sqrt{(x_w - x_{fl})^2 + (y_w - y_{fl})^2}, \tag{4.4}$$

where (x_w, y_w) and (x_{fl}, y_{fl}) denote coordinates obtained from WC_1 and FP_{light}, respectively.

For example, let the estimation locations by WC_1 and FP_{light} $(k = 3)$ be (x_w, y_w), (x_{f1}, y_{f1}), (x_{f2}, x_{f2}), and (x_{f3}, x_{f3}), respectively as presented in Figure 4.4. Now, in the second estimation step, the three calculated distances (marked by the dotted red arrow in Figure 4.4) are converted to their respective weights (w_1, w_2, and w_3) using degree (g).

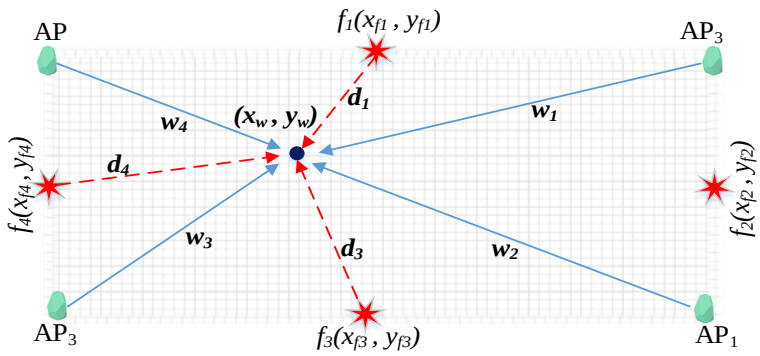


Figure 4.4: Localization estimation of the proposed method.

The estimated weights and the location coordinates of the k RPs are further utilized for WC_2 operation to yield the final location estimation of the proposed localization method as shown in (4.5).

$$P(x,y) = \left(\frac{\sum_{l=1}^{k} x_{fl} \times w_l}{\sum_{l=1}^{k} w_l}, \frac{\sum_{l=1}^{k} y_{fl} \times w_l}{\sum_{l=1}^{k} w_l} \right), \; w_l = \frac{1}{d_l^g} \tag{4.5}$$

The overall process flow in the proposed method is summarized in Figure 4.5.

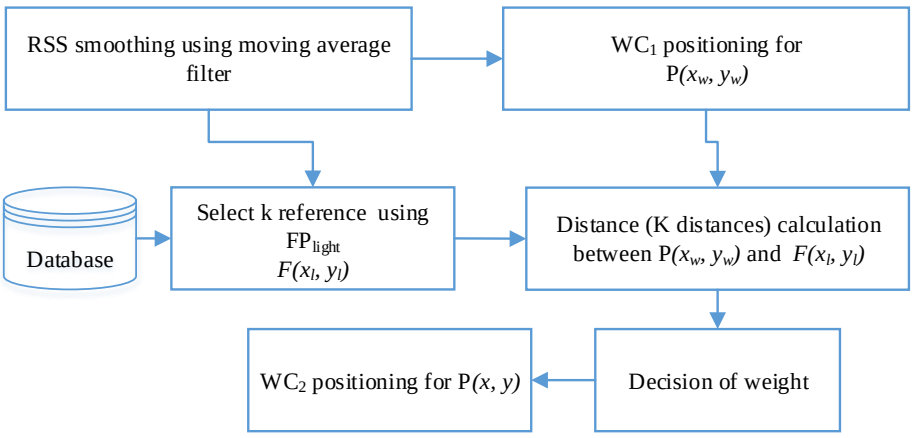


Figure 4.5: The process flow of the proposed localization method.

4.3.2 Experimental Setup

The proposed method was experimented in two different testbed environments. The first testbed is a hallway corridor and next is a fully furnished computer lab. The APs (BLE beacons in this experiment) was deployed in a rectangular fashion with an orientation towards their opposite walls at the corridor whereas, in the lab, the APs were deployed at the four corners and center of the room. The experimental conditions for the evaluation of the proposed fingerprinting method are tabulated in Table 4.1 and 4.2:

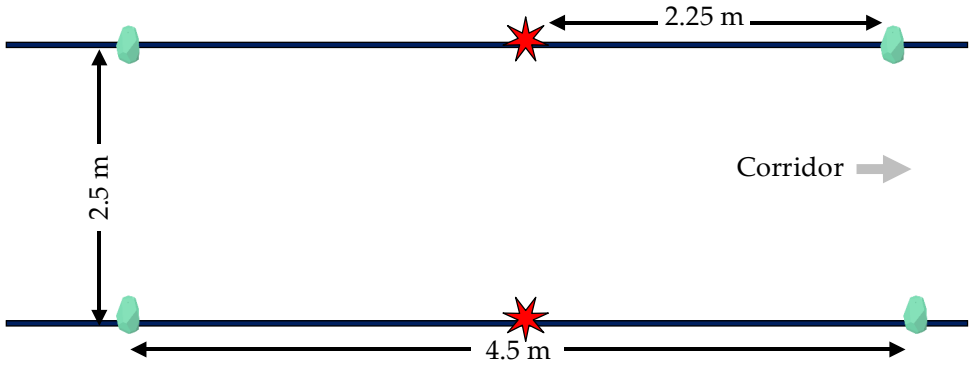
Table 4.1: Experiment conditions for the evaluation of the proposed fingerprinting method at the corridor.

Parameters	**Value**
Total number of APs (B)	14
Space between two adjacent APs	4.5 m
Breadth of corridor	2.5 m
The height of AP deployment	2.5 m
AP transmission power	+4 dBm
AP advertisement interval	300 milliseconds
Tx-Rx devices	Estimote beacons, iPhone-4S
A[dBm]	-60.85
Path loss exponent (n)	1.3
Degree (g) in WC estimation	0.5
No. of reference points (N)	12, 19, 26, and 36

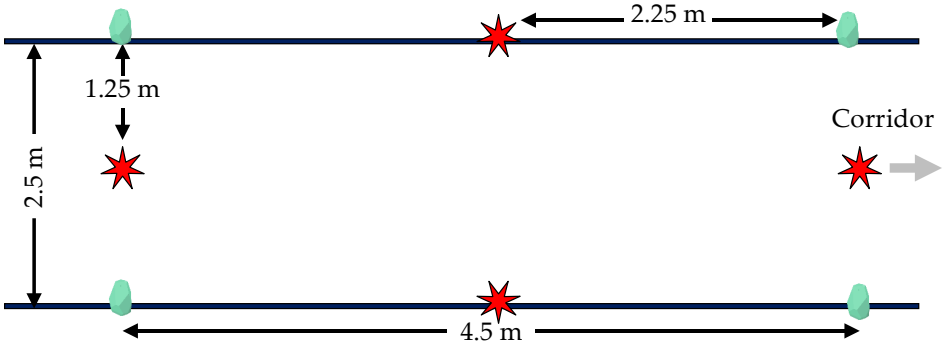
Table 4.2: Experiment conditions for the evaluation of the proposed fingerprinting method at the room.

Parameters	**Value**
Total number of APs (B)	5
Length breadth of the room	7.28 m $\times$ 7.24 m
The height of AP deployment	2.7 m
AP transmission power	+4 dBm
AP advertisement intervalr	300 milliseconds
Tx-Rx devices	Estimote beacons, iPhone-4S
A[dBm]	-64.97
Path loss exponent (n)	1.6
Degree (g) in WC estimation	0.5
No. of reference points (N)	20 and 32

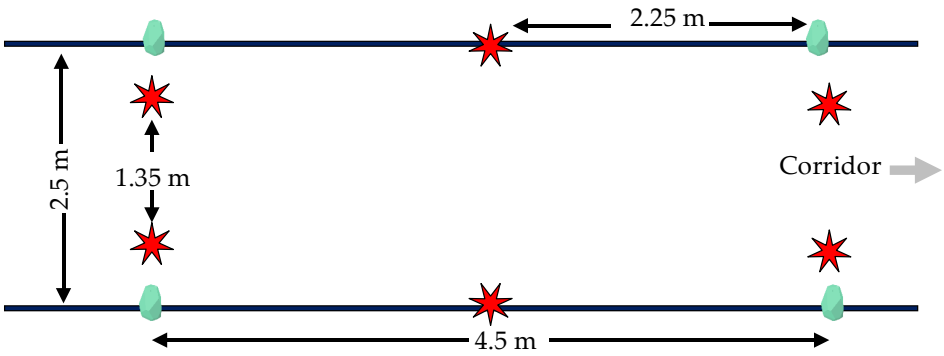
As the two opposite walls confine the corridor, it is divided into 62 uniform girds of length 1.25 m and breadth 0.9 m, forming 62 fingerprinting RPs. For analysis of the proposed method with different numbers of RPs, different sets of RPs are considered. At first, 12 RPs were taken such that there was an RP in between two APs along the corridor. Later, the number of RPs were uniformly increased over the testbed to 19, 26, and 36, as illustrated in Figure 4.6.



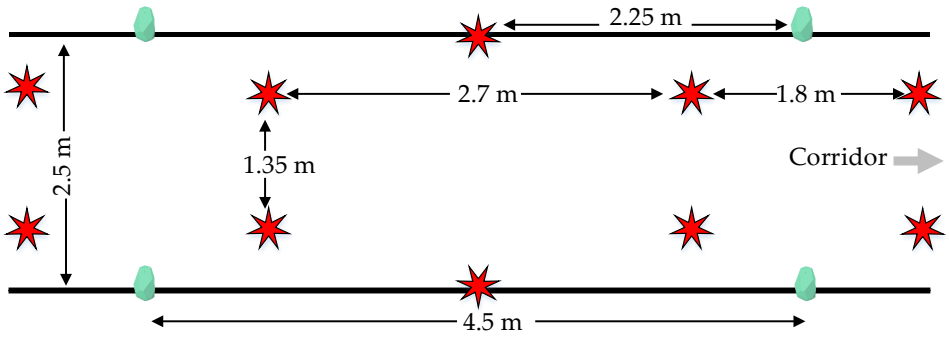
(a)



(b)



(c)



(d)

Figure 4.6: Testbed conditions for fingerprinting positions of (a) 12, (b) 19, (c) 26, and (d) 36 reference point distribution pattern over the testbed where red stars represent reference points.

Figure 4.7 shows the graphical representation of the testbed (corridor).

Figure 4.7: Experimental environment: beacons deployed in the corridor where measurement places are marked on the floor as A, B, and C.

However, the computer lab is not constrained by two opposite walls as in the corridor; it is divided into a uniform grid with 65 RPs of length and breadth 0.9 m. For the evaluation of the proposed method, the size of the RP grid was increased to 1.35 m and 1.8 m forming 28 and

16 fingerprinting RPs. In addition, to mitigate the high localization error of WC estimation near the walls, one more RP was added in the middle of each wall in the room. Hence, the proposed method is evaluated using 20 and 32 RPs in this testbed.

The graphical representation of the computer lab testbed is shown in Figure 4.8.

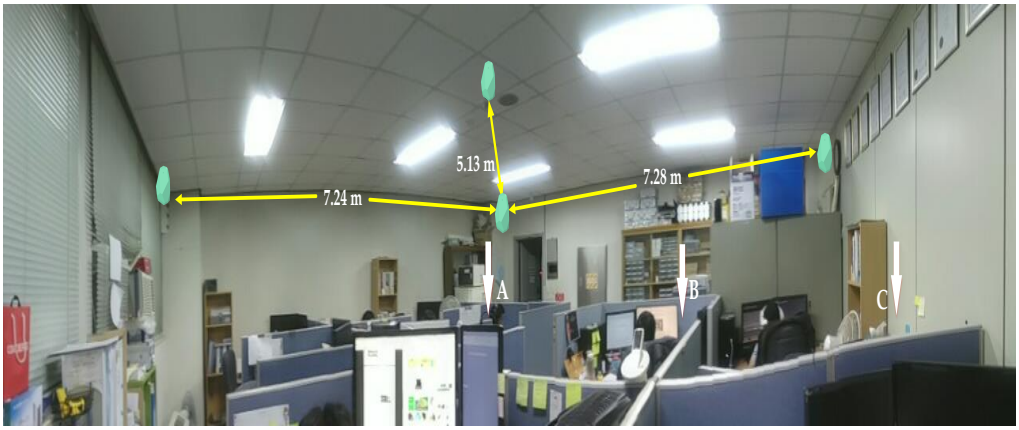


Figure 4.8: Panoramic view of the computer lab where measurement places are marked with white arrows (A, B, and C).

4.3.3 Experimental Results and Discussion

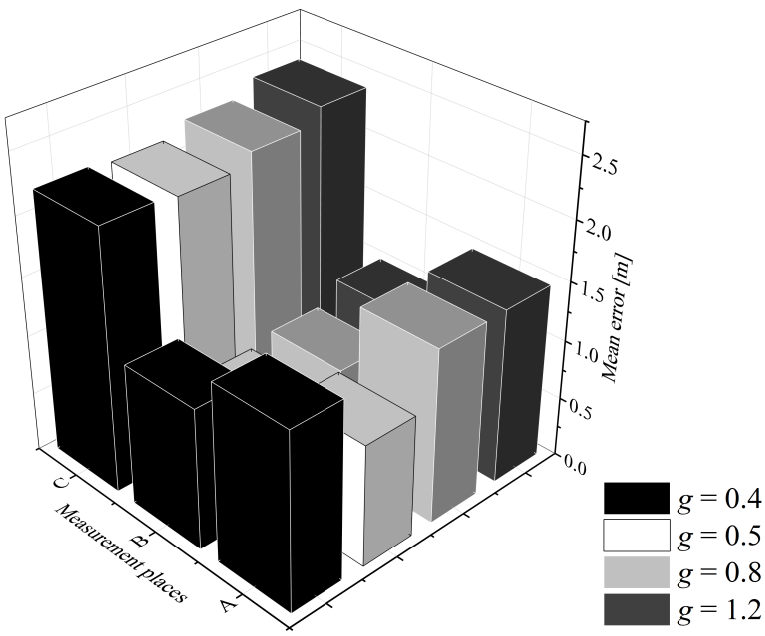
Considering the deployment of APs in a rectangular fashion, four APs were selected for WC operation. Moreover, researchers usually recommend different values for k such as $k = (2)$ [52] and $k = (3, 4)$ [67]. In the proposed approach, it was observed that for larger values of k (eg. 4 or 5), the RPs located far from the original tag position also was selected (owing to larger space among the RPs) that increased localization error. The least localization error was observed with $k = 3$, hence the value of k was set to 3 in this experiment. Note that in a traditional fingerprinting localization, large space between RPs reduces granularity or accuracy of the positioning system and small space increases accuracy. However, the small space does not increase the probability of correctly matching the fingerprints owing to similar fingerprinting of the close RPs [68].

The Decision of Degree of Weight

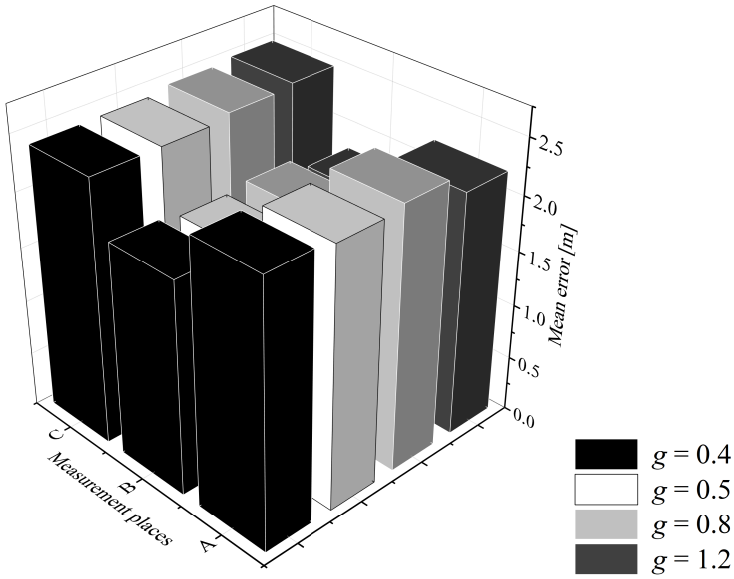
The degree of weight (g) in (2.3) can be adjusted for best positioning result in WC localization. If g is kept high, the WC estimation moves to the closest AP real coordinate. On the other hand, very low value (near to zero) may yield the WC estimation as a centroid point of the APs. Hence, the localization estimation error by WC procedure was evaluated with varying g $(0.4, 0.5, 0.8,$ and $1.2)$ in the testbed (hallway corridor). Considering the properties of WC estimation as illustrated

in Section 2.4.2 in Chapter 2, different regions across the corridor were chosen for the decision of degree of weight. Hence, the measurements were taken both at the border of the rectangular polygon formed by the deployed APs and at central regions of the polygon. Moreover, at any region, measurements were taken at three distinct places such as away from the wall (A), midway of the corridor (B), and near the wall (C), as shown in Figure 4.7. At a measurement place, 100 measurements were taken and the localization estimation error was averaged. The obtained result is presented in Figure 4.9.

As seen in Figure 4.9, the performance of WC estimation with the APs deployment scenario at degree 0.5 is better compared to others. Therefore, WC is operated with a degree of 0.5 for the evaluation of the proposed method. The same value of the degree is used in computer lab too, owing to the fact that the AP deployment height is similar at both testbeds and distance from the tag device to the beacon is in a similar range during the positioning operation.



(a)

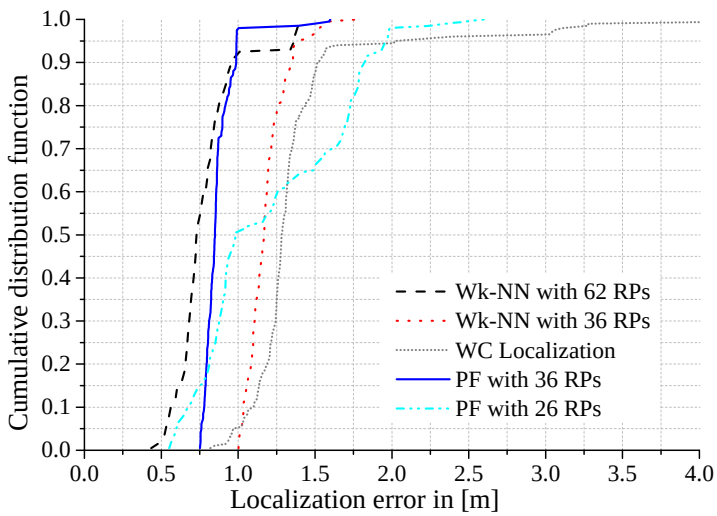


(b)

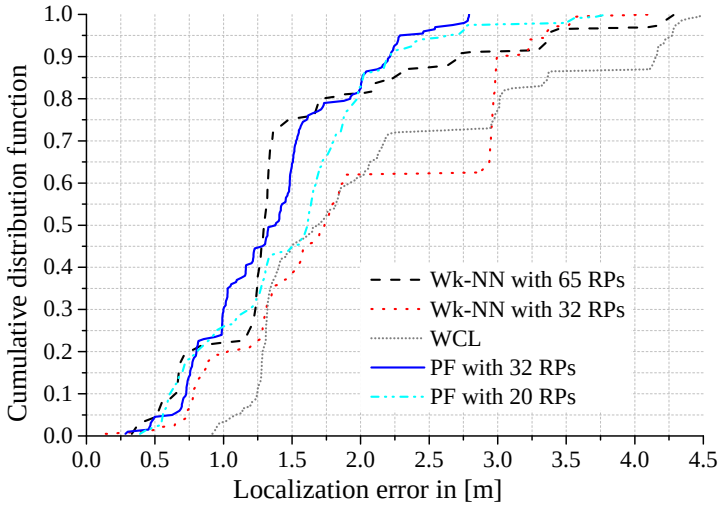
Figure 4.9: Average location error at different measurement places (A, B, and C) in the (a) central and (b) border regions of a rectangular polygon with a respective degree of weight.

Cumulative Distribution Function

The CDF of localization estimation error at a fixed point in both the testbeds by various positioning methods are given in Figure 4.10a and 4.10b.



(a)



(b)

Figure 4.10: CDF of localization estimation error where PF denotes the proposed fingerprinting method. (a) Corridor. (b) Computer lab.

Here, the performance of the proposed method (with 36 RPs in the corridor and 32 RPs in the computer lab) is comparatively similar with the Wk-NN fingerprinting (with 62 RPs in the corridor and 65 RPs in the computer lab). The performance of Wk-NN with the reduced number of RPs is degraded drastically.

Average Localization Error

For an exhaustive study, localization error was estimated at three regions at the corridor: (i) the center of the polygon formed by APs, (ii) the border of the rectangular polygon, and (iii) the edge of the corridor or end of AP deployment as shown in Figure 4.11.

At each region, the localization estimation error is measured at three different places (A, B, and C), as shown in Figure 4.7. Moreover, since the computer lab is not bounded by opposite walls as in the corridor, the measurements were taken at three different measurement places inside the computer lab (A, B, and C) as shown in Figure 4.8. At each measurement place, 200 samples of localization estimation error were taken, with 50 samples in each direction (0°, 90°, 180°, 270°). The obtained results at different measurement regions of the corridor are presented in Figure 4.12, 4.13, and 4.14.

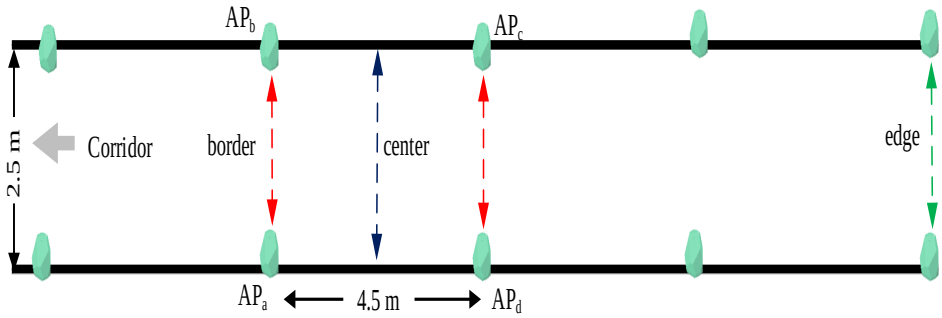
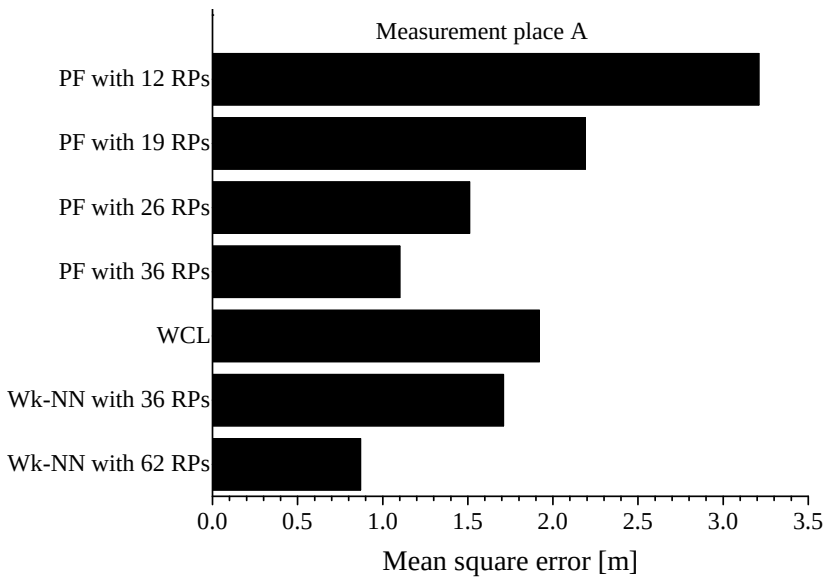
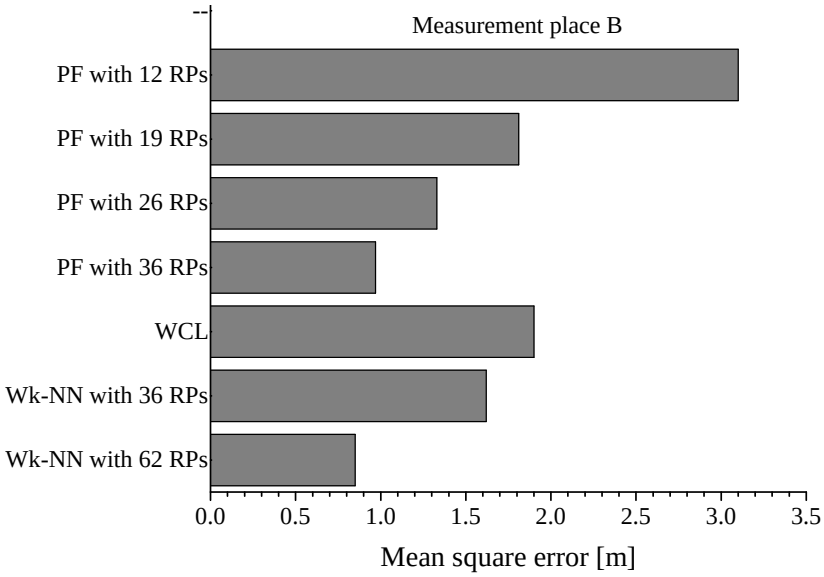


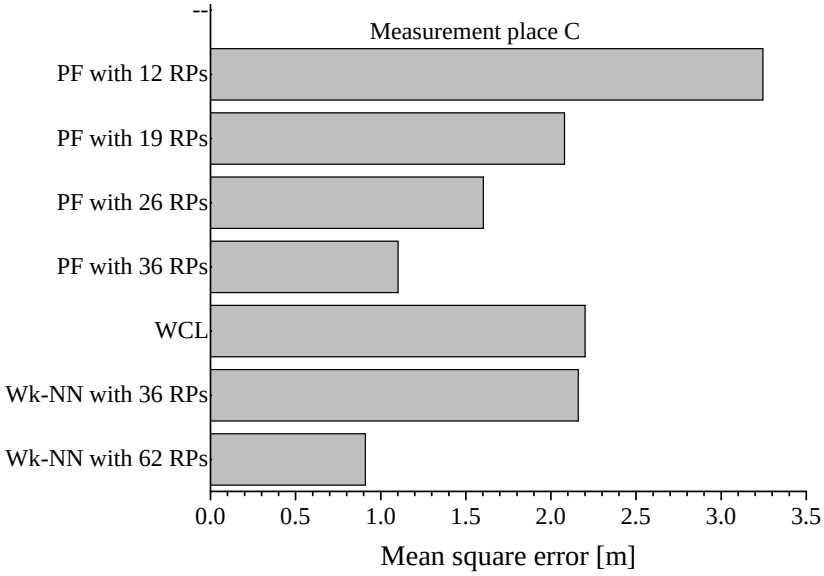
Figure 4.11: Location error measurement regions in the corridor where any four adjacent APs form a rectangular polygon (for example, the rectangle formed by APs a, b, c, and d). The red, blue, and green lines indicate the border of the polygon, the center of the polygon, and the edge of the corridor, respectively.



(a)

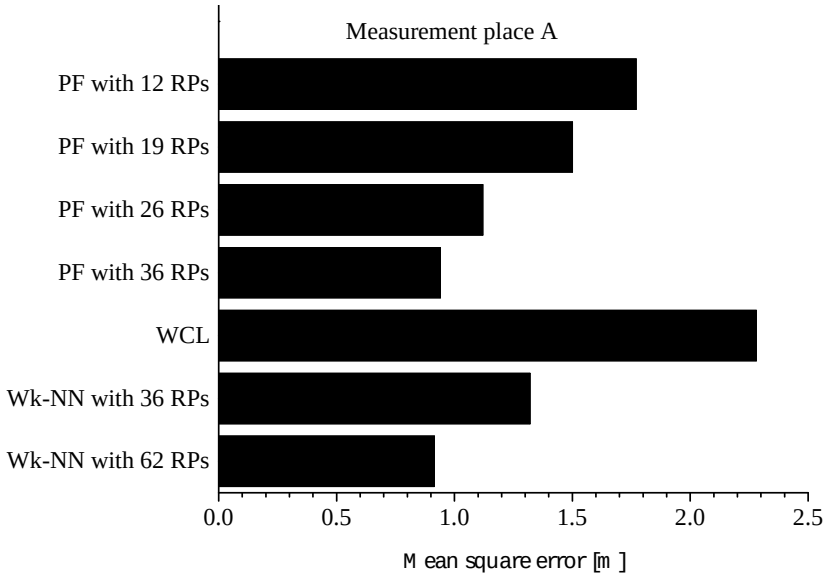


(b)

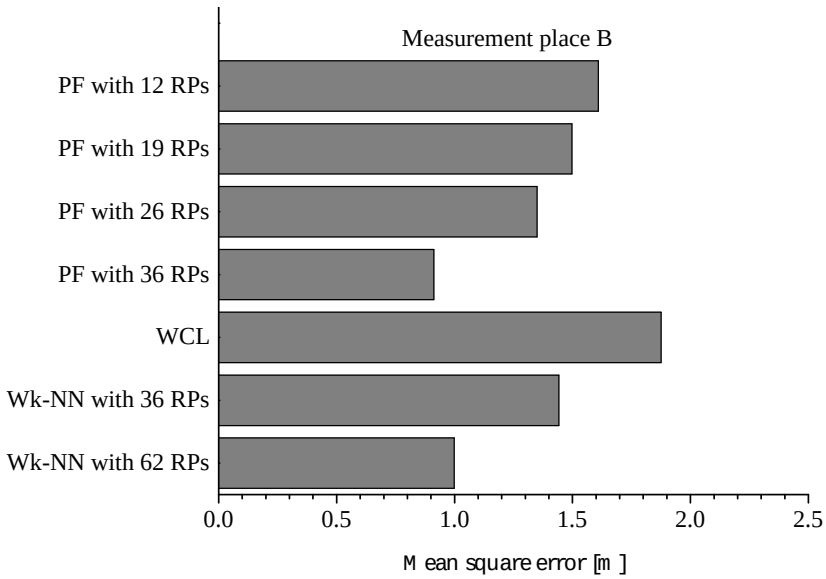


(c)

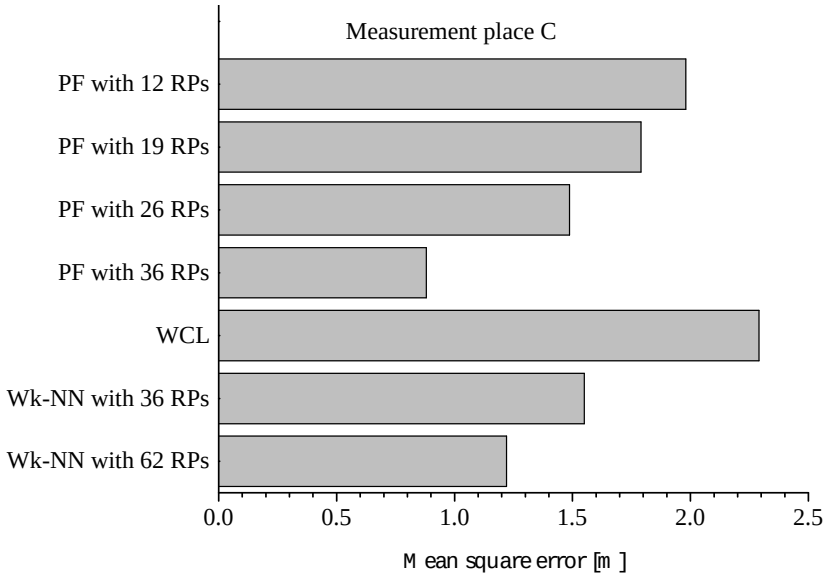
Figure 4.12: Average location error at different measurement places at the edge of the testbed corridor: (a) measurement place A, (b) measurement place B, and (c) measurement place C. PF: proposed practical fingerprinting; Wk-NN: weighted k-nearest neighbor; WCL: weighted centroid localization; RP: reference points.



(a)

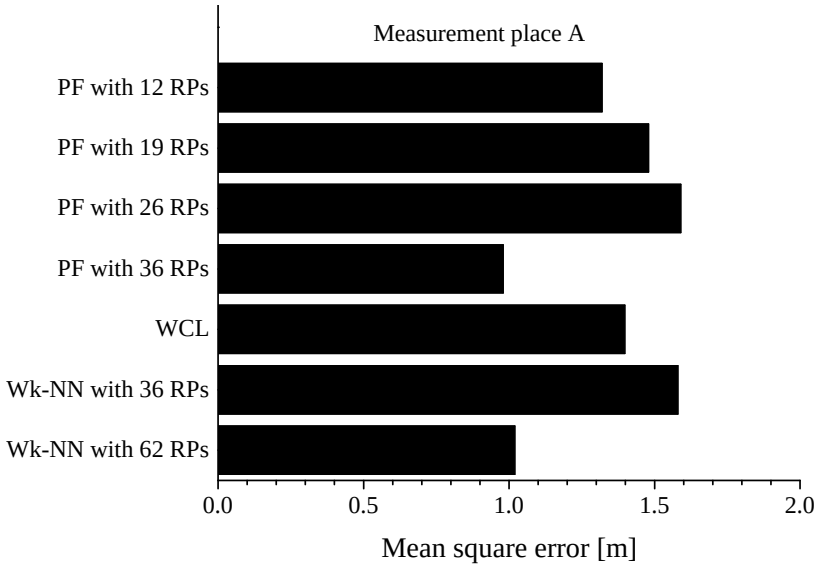


(b)

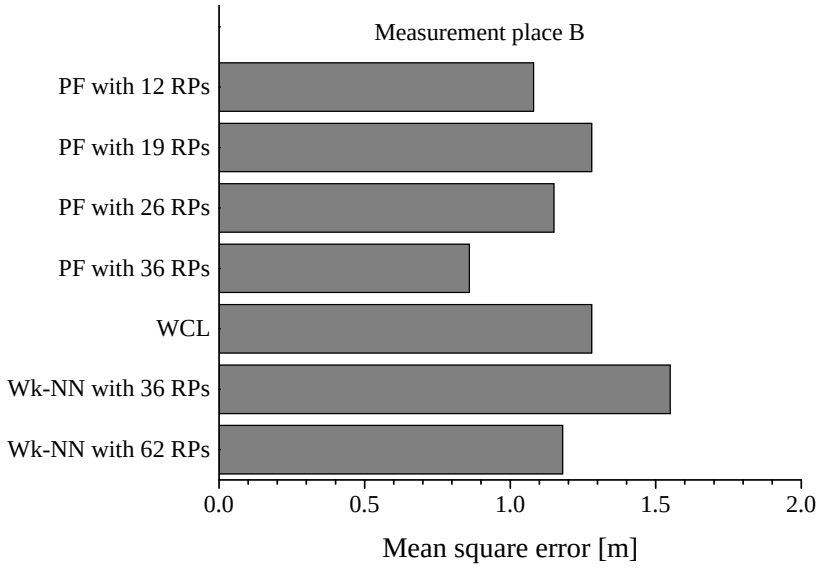


(c)

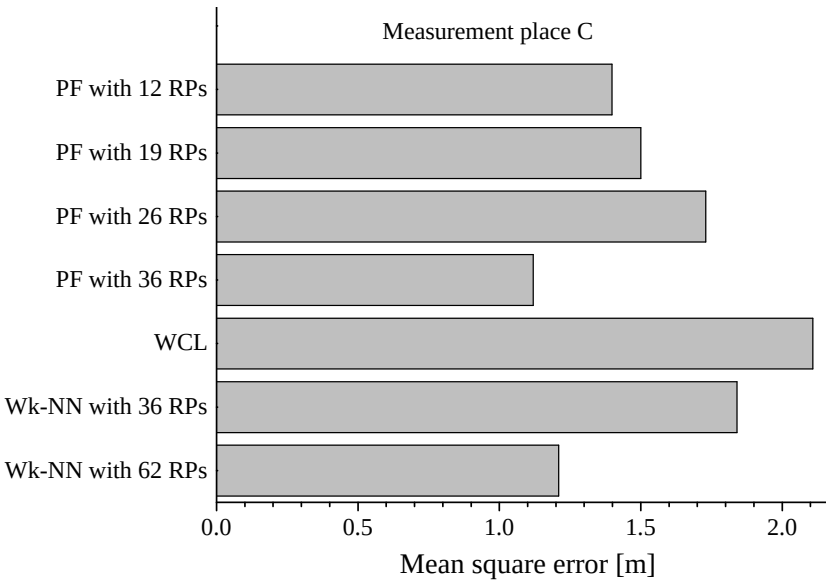
Figure 4.13: Average location error at different measurement places at the border of the rectangular polygon: (a) measurement place A, (b) measurement place B, and (c) measurement place C. PF: proposed practical fingerprinting; Wk-NN: weighted k-nearest neighbor; WCL: weighted centroid localization; RP: reference points.



(a)



(b)



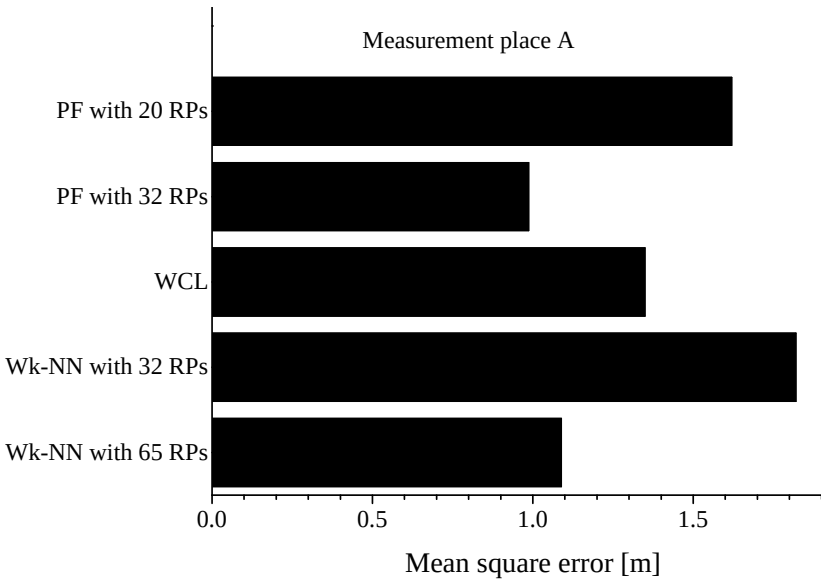
(c)

Figure 4.14: Average location error at different measurement places in the center of the rectangular polygon: (a) measurement place A, (b) measurement place B, and (c) measurement place C. PF: proposed practical fingerprinting; Wk-NN: weighted k-nearest neighbor; WCL: weighted centroid localization; RP: reference points.

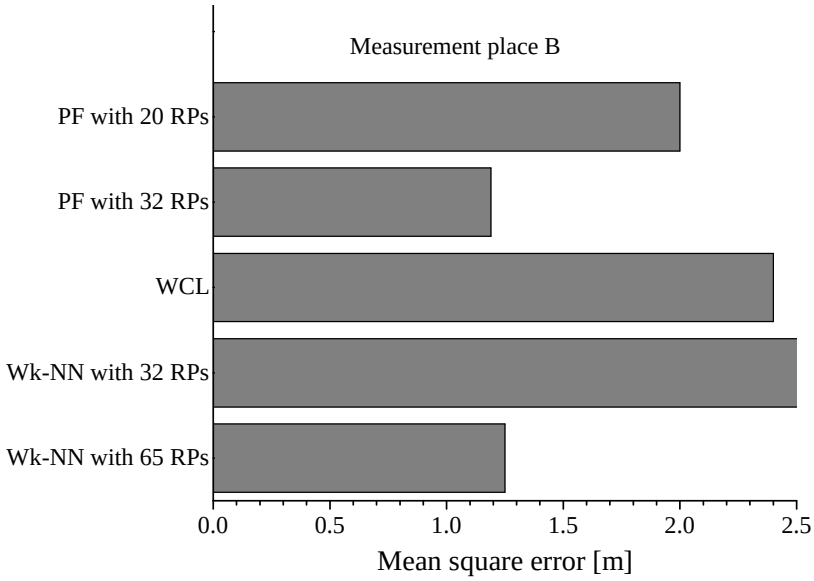
In Figure 4.12, 4.13, and 4.14, the average localization error of the proposed localization method in different measurement places is compared with Wk-NN fingerprinting and WC localization. As shown in Figure 4.12, the proposed method with less number of RPS (12 and 19) or high space among the RPs results to a high localization estimation error. At the edge of the corridor, the selected $k-NN$ RPs (for 12 and 19 RPs) are far away from the measurement place, where localization result of WC estimation is better than the proposed one. However, the proposed approach with 36 RPs yielded similar localization error compared to Wk-NN fingerprinting.

Figure 4.13 and 4.14 show the localization result at the measurement places in the border and center region of the rectangular polygon, respectively. Here, the proposed approach with 36 RPs has almost similar or lesser localization estimation error to Wk-NN (62 RPs) fingerprinting. In all regions of the corridor, the localization estimation error is lower at the midway (measurement place B) and larger near the walls. This condition has less effect on real IPS applications because people tend to walk down the mid-way of the corridor and normally avoid the places near the walls. The implementation of the proposed method in corridor testbed helps to reduce the number of RPs by 41.49%.

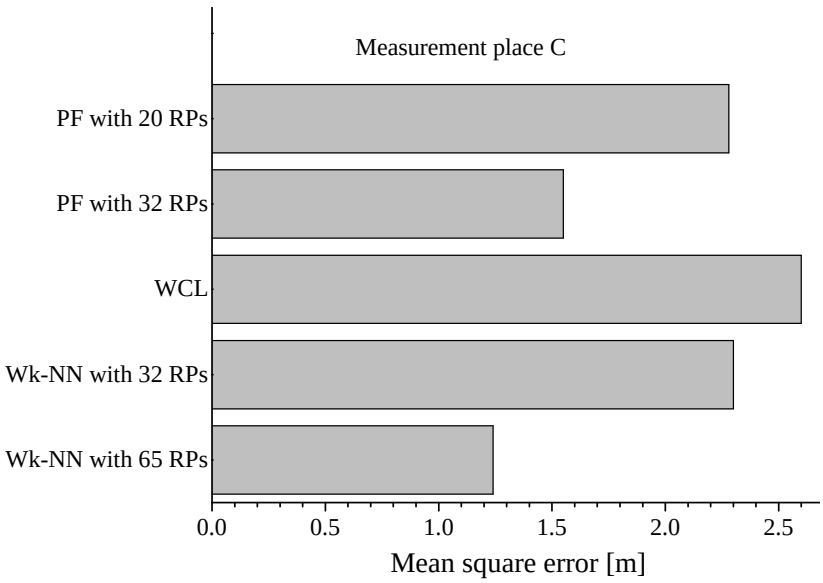
Similarly, the localization estimation result at the computer lab is presented in Figure 4.15.



(a)



(b)



(c)

Figure 4.15: Average location error at different measurement places in the room: (a) measurement place A, (b) measurement place B, and (c) measurement place C. PF: proposed practical fingerprinting; Wk-NN: weighted k-nearest neighbor; WCL: weighted centroid localization; RP: reference points.

Here, the localization estimation error is higher owing to a fully furnished testbed condition. At the center of the room (measurement place A), the localization error of WC estimation is the least due to the presence of an AP just above the tag device and all other four APs in equidistant to this place. The proposed method with 32 RPs closely follows the average localization error of WK-NN everywhere. Similar to the corridor, the localization estimation error is increased as we move towards the wall of the room. The proposed method reduced the number of RPs by 49.23% in computer room testbed.

4.4 Proposed Method-II

The conventional Wk-NN fingerprinting is also called flat Wk-NN where all the RPs on the testbed are considered for final localization estimation. The flat Wk-NN can be improvised into two-step Wk-NN by employing clustering where RPs are divided among the clusters. The second approach on improvising the traditional Wk-NN fingerprinting uses multiple features to represent an RP's fingerprint and utilizes clustering method, which enhances the localization accuracy, reduces the physical size of the radio map database, and minimizes the computational complexity.

The suggested method uses a propagation model to convert received RSS of APs to distance and estimate weight the WC of nearby APs. The estimated WCs, along with RSS and rank of nearby APs are stored in the server database for localization instead of average RSS from all deployed APs. Here, the rank of the APs refers to a sorted list of APs in descending order where first u APs in the sorted list is used for WC estimation. First, the proposed method makes use of diverse fingerprinting features to increase positioning accuracy that also decreases both the physical size of the database and the amount of data communication with the server in the online phase; second, APC minimizes the searching space of RPs and reduces the computation time; third, EA is utilized to smooth the noisy RSS.

4.4.1 Affinity Propagation Clustering

The APC joins all of the points in the large space and makes each point node (RP in this study) a potential exemplar. The points launch responsibility message and receive availability message constantly that continue to extend the gap between the exemplar and supplementary points until the exemplar is decided. The radio map ($\mathbf{\Psi}_{RSS}$) is represented as follows:

$$\mathbf{\Psi}_{RSS} = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,N} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{B,1} & \psi_{B,2} & \cdots & \psi_{B,N} \end{pmatrix}, \tag{4.6}$$

where $\psi_{b,j} = \overline{y^*}_{b,j}\big(b = 1, 2, 3, ...B; j = 1, 2, 3, ..., N\big)$ is the mean RSS of the bth AP (AP_b) signal at the jth RP (RP_j). At j^{th} RP, the RSS fingerprint from the deployed APs is $\Psi_j = \big(\psi_{1,j}, \psi_{2,j}, ..., \psi_{B,j}\big)^T$.

Let Ψ_i and Ψ_j denote the mean RSS vectors of any two RPs, the similarity between the RPs indicated by $sim\big(i, j\big)$ is given as:

$$sim\big(i, j\big) = -\|\Psi_i - \Psi_j\|^2, \forall i, j \in (1, 2, ..., N) \tag{4.7}$$

The similarity values form a $N \times N$ similarity matrix Z, where N is the total number of RPs to be clustered. The value $sim\big(j, j\big)$ on the diagonal of the matrix Z is called self-similarity, is used to judge whether RP_j can become the exemplar. The self-similarity is also known as a preference ($Pref$) given as:

$$Pref = median\big\{sim(i, j), \forall i, j \in 1, 2, ..., N\big\} \tag{4.8}$$

The responsibility and availability messages transmitted by RPs are denoted as $r\big(i, j\big)$ and $a\big(i, j\big)$, respectively. Note that both of them are set to zero initially. The responsibility message represents the confidence level of RP_j as an exemplar of RP_i, which is update by the following relation:

$$r(i, j) = sim(i, j) - \max_{j' \neq j}\Big\{a(i, j') + sim(i, j')\Big\} \tag{4.9}$$

The availability message represents that RP_i selects RP_j as the confidence center of its exemplar, which is update by the following relation:

$$a(i, j) = min\Bigg\{0, r(i, j') + \sum_{i' \neq i,j} max\big\{0, r(i', j')\big\}\Bigg\} \tag{4.10}$$

Moreover, $a(j, j)$ is a self-availability message, which represents the cumulative evidence for RP_j as the exemplar that is calculated as follows:

$$a(j, j) = \sum_{i' \neq j} max\big\{0, r(i', j)\big\} \tag{4.11}$$

In order to avoid the possible ringing oscillation while updating (4.9) and (4.10), a damping factor $\gamma \in [0.5, 1)$ is exploited resulting into the following equations:

$$r_t(i, j) = \gamma * r_{t-1}(i, j) + (1 - \gamma) * r_t(i, j) \tag{4.12}$$

$$a_t(i, j) = \gamma * a_{t-1}(i, j) + (1 - \gamma) * a_t(i, j) \tag{4.13}$$

Here, r_t and a_t corresponds to the value of responsibility and availability of the current iteration, respectively. Similarly, r_{t-1} and a_{t-1} are the value of responsibility and availability of the

last iterations. The exemplar is updated according to the value of $r(i,j)+a(i,j)$. If $r(i,j)+a(i,j)$ is largest, it denotes that RP_j is the exemplar of RP_i. Else, RP_i will be selected.

4.4.2 Process Flow of the Proposed Method

The proposed method has two phases of operation; training, and testing. In the training phase, the testbed is divided into uniform N imaginary grids where the center of each grid represents an RP. Hence, the testbed contains N RPs in locations $L_r(r \in 1,2,\ldots,N)$ and the RSS data of B APs are acquired at r^{th} RP q times; that is: $AP_r = \{ap_{(r,1)}, ap_{(r,2)}, \ldots\ldots, ap_{(r,B)}\}$, where $ap_{(r,i)} = Avg\big(ap^1_{r,i}, ap^2_{r,i}, \ldots\ldots, ap^q_{r,i}\big)$.

The acquired RSS data of B APs are used for APC and only u APs ($u < B$) that are selected based on the largest RSS value are used for fingerprinting information. Here, the RSS data is acquired from four directions at an RP as shown in Figure 4.2 with a tag device in $texting/messaging$ position.

Figure 4.16 illustrates the localization estimation procedure of the proposed system. The observed testing RSS data at the unknown location r' can be represented as:

$AP_{r'} = \{ap_{(r',1)}, ap_{(r',2)}, \ldots, ap_{(r',B)}\}$, where $ap_{(r',i)} = Avg\big(ap^1_{r',i}, ap^2_{r',i}, \ldots, ap^q_{r',i}\big)$.

Here, t is the time instant and EA is the exponential averaging given by (3.11). Let H be the set of cluster head and N' ($N' < N$) be the number of RPs that are members of the cluster head. The cluster-head (RP_H) is determined in the following way:

$$RP_H = argmin_{m \in H} \left\| WC_{r'} - WC_m \right\|, \tag{4.14}$$

where $WC_{r'}$ and WC_m denote online observed weighted centroid from nearby u' beacons at unknown place r' and the stored weighted centroid of mth cluster-head RP, respectively. Note that u and u' are seen APs (BLE beacons) of same size at training and testing phase, respectively. When the cluster-head RP is chosen, Euclidean distances are calculated based on online-observed fingerprinting data and the stored fingerprinting data for N' RP as given below:

$$D_{WC_l} = \left\| WC_{r'} - WC_p \right\|, p \in N' and\ l = (1,2,3,\ldots\ldots,N') \tag{4.15}$$

$$D_{RSS_l} = \sum_{b=1}^{u} \sqrt{\left(ap_{b,r'} - ap_{b,p}\right)^2}, p \in N' and\ l = (1,2,3,\ldots\ldots,N') \tag{4.16}$$

where p is a member of the selected cluster head.

The member RPs (N') grouped under the chosen cluster-head are arranged with ascending order of D_{wc_l} and the first k RPs with their known positions $J_z[x_z, y_z]$ are selected to estimate T_{WC} as follows:

$$T_{WC} = \frac{\sum_{z=1}^{k} J_z G_{WC_z}}{\sum_{z=1}^{k} G_{WC_z}}, where\ G_{WC_z} = \frac{1}{D_{WC_z}}. \tag{4.17}$$

The similar process is repeated using D_{RSS_l} (u in (4.15) is the stored AP at the RP for least D_{WC_l}) to estimate T_{RSS} using the following relation:

$$T_{RSS} = \frac{\sum_{z=1}^{k} J_z G_{RSS_z}}{\sum_{z=1}^{k} G_{RSS_z}},\ where\ G_{RSS_z} = \frac{1}{D_{RSS_z}}. \tag{4.18}$$

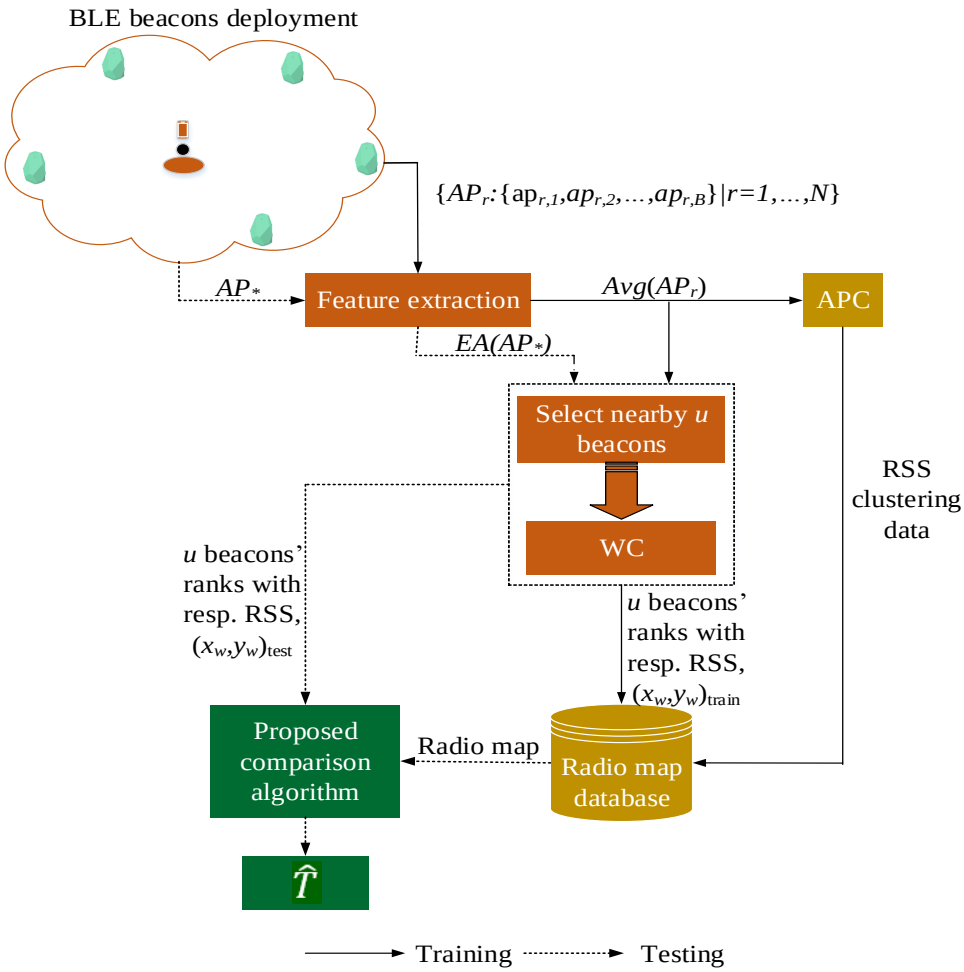


Figure 4.16: Process flow of the proposed fingerprinting IPS.

Hence, the final location is estimated utilizing the following relation:

$$\hat{T} = (1 - \alpha) \times T_{RSS} + \alpha \times T_{WC}, \tag{4.19}$$

where $\alpha = \frac{u''}{u}$. Note that u'' represents the total number of matched APs, i.e., number of online observed beacons at location r' that matches its rank to the stored ones for the RP with least D_{WC_l}.

4.4.3 Experimental Setup

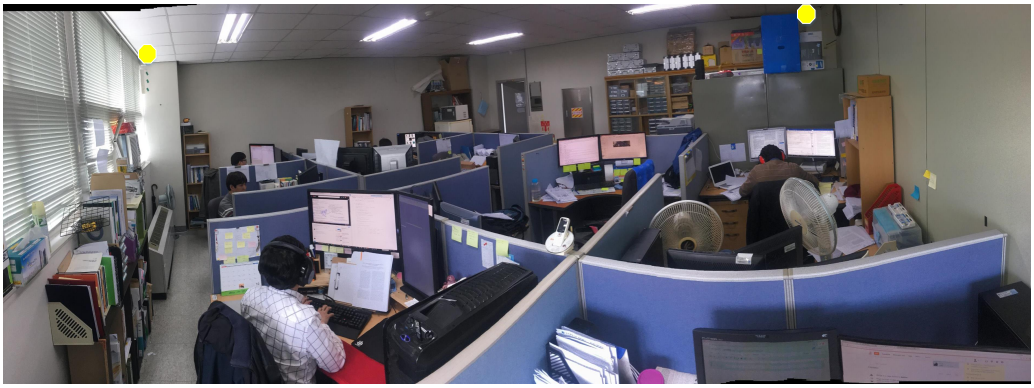
The Testbed

The APs were deployed in the rectangular fashion on the ceiling at 2.7 m height. The advertisement interval of the APs (BLE beacons) was set as 300 ms (three beacon packets per second), and the tag device (iPhone) scans APs with a 1-second interval on average. The distance between the two adjacent APs was 6.8 m and the testbed was divided into girds of size 1.35 m length and breadth. The calibrated values of parameters (A and n) of the propagation model given by (3.1) were $-62.91dBm$ and 1.45, respectively. Note that the parameters were calibrated with representative data from all testbed environments and used uniformly in all the testbed environments owing to the fact that the same set of parameters are employed to compare offline and online WC in the proposed method. Furthermore, based on the deployment of APs in rectangular fashion and distance among the adjacent APs, the value for the degree (g) and a number of nearby APs (u) in WC estimation was assumed as 0.7 and 4, respectively for training and testing phases.

The testbed representing a research lab and the hallways are presented in Figure 4.17.

(a)

(b)



(c)

Figure 4.17: Testbeds where the yellow dots represent BLE APs. (a) Left the center of the corridor. (b) Right from the center of the corridor. (c) Panoramic view of a fully furnished computer lab.

Figure 4.18 illustrates the graphical representation of the testbed.

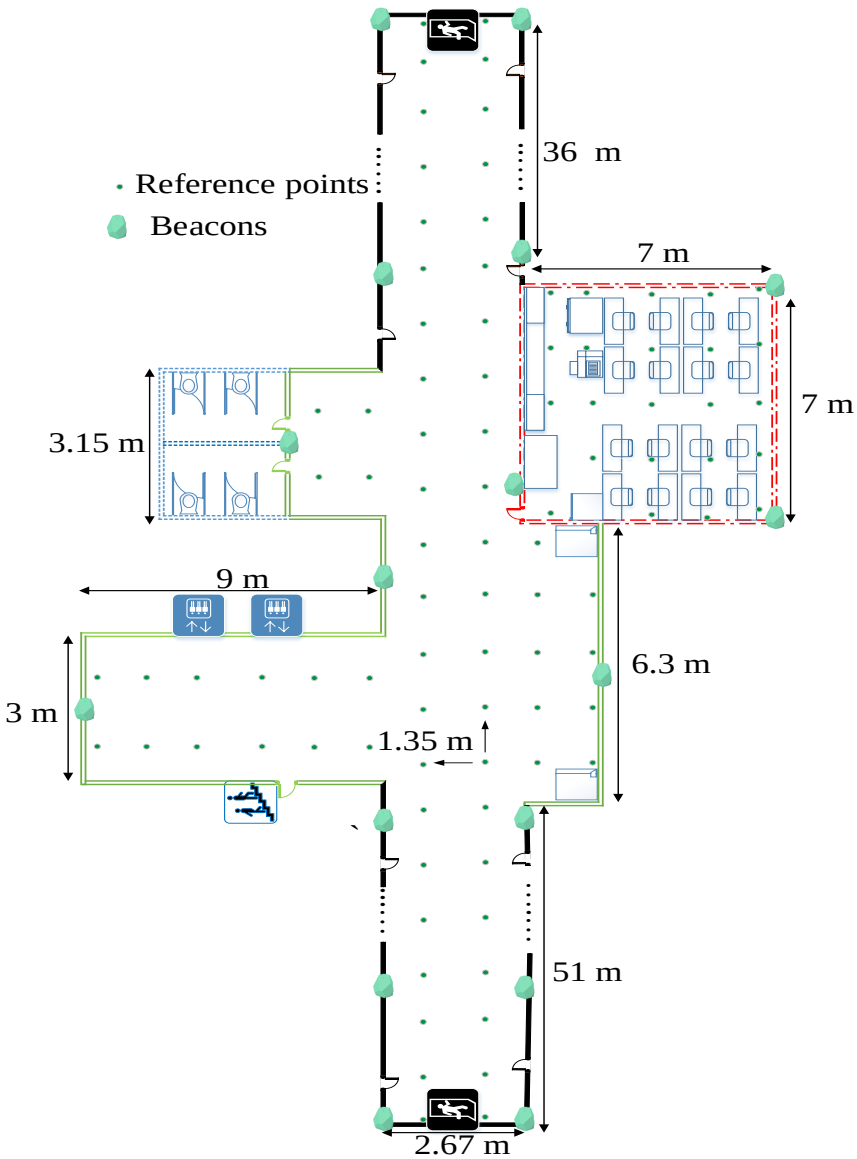


Figure 4.18: Graphical representation of the testbed.

RSS Clustering using APC

The APC operation was run after acquiring the RSS data at all the RPs from all the deployed APs. Since the damping factor is introduced in (4.11) and (4.12) while updating the message in APC, the APC operation was analyzed by varying the damping factor $(\gamma) \in [0.5, 0.9]$ to obtain the minimum number of iterations. The obtained result is presented in Figure 4.19.

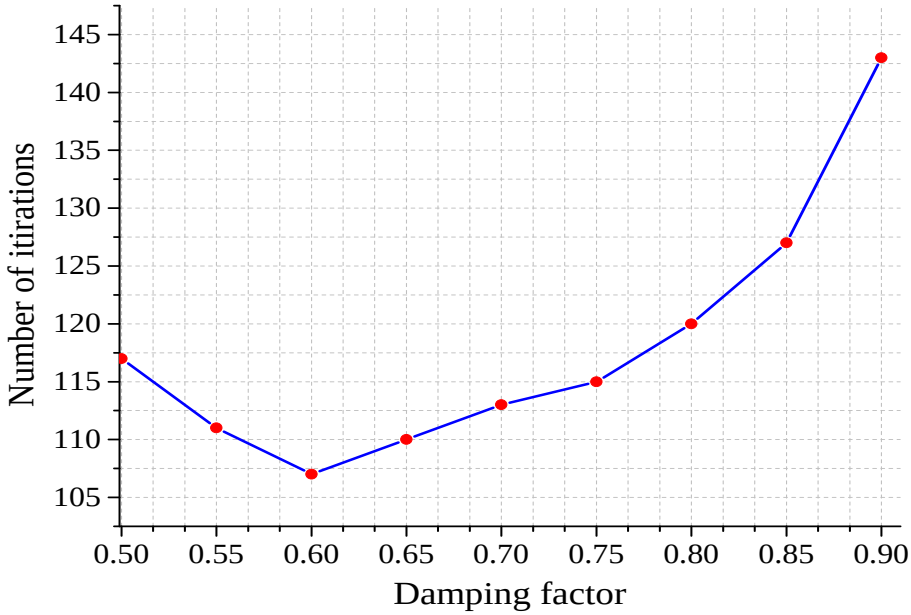


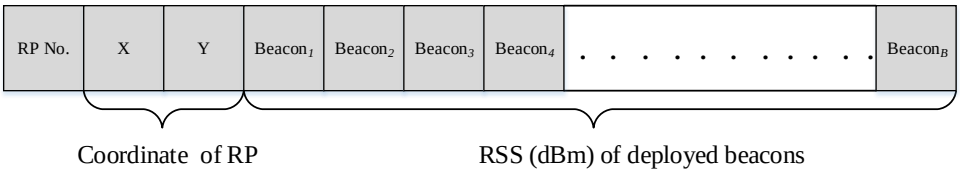
Figure 4.19: Variation on a number of iterations with respect to the damping factor.

Figure 4.19 shows that the damping factor of 0.6 performs the RSS clustering with the least number of iterations. Since least number of iteration results in the least computational cost, $\gamma = 0.6$ was used for APC operation in the experiment.

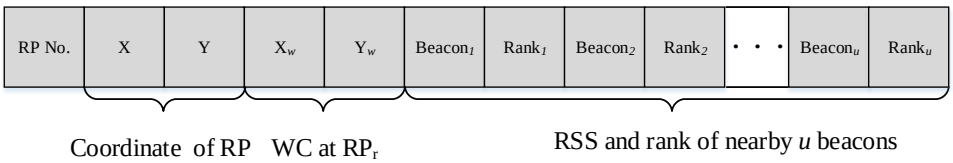
4.4.4 Experimental Result and Discussion

Size of Database with Proposed Fingerprinting Positioning

In a conventional fingerprinting positioning, a small default RSS value (<-90 dBm) is assigned to the undetected APs that are normally referred to as invalid data [69]. This kind of invalid data is entirely omitted in the radio map database of the proposed method that helps to minimize the size of the database. Moreover, as compared to typical fingerprinting localization, the RSS from $B - u$ (B as a total deployed APs) APs at the RPs are redundant in the suggested method. Figure 4.20 shows the columns of the tables inside the database where each row in the database table contains the fingerprinting data of an RP.



(a)



(b)

Figure 4.20: Columns of tables inside the fingerprinting database. (a) Typical RSS-based fingerprinting. (b) Proposed fingerprinting method.

The Oracle Database 10g Express was used to store the radio map data in the proposed system [70]. Each cell in the database table is specified using numeric datatype. The physical memory size occupied by the radio map database was obtained using SQL query as illustrated in Table 4.3.

Table 4.3: Physical memory occupied by fingerprinting database.

Parameter [m]	**Typical method**	**Proposed method**
Avg. size of an RP	158.9 bytes	48.2 bytes
Size of fingerprinting database	32,413 bytes	9,830 bytes

It can be seen from Table 4.3 that the physical size of the radio map database of the proposed method ($u = 4$) can be reduced by 69.7% compared to the conventional fingerprinting method ($B = 37$).

Localization Performance of the Proposed Method

The proposed approach of IPS is evaluated and compared in terms of positioning time (processing time required for positioning) and the localization estimation error. To calculate the processing time, the objects of $NSDate$ [71] were declared called $positioningStartTime$ and

$positioningFinishTime$ at the beginning and end of the positioning application, respectively. Later, $NSTimeInterval$ is utilized to calculate the positioning time as the proceed time between the $positioningStartTime$ and the $positioninFinishTime$ in seconds. Figure 4.21 illustrates the positioning time of the proposed method compared to existing methods with the positioning time of the proposed method represented in t seconds.

Figure 4.21 shows that the positioning time of the proposed method is lesser compared to the time taken by Wk-NN with APC (Wk-NN+APC). It concludes that the proposed method is computationally effective compared to Wk-NN+APC because the proposed method not only implements clustering, it also significantly reduces the number of APs for positioning. In addition, the positioning time of Wk-NN and NN fingerprinting is very large compared to the proposed method and Wk-NN+APC. Moreover, it is noted that the positioning time incurred by Wk-NN is slightly larger than the NN fingerprinting.

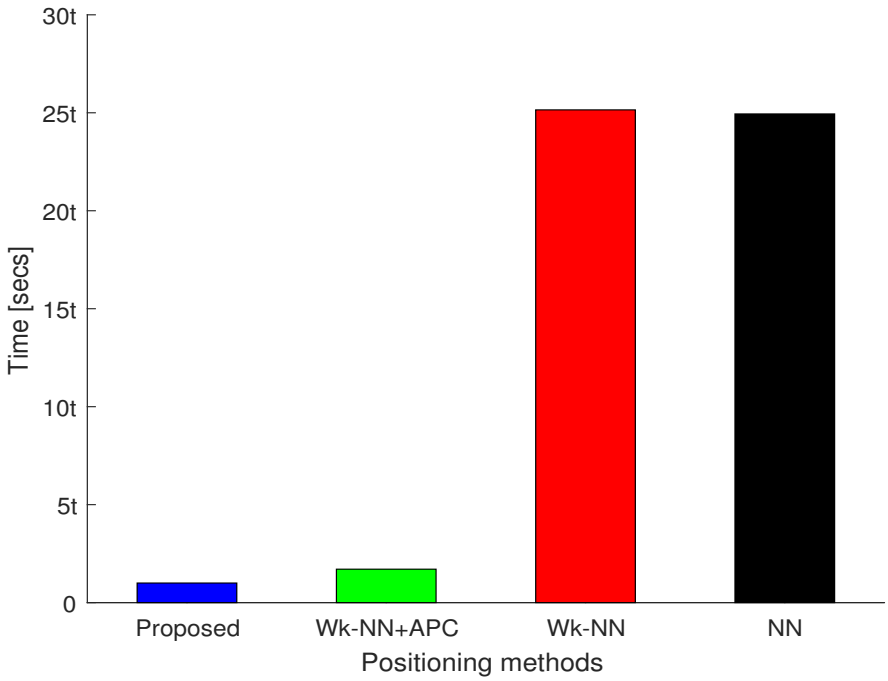


Figure 4.21: Performance comparison between the positioning methods in terms of positioning time.

The presented positioning time result can be verified in terms of $Big - O$ notation [72] too. The computational complexity of the conventional fingerprinting or flat-based fingerprinting localization grows with the radio map size (number of RPs across the testbed). Considering N RPs on the testbed, the computational complexity of the typical fingerprinting localization or linear

matching process is $O(N)$ [73, 74]. Meanwhile, the complexity of the cluster-based fingerprinting where every cluster (H being the total number of cluster-centers) have a uniform number of RPs is $O(N/H)$. However, in the proposed method, the RPs under a cluster head $N'(N' << N)$ can vary and has a computational complexity of $O(N')$. That is, the coarse localization stage of the suggested method minimizes the area of interest from N RPs to N' RPs that reduces the number of fingerprints to be compared for fine localization and finally reduces the computational complexity of the IPS.

As seen in Figure 4.18, the testbed consists of a long hallway/corridor and a computer lab where elevators, stairs, vending machine, and lavatories are present at the hallway. Hence, it can be inferred that the human movement at the center of the hallway is more frequent compared to another part of the corridor. Therefore, the experimental testbed was categorized into three areas; the well-furnished computer lab, the center of the hallway, and the remaining part of the corridor. In Figure 4.18, red, black, and green boundaries represent the computer lab, hallway/corridor, and the center of the hallway, respectively. For the best positioning result, k was set to 3 in all experiments. The evaluation of the proposed method in terms of localization error by considering three cases in the testbed is as follows:

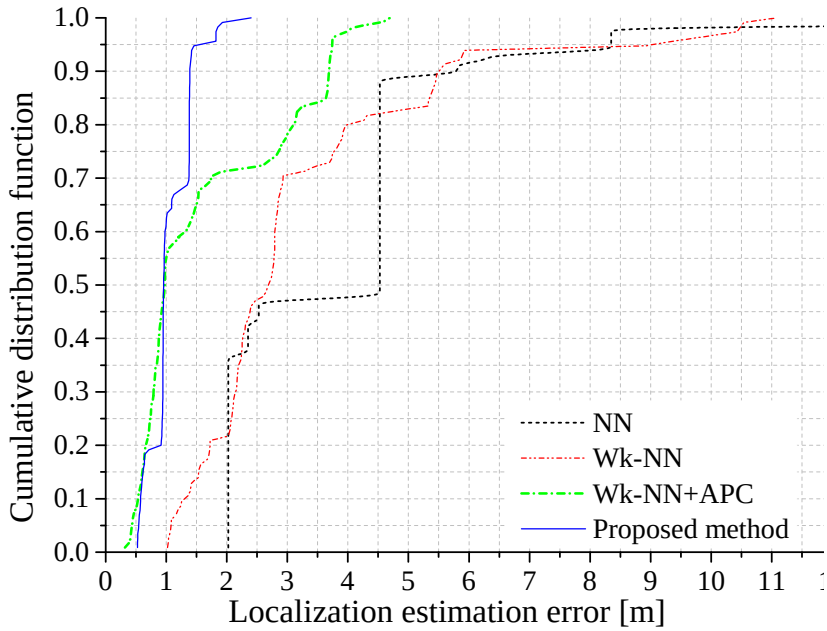
1. Hallway/Corridor: The CDF and average localization estimation error of the proposed method compared to existing methods is presented in Figure 4.22.

   Here, the proposed method has a probability of 63% that the localization error is below 1m. Similarly, the Wk-NN with and without clustering have a probability of 55% and 22.5% for the same 1m positioning error condition, respectively. It clearly shows the effectiveness of RSS clustering in IPS. Moreover, the average error of localization estimation of the proposed method, Wk-NN+APC, Wk-NN only and NN are 1.06 m, 1.63 m, 3.59 m, and 4m, respectively.

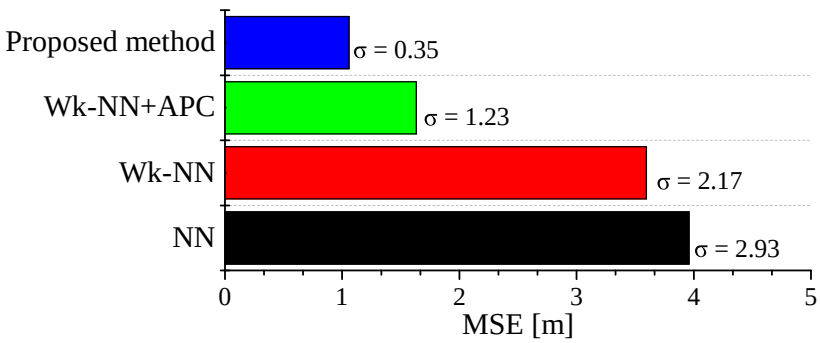
2. Center of the Hallway: The presence of different public amenities at the center of the hallway may result in different RF attenuation compared to another part of the hallway. In addition, the center of the corridor (marked by a green boundary line in Figure 4.18) is wider and crowded. The localization result at the center of the hallway is presented in Figure 4.23. The average localization estimation error is slightly increased as compared to the other parts of the corridor. The proposed method and Wk-NN+APC have 1.32 m and 2.46 m of error, respectively.

3. Computer Lab: The third case of the experiment is the computer lab. As the computer lab is well furnished, more RF attenuation is expected as compared to the corridor. The CDF and the average localization estimation error is illustrated in Figure 4.24. The CDF and the average error of localization estimation clearly show the advantage of using clustering in IPS. It is seen that the localization accuracy of Wk-NN and NN fingerprinting are significantly de-

teriorated whereas Wk-NN+APC and the suggested approach yielded an acceptable result. Note that the proposed method has a probability of 2.5% that the localization estimation error is above 2.5 m. It can be due to RF attenuation in the room and the wrong choice of exemplar on the testing phase.

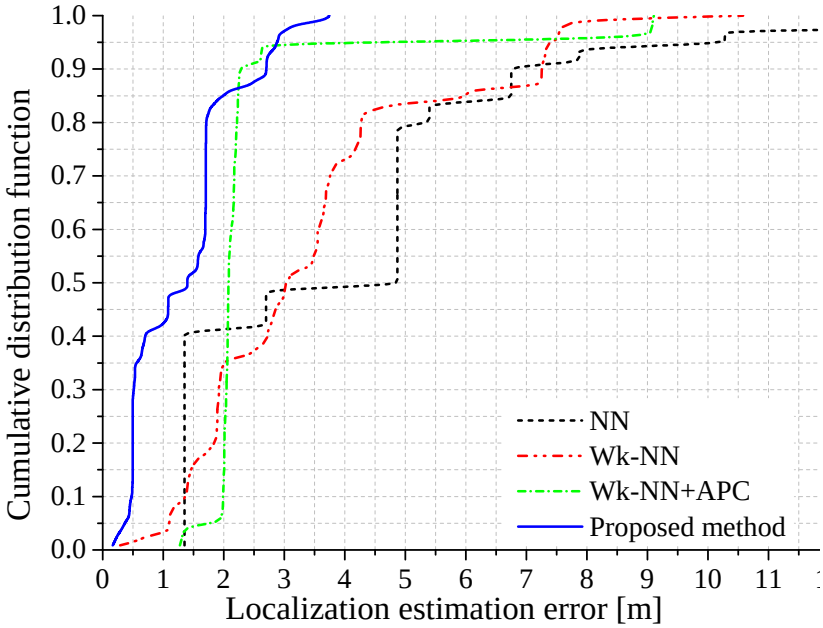


(a)

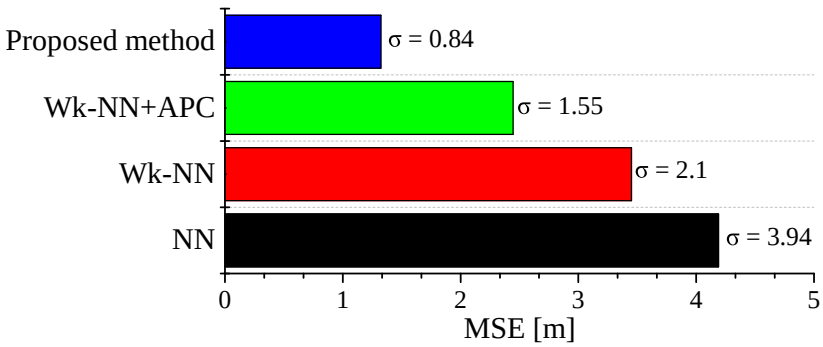


(b)

Figure 4.22: Localization estimation result at the hallways in the testbed. (a) Cumulative probability function. (b) Average localization estimation error.

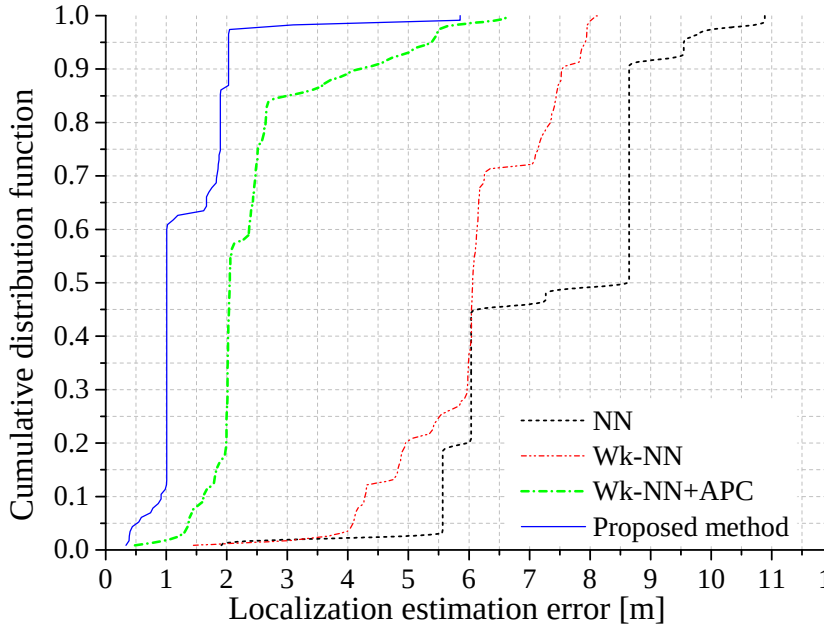


(a)

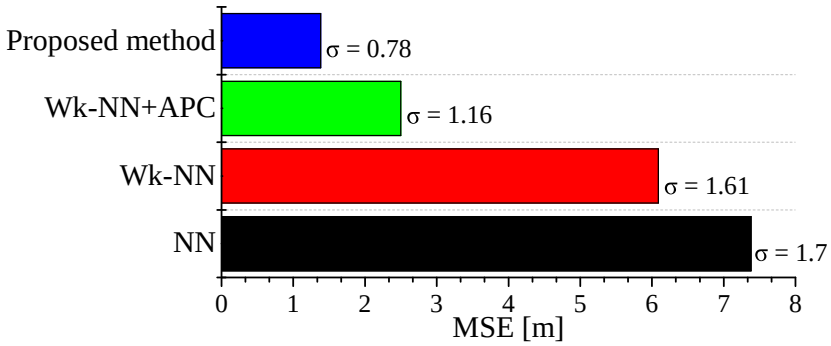


(b)

Figure 4.23: Localization estimation result at the center of hallways in the testbed. (a) Cumulative probability function. (b) Average localization estimation error.



(a)



(b)

Figure 4.24: Localization estimation result at the computer lab in the testbed. (a) Cumulative probability function. (b) Average localization estimation error.

4.5 Concluding Remarks

This chapter elaborates two different improvisation approaches on conventional Wk-NN fingerprinting. The first approach is intended to address the problem of the time-consuming offline phase of typical Wk-NN fingerprinting localization. However, the second suggested method helps to reduce the physical size of the radio map database, minimizes computational complexity of the

positioning system, and increases localization accuracy.

The proposed method-I combines the traditional Wk-NN fingerprinting with the WC estimation in a pipeline such that the system has similar localization accuracy with reducing the number of RPs as compared to Wk-NN fingerprinting. This approach uses Gaussian filter to estimate RSS in the offline phase whereas on the online phase it uses a moving average filter as a low pass filter. The obtained result by real filed deployment shows that the suggested method can reduce the number of RPs by 40% as compared to the Wk-NN fingerprinting for similar localization estimation error.

The proposed method-II proposes to use multiple features to represent the fingerprint of an RP. Particularly, this approach uses the WC estimation, the rank, and the signal strength of nearby APs that are at the vicinity of the tag device. Moreover, this method also uses APC for RSS clustering where the testbed is clustered into many groups where each group of RPs has a clusterhead. On the online/testing phase, the fingerprinting features are wisely utilized to yield better positioning result. In addition, as the fingerprinting data consists of features extracted from the signal strength of few APs, it helps to significantly reduce the physical size of radio map that eventually minimizes the amount of data communication with the server.

Chapter 5

The Practical Fingerprinting Localization

5.1 Probabilistic Approach of Fingerprinting Localization

The fingerprinting localization technique is predominantly realized in IPS applications owing to its high reliability. However, the learning methodology of radio signals in fingerprinting localization is costly in terms of time-consumption and workload. The Wk-NN fingerprinting localization helps to improve the localization accuracy of the conventional NN fingerprinting localization; however, the problems with learning procedure persists. Meanwhile, the fingerprinting technique can also be implemented using a probabilistic approach that yields better localization result compared to deterministic-based or Wk-NN fingerprinting localization. The probabilistic approach of fingerprinting yields localization information by estimating a probability distribution over the RPs. Here, a matching probability is calculated between the online-observed RSS readings and the pre-stored fingerprinting data in the radio map database.

As it is observed in Figure 3.3 and 3.4 that the statistical distribution of RSS at a particular RP can be thought of as a Gaussian probability distribution, the RSS values should obey the normal distribution $N\ (\mu, \sigma^2)$ where μ and σ^2 are mean and variance of RSS data [75]. Now, the likelihood function ($L(\mu, \sigma^2)$) is given by the following relation:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^{q} \frac{1}{\sqrt{2\pi}\sigma} exp^{-\frac{(RSS_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{\frac{-q}{2}} exp^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{q}(RSS_i-\mu)^2\right)} \end{aligned} \tag{5.1}$$

We can obtain the logarithmic equation of (5.1) as follows:

$$log[L(\mu, \sigma^2)] = -\frac{q}{2}log(2q) - \frac{q}{2}log(\sigma^2) - \frac{q}{2\sigma^2}\sum_{i=1}^{q}(RSS_i - \mu)^2 \tag{5.2}$$

Hence, the likelihood equations can be written as:

$$\begin{aligned} \frac{\partial log[L(\mu, \sigma^2)]}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{q}(RSS_i - \mu)^2 = 0 \\ \frac{\partial log[L(\mu, \sigma^2)]}{\partial \sigma^2} &= -\frac{q}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{q}(RSS_i - \mu)^2 = 0 \end{aligned} \tag{5.3}$$

From (5.3), we get,

$$\mu* = \overline{RSS} = \frac{1}{q}\sum_{i=1}^{q} RSS_i \tag{5.4}$$

$$\sigma^{*2} = \frac{1}{q}\sum_{i=1}^{q}(RSS_i - \mu)^2 \tag{5.5}$$

The unique solution (μ^*, σ^{*2}) of the likelihood equations should also be a local maximum point. In other words, when $|\mu| \rightarrow \infty$ or $\sigma^2 \rightarrow \infty$ or $\sigma^2 \rightarrow 0$, the non-negative function $L(\mu, \sigma^2) \rightarrow 0$. Hence, the maximum likelihood equation of μ and σ^2 will be (5.4) and (5.5), respectively.

Hence, with this idea, the average of RSS reading and its corresponding variance from each AP is calculated for the construction of radio map in the offline phase of the probability-based fingerprinting localization. Later, in the online phase, after acquiring new RSS readings ($RSS_i, i = 1, 2, 3, \ldots, B$) at an unknown location from B APs, we can estimate the probability of the RP (x, y) with respect to i^{th} AP ($P_i(x, y)$) as follows:

$$P_i(x, y) = \frac{1}{\sqrt{2\pi}\sigma_i} exp^{\frac{-(RSS_i - \mu)^2}{2\sigma_i^2}}, \tag{5.6}$$

where μ_i and σ_i^2 are the stored average RSS and its corresponding variance from the ith AP, respectively. This way the probability of each RP can be established where tag's location could be located at the RP that has maximum probability.

5.2 Problem Statement

Although the positioning accuracy of fingerprinting localization improves with a probabilistic approach, this approach suffers from two major problems:

1. The radio map needs to be updated from time to time to achieve good localization estimation owing to the dynamic and unpredictable nature of the radio channel and changes in the surrounding environment [76]. On top of that, the construction of every set of the radio map is time-consuming and labor-intensive.

2. The positioning result of probability-based fingerprinting is better than deterministic based fingerprinting (Wk-NN); however, the computational complexity of the probabilistic approach of fingerprinting is higher than the deterministic approach [59].

These issues on fingerprinting-based localization intensify practical limits and challenges in realizing a reliable and scalable IPS to meet the required accuracy of practical IPS.

The use of regression is proposed to predict the signal strength of the deployed APs at locations with no prior measurements. For computational complexity, the use of clustering is proposed that helps in minimizing the searching space of RPs on the execution phase.

In particular, the use of Gaussian process regression (GPR) as a signal estimation or regression component and APC as a clustering component are proposed. The motivation for using GPR for regression and APC for RSS clustering in this work stems from the facts that GPR not only predicts the mean RSS but also infers the variance at each location and initialization-independent property of APC, respectively.

5.3 Related Work

Although fingerprinting localization yields good localization estimation, an up-to-date radio map database is expected for practical fingerprinting localization. In the conventional fingerprinting localization, radio map was updated by manually collecting the RSS data. The manually acquiring of RSS data increases the human workload and consumes more time. Various approaches have been proposed to reduce the offline workload of fingerprinting localization.

5.3.1 Offline Workload Reduction without Regression

The simultaneous localization and mapping (SLAM) endeavors to eliminate the effort to build the radio map [77, 78]. However, localization estimation error by SLAM (generally 6.7 to 8 m) is not good enough for practical IPS. The study in reference [79] employs a self-guided robot armed with an inertial measurement unit (IMU) sensors to explore the testbed and acquired RSS data. This approach is able to reduce the human workload; however, it is not practically feasible. Other approaches to reducing offline workload are crowdsourcing and a fusion of similarity-based sequence and dead reckoning [80, 81].

5.3.2 Offline Workload Reduction with Regression

The prediction of RSS across the testbed with little training data helps to significantly reduce the human workload. Reference [76] presents a GPR-based fingerprinting IPS using indoor Wi-Fi APs. This work uses a few data points to train the GP where the firefly algorithm is used to estimate the hyperparameters of the GP. Moreover, it also shows that the probabilistic-based localization performs better compared to deterministic-based localization using the predicted radio map. Liu et al. proposed a GPR plus method with Bluetooth transmitters using a naïve Bayes algorithm [82]. They compare their method with [76] and claim that their method is computationally cheaper. Another example of GPR-based fingerprinting is put forward in [83]. This work estimated the hyperparameters by using subspace trust-region method and shows that location estimation with a radio map built using GPR is better than that of Horus [84] fingerprinting method.

5.3.3 Computational complexity reduction with APC

The conventional fingerprinting is also termed as flat fingerprinting and can be converted to two-step fingerprinting using RSS clustering. As the name suggests, the two-step fingerprinting is realized with coarse localization step and fine localization step. Clustering helps to reduce the searching space of RPs on the online phase of fingerprinting, which eventually reduces the computational cost of the system. Moreover, it also helps to reduce the localization estimation error by removing the outliers. APC has widely used clustering technique in IPS owing to its initialization-independent and better cluster head selection characteristics. Many kinds of literature on IPS have employed APC for RSS clustering where their fine localization is either probabilistic-based or deterministic-based [62, 69, 85, 86]. Irrespective to their fine localization methods, all of this literature uses closest RSS distance as a metric to decide the cluster head in their coarse localization step.

5.4 Gaussian Process Regression

A GP is a collection of random variables $f(\mathbf{x})|\mathbf{x} \in \chi$ where any finite subset of realizations of the process $\mathbf{f} = \left\{f(\mathbf{x}_i)\right\}_{i=1}^{n}$ is jointly Gaussian distributed [87]. It is also a generalization of normal multivariate distribution into a finite dimension where it defines a distribution over functions from the view of function space [76, 87]. The GP is fully specified by its mean function $m(x) = \mathbb{E}[f(\mathbf{x})]$ and covariance function $k(\mathbf{x}_m, \mathbf{x}_n) = \mathbb{E}[(f(\mathbf{x}_m) - m(\mathbf{x}_m))(f(\mathbf{x}_n) - m(\mathbf{x}_n))]$. The set χ is the index set of possible input values and $\mathbb{E}(.)$ is the expectation operator.

The GP is utilized to model the relationship between the observed RSS data (y) and RPs locations coordinate ($\mathbf{x}$) considering the following observation model:

$$y = f(\mathbf{x}) + \nu, \tag{5.7}$$

where ν denotes an independent and identically distributed Gaussian noise with zero mean and variance (σ_ν^2), summarized as $\nu \sim N(0, \sigma_\nu^2)$. So the regression problem consists of constraining the non-linear function $f(.)$ from noisy observations that surfaces from σ_ν^2. The interference of $f(\mathbf{x})$ corresponds to estimating the Bayesian posterior distribution given as:

$$p(f(\mathbf{x})|y, X) = \frac{p(y|f(\mathbf{x}), X)p(f(\mathbf{x}))}{p(y|\mathbf{x})}, \tag{5.8}$$

where $p(y|f(\mathbf{x}), X)$ denotes the probability of obtaining the y given the function $f(\mathbf{x})$, which is also known as the likelihood function. Note that $p(f(\mathbf{x}))$ represents the prior distribution of the latent function and itself is a GP. The GPR utilizes the covariance of neighboring RSSs to predict the RSSs at arbitrary locations. The covariance functions are based on kernel functions. The kernel function decreases with increased distance of the input points and peaks when the distance

in input space is minimal. Some of the common kernel functions are Gaussian, Laplacian, and Exponential where localization accuracy is relatively insensitive to the choice of these functions [84]. The dissertation uses the most commonly adopted squared exponential kernel in this study:

$$k(\mathbf{x}_m, \mathbf{x}_n) = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x}_m - \mathbf{x}_n\|}{2l^2}\right), \tag{5.9}$$

where σ_f^2 is the signal variance and l is the length scale parameter. Let us denote the covariance matrix of all pairs of training targets as $K(X, X)$. Owing to the noisy RSS observations, the covariance function of the prior distribution $k(y) = K(X, X) + \sigma_\nu^2 I$, where I is an identity matrix.

The target values (y) of the training data are correlated among each other, as they are also correlated with any random target values. It enables to predict RSSs as $\boldsymbol{y}^* \equiv f(X^*)$. Since GP modeling assumes that the data can be represented as a sample from a multivariate Gaussian distribution, the joint Gaussian distribution of target values assuming zero mean prior distribution is given as:

$$\begin{bmatrix} y \\ \boldsymbol{y}^* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_\nu^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right), \tag{5.10}$$

where $K(X, X^*)$ is a $N \times N^*$ matrix containing the covariance between all pairs of y and $\boldsymbol{y}^*$. Here, y and $\boldsymbol{y}^*$ are N training and N^* test RSS data, respectively. The conditional probability of $\boldsymbol{y}^*|y$ is computed as:

$$\begin{aligned} P(\boldsymbol{y}^*|y) \sim N\Big(&K(X^*, X)[K(X,X) + \sigma_\nu^2 I]^{-1} y), \\ &K(X^*, X^*) - K(X^*, X)[K(X,X) + \sigma_\nu^2 I]^{-1} K(X, X^*)\Big) \end{aligned} \tag{5.11}$$

Note that, $P(\boldsymbol{y}^*|y)$ indicates how likely a prediction $\boldsymbol{y}^*$ is, given the training data y. From (5.11), the mean function and the covariance function can be inferred as follows:

$$\overline{\boldsymbol{y}^*} = K(X^*, X)[K(X,X) + \sigma_\nu^2 I]^{-1} \mathbf{y} \tag{5.12}$$

$$var\big(f(X)\big) = K(X^*, X^*) - K(X^*, X)[K(X,X) + \sigma_\nu^2 I]^{-1} K(X, X^*) \tag{5.13}$$

The principle structure of the GP model is determined by the mean and covariance functions, however, to fit the observations properly, the function's optimal hyperparameters, $\boldsymbol{\theta} = [\sigma_f, l]$ in (5.9) need to be established. There are various ways of inferring the GP hyperparameters such as marginal likelihood, Bayesian optimization, and cross-validation [88, 89, 90]. Marginal likelihood is the optimal and computationally efficient when the data truly follows the GP model. In addition, marginal likelihood is implemented with approaches like Maximum a Posteriori estimator and

Minimum Mean Square Error estimator. The logarithmic form of the marginal likelihood function for Gaussian distributed noise is as follows:

$$log\big(p(y|X,\boldsymbol{\theta})\big) = -\frac{1}{2}y^T\Big[K(X,X)+\sigma_\nu^2 I\Big]^{-1}y - \frac{1}{2}log|K(X,X)+\sigma_\nu^2 I| \\ -\frac{\nu}{2}log2\pi, \tag{5.14}$$

where T denotes the matrix transpose. The optimization problem was solved employing the limited memory BFGS-B algorithm (LM-BFGS) [91, 92]. Note that LM-BFGS is an optimization algorithm in the family of quasi-newton methods, which approximates the BFGS [93] algorithm using the limited amount of system memory. In addition, LM-BFGS is the widely used algorithm for parameter estimation in machine learning.

5.5 Proposed Practical Fingerprinting Method

The suggested practical fingerprinting method is a lightweight workload fingerprinting localization that uses APC for clustering and GPR for regression. The proposed approach uses very little training data to model the GP that estimates the RSS at locations without prior measurements. The prediction of RSS data across the testbed helps to significantly reduce the time consumption and workload for construction of the radio map database. In addition, the fingerprinting database constructed by limited training data is further divided into sub-areas or clusters applying APC.

5.5.1 Process Flow of the Proposed Practical Positioning Approach

The framework of the proposed method that illustrates the operations at offline and online phases is presented in Figure 5.1.

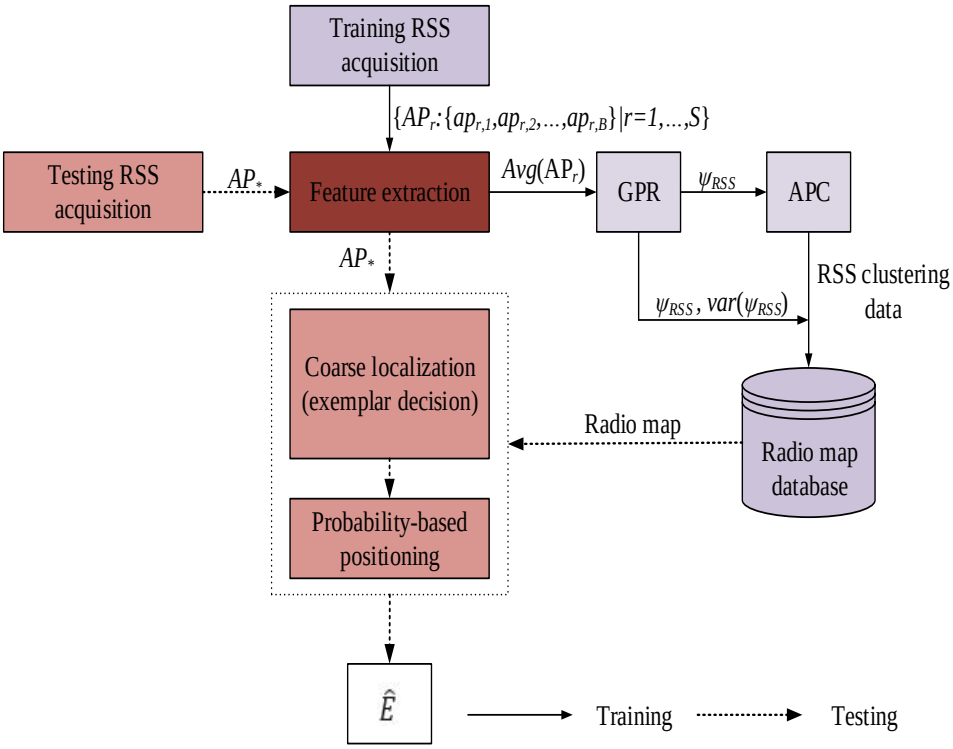


Figure 5.1: Framework of the proposed positioning method.

At first, the training data from a few sparse S RPs (that are subset to the total N RPs on the testbed, $S \subset N$ and $S < N$) are collected on the offline phase. The RSS of B APs (B BLE beacons on the testbed) are acquired at the rth RP ($r \in (1, 2, \ldots, N)$) q times, that is AP_r= $(ap_{r,1}, ap_{r,2}, ap_{r,3}, ..., ap_{r,B})$, where $ap_{r,1} = Avg(ap^1_{r,i}, ap^2_{r,i}, \ldots, ap^q_{r,i})$. Here, the average value of q RSS readings are utilized to minimize the time varying fluctuation of the RSS data. The acquired training data is used to model the GPR framework. The predicted mean RSS data from the GPR is further preprocessed with APC to cluster it into different cluster-centers. Hence, the radio map consists of the predicted mean RSS, its corresponding variances, and the clustering information.

On the online phase, the observed RSS data at an unknown location inside the testbed can be represented as: $AP_* = (ap_{*1}, ap_{*2}, \ldots, ap_{*B})$. Here, each RP is assigned with a probability computed as the probability of RP_m, given the online RSS observation vector AP_*.

$$P_r(RP_m|AP_*) = \frac{Pr(AP_*|RP_m)P_r(RP_m)}{\sum_{n=1}^{N} P_r(AP_*|RP_n)P_r(RP_n)}, \tag{5.15}$$

where $Pr(AP_*|RP_m)$ is the likelihood defined by the following relationship:

$$Pr(AP_*|RP_m = \prod_{b=1}^{B} \frac{1}{\sqrt{2\pi\sigma_b^2}} exp^{\left(\frac{-|AP_b-\mu_b|}{2\sigma_b^2}\right)} \tag{5.16}$$

Note that relation (5.16) is derived from (5.6) owing to the fact that the RSS information from the deployed APs is independent of each other. Moreover, μ_b and σ_b^2 in (5.16) are the predicted mean RSS and its corresponding variance of AP b signal at the RP location, respectively.

Let us define Δ and N' as the set of cluster-centers and the number of RPs that are grouped under a cluster center, respectively. Here, N' may vary with the cluster, where $N' < N$. At first, coarse localization is performed to find the exemplar (cluster-head) RP. The exemplar RP (RP_e) is determined in two different ways in the proposed approach as follows:

$$RP_e = argmax_m[P_r(RP_m|AP_*)],\ m \in \Delta \tag{5.17}$$

$$RP_e = argmin_{m\in\Delta}\left\|\overrightarrow{\psi_r} - \overrightarrow{\psi_m}\right\|^2,\ \overrightarrow{\psi_r} = [\psi_{1,r}, \psi_{2,r}, \ldots, \psi_{B,r}]^T \tag{5.18}$$

where $\overrightarrow{\psi_r}$ and $\overrightarrow{\psi_m}$ are the stored and online observed RSS vectors, respectively. In (5.17), given the online-acquired RSS data, any RP in Δ that has the largest posterior probability is chosen as the cluster head, whereas in (5.18), the RP in Δ that has the least RSS distance with the online-observed RSS data is determined as the cluster head.

Finally, the fine localization estimation is estimated as a weighted sum of N' RPs with their respective probabilities as follows:

$$\hat{E} = \sum_{i=1}^{N'} P_r(RP_i|AP_*) \times Rp_i \tag{5.19}$$

5.5.2 Experimental Setup

The training data were obtained on the third floor of the IT building, Chosun University. The testbed is an academic area with frequent movement of people along the hallways. The graphical representation of the testbed is illustrated in Figure 5.2.

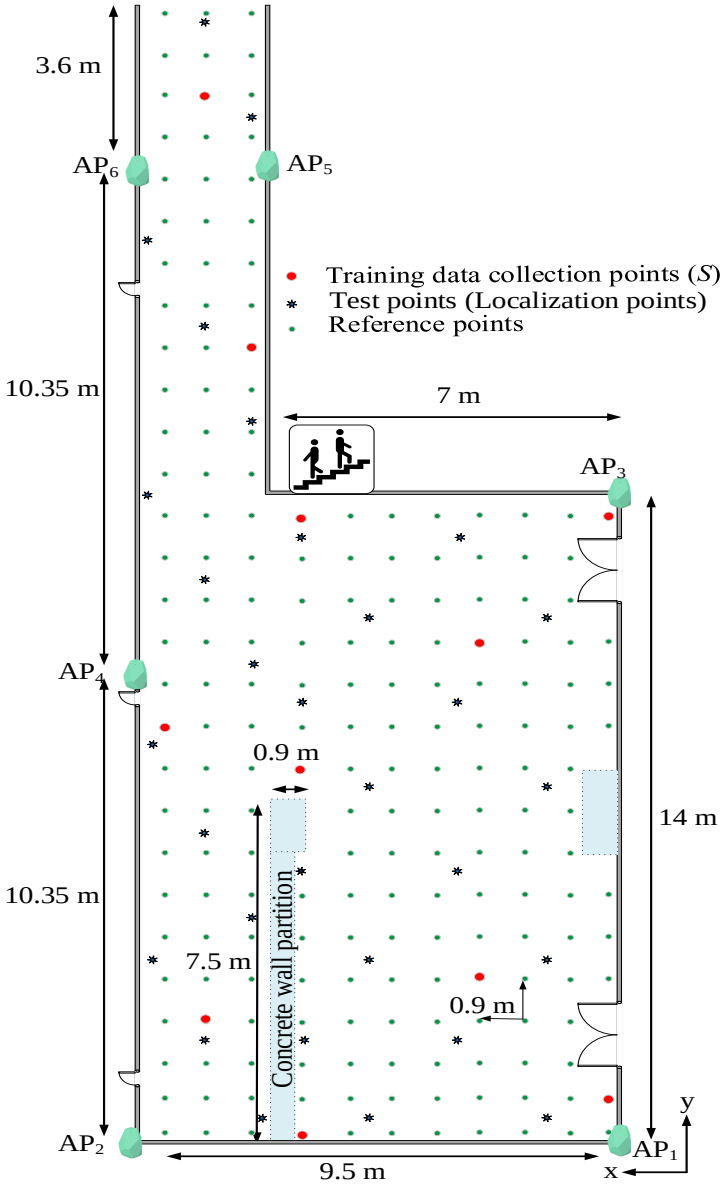


Figure 5.2: Graphical representation of the testbed.

Six APs (BLE beacons) were deployed to cover the whole range of the testbed area where each AP was set to have an advertisement interval of 300 ms and transmission power of $+4$ dBm.

The testbed was divided into hypothetical uniform grids with side measurements of 0.9 to form 203 RPs. An iOS application on $iPhone - 6S$ was used to acquire the RSS data from the deployed APs. The training data were collected at 11 randomly selected sparse locations

($S = 11$) where 10-time samples of RSS in four different directions ($q = 40$) were recorded at each measurement location. For the locations with undetected APs or null reading from the BLE beacons, a default value ($invalid\ data$) of -95 dBm was used [69, 76].

5.5.3 Experimental Results and Discussion

RSS Prediction using GPR

The GPR was employed to build the posterior mean RSS and its variance at each of the RPs on the testbed. To compare the accuracy of the predicted RSS data, RSS data was manually measured across all the RPs on the testbed. Figure 5.3 presents the predicted mean RSS as well as the manually recorded RSS data of AP_4 (see Figure 5.2) at the testbed. Similarly, Figure 5.4 presents the corresponding standard deviation of the predicted mean RSS.

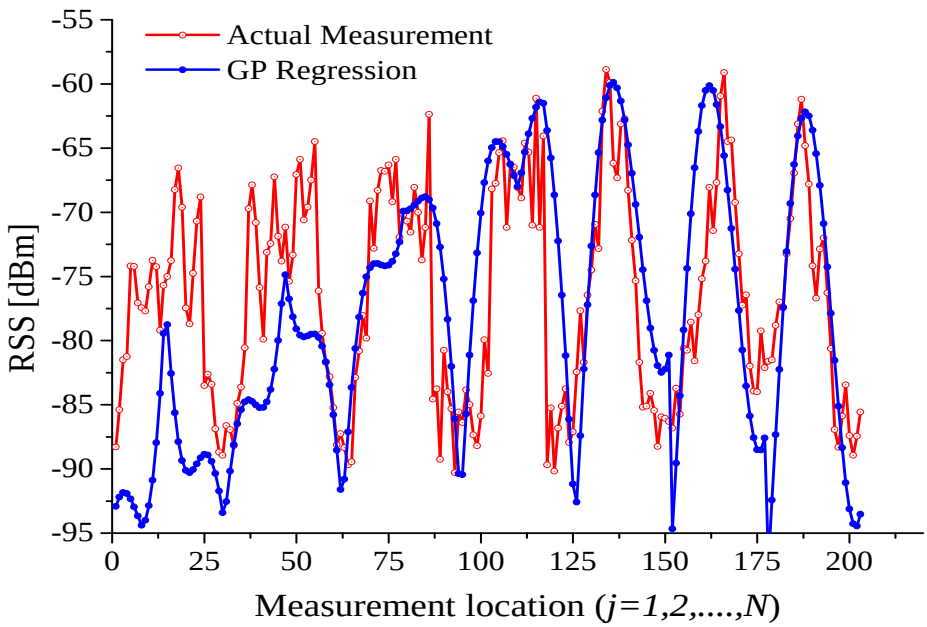


Figure 5.3: Comparison of predicted and manually measured RSS of AP_4 at the measurement locations.

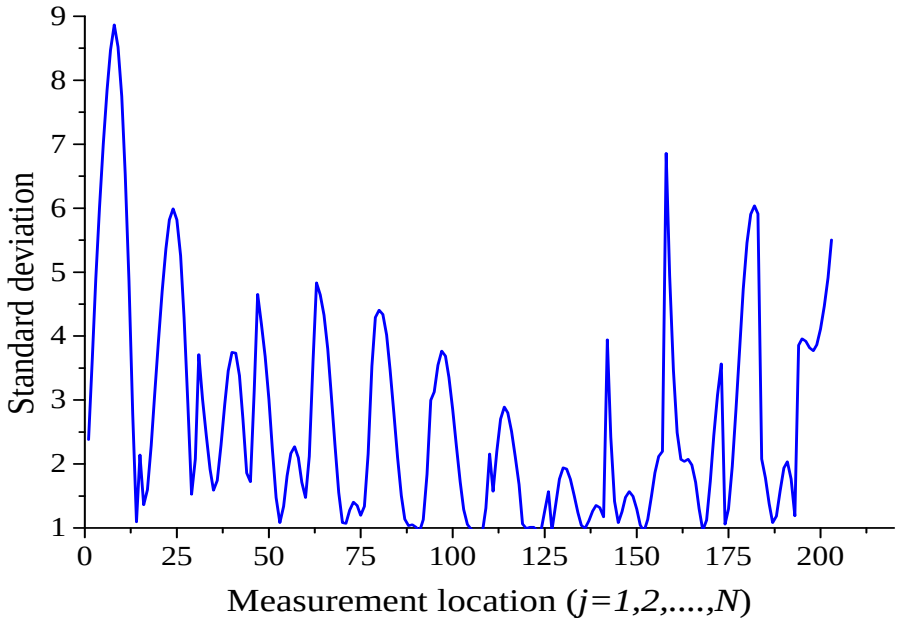


Figure 5.4: Corresponding standard deviation of the predicted RSS of AP_4 at the measurement locations.

The comparison of the predicted and manually acquired RSS shows that the predicted RSS is almost similar to the real measured data. In particular, the mean difference between the predicted and manually measured RSS data was 6.50 dBm. The surface plot of the predicted RSS of AP_4 on the testbed is shown in Figure 5.5. The surface in red on Figure 5.5 depicts the region on the testbed close to the true location of AP_4. Similarly, the surface plot of the corresponding standard deviation is presented in Figure 5.6. The surface in blue color in Figure 5.6 indicates that the measurement (training) data are available in the surroundings whereas the red/yellow surface indicates the lack of measurement data around it.

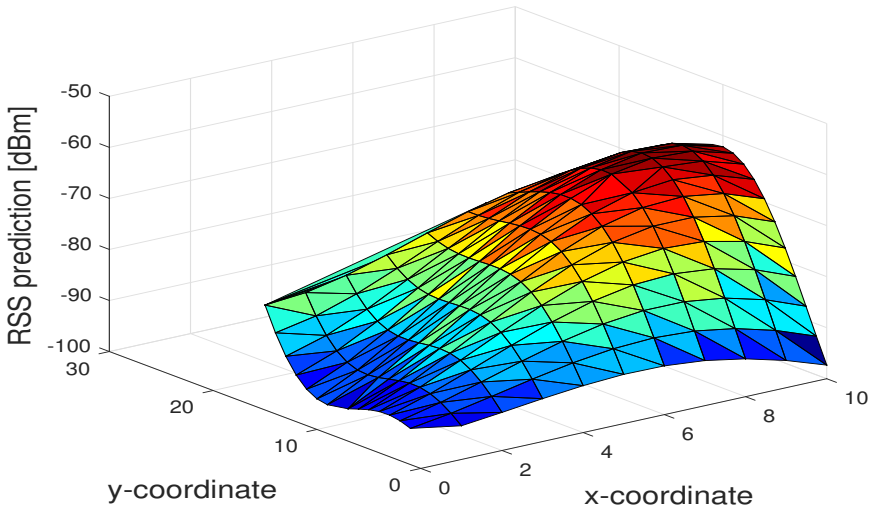


Figure 5.5: Surface plot of the predicted RSS of AP_4 on the testbed.

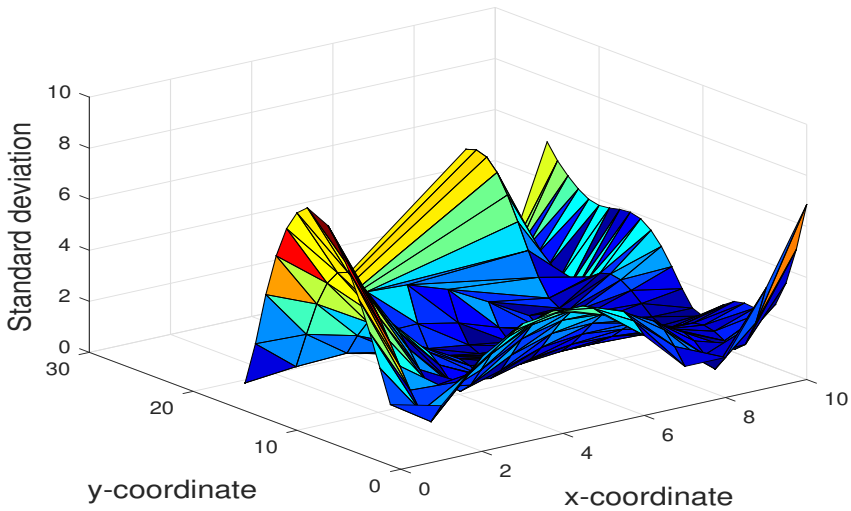


Figure 5.6: Surface plot of the corresponding standard deviation of the predicted RSS of AP_4 on the testbed.

The proposed method is able to minimize the time-consuming and labor-intensive problems of conventional fingerprinting localization. For an instant, the GP model was trained using training data from 11 measurement places to populate the radio map database for the entire testbed (203 RPs) environment.

RSS Clustering using APC

The predicted mean RSS obtained from the GPR operation was fed to the APC for RSS clustering. The results of the APC operation with varying numbers of APs on the testbed is presented in Figure 5.7 and 5.8.

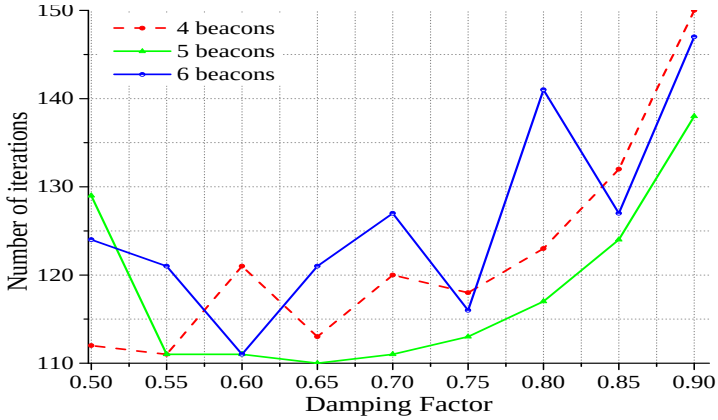


Figure 5.7: The variation on the number of iterations with respect to the damping factor.

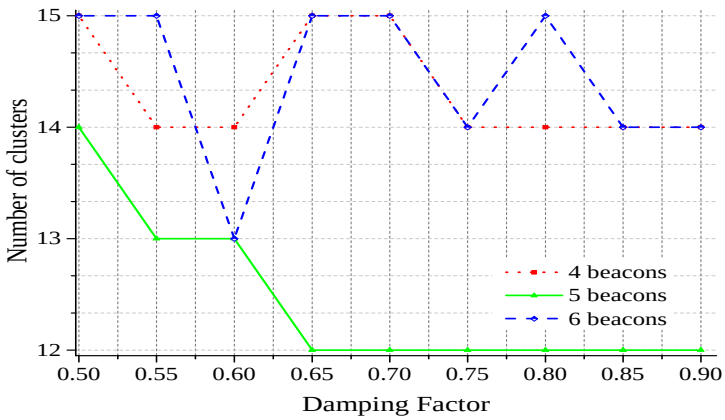


Figure 5.8: The variation on the number of clusters with respect to the damping factor.

The minimum number of iteration was observed at $\gamma = 0.6$ for six APs. Since the minimum number of iterations also signifies the minimum computational cost, γ with the lowest number of iterations was chosen for each size of APs deployment.

Note that the computational complexity of conventional probabilistic approach of the fingerprinting method is $O(NB)$, where N is the total number of fingerprints to be compared and B is the total number of deployed APs in a given indoor environment. It shows that the computational complexity increases with the size of N. However, in the proposed method, the coarse localiza-

tion stage reduces the area of interest from N RPs to $N'(N' << N)$, which reduces the total number of fingerprints to be compared for fine localization. The computational complexity of the proposed method becomes $O(N'B)$ owing to APC clustering that minimizes the searching space of RPs and reduces computational complexity.

Localization Performance of the Proposed Method

The proposed method is evaluated under the deployment of different numbers of APs on the testbed ($4, 5$, and 6 APs). For the deployment scenarios of four and five APs in the testbed, (AP_4, AP_5) and (AP_5) were removed from the testbed, respectively (see Figure 5.2). Both decision rules were implemented for the cluster-head estimation given by (5.17) and (5.18). The suggested method was compared with a typical GPR-based probabilistic method [83] and the Hours method [75]. Note that the Horus method estimates the location of the tag device that has the largest posterior probability through Bayesian interference [94]. In addition, the Horus method uses a clustering module where any cluster is a set of RPs sharing a common set of Wi-Fi APs. In the experiment, APC was applied to the manually measured RSS data for the Horus method where (5.18) determines the cluster-head. Figure 5.9 presents the CDF of the localization estimation error where only the Horus method uses the manually constructed radio map database and other methods under comparison use the predicted RSS data.

As presented in Figure 5.9, the proposed method using RSS distance as the cluster-head decision rule has a probability of 48% that the localization error is below 2 m. Similarly, for the typical GPR and Horus method, the corresponding probabilities are 15% and 34%, respectively. The proposed method utilizing the RSS distance given by (5.18) as the exemplar decision rule performed better compared to the largest posterior probability given by (5.17). The use of (5.18) is recommended for the proposed method, whereas (5.17) can be used when (5.18) is not available.

Furthermore, note that approximately 7.5% of the localization estimation error in above 4.5 m in the proposed method owing to the following reasons:

1. Wrong exemplar estimation in the online phase: Due to the random fluctuation in the received RSS, the positioning system may lead to the estimation of an RP as an exemplar that is physically far away from the real position of the tag device. Since the RPs grouped under the cluster-head are responsible for the final localization estimation, the wrong estimation of the exemplar eventually adds some localization estimation error.

2. The outliers at the preprocessing in the offline phase: Although APC has initialization-independent property, the clustering algorithm might lead to an RP that belongs to a cluster but is physically away from the cluster-head [62]. This problem can be minimized by considering the benefit of the known position of each RP where each outlier RP can be forced to join the cluster characterized by the exemplar at the minimum distance from the outlier itself.

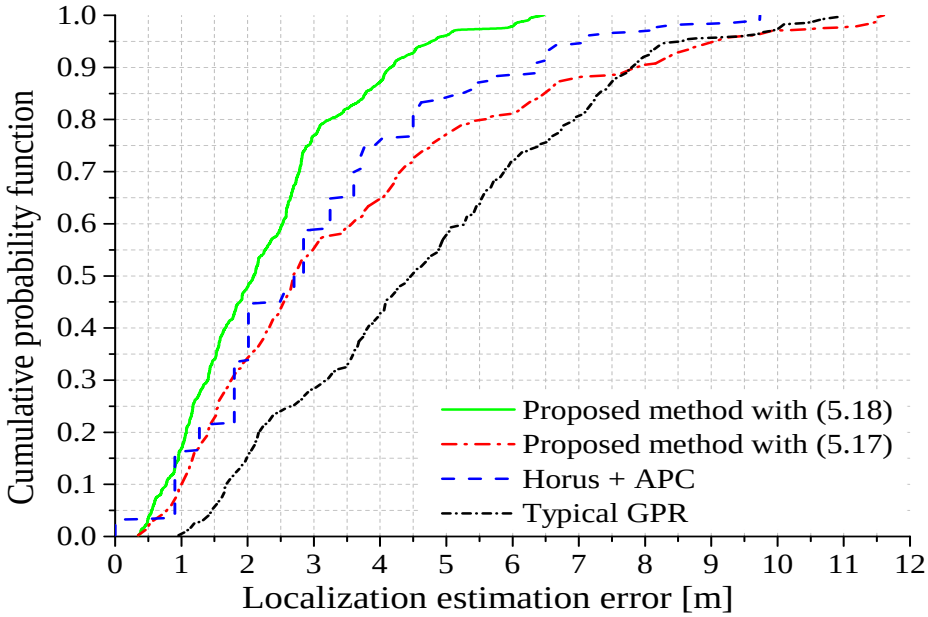


Figure 5.9: Cumulative probability function of the localization estimation error.

The average error of localization estimation with different numbers of APs deployment on the testbed is shown in Figure 5.10.

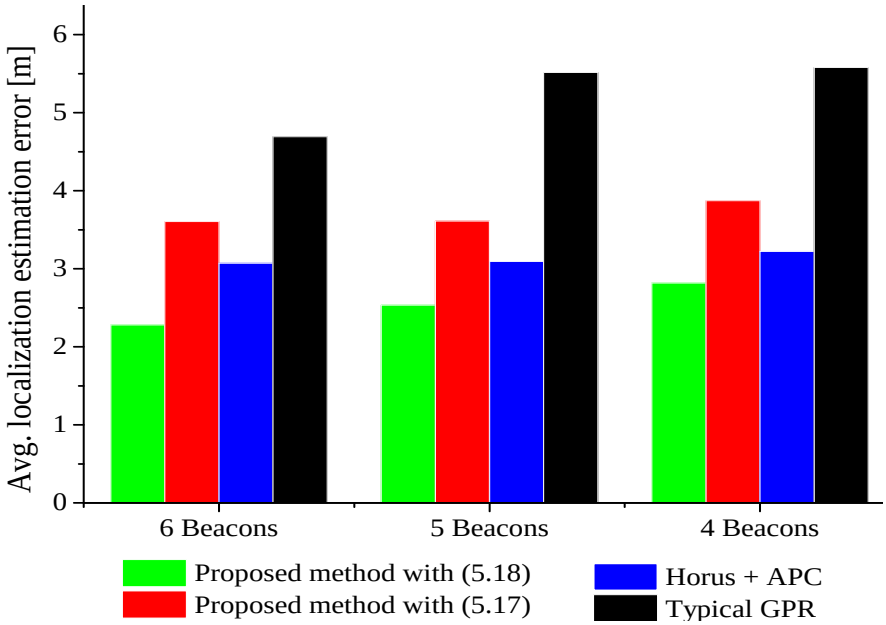


Figure 5.10: Average localization estimation error from different positioning methods at different AP density scenario.

Figure 5.10 shows that the decrease in AP density at the testbed environment increases the localization estimation error. The suggest IPS approach outperforms the existing methods in terms of localization accuracy and computational complexity.

5.6 Concluding Remarks

This chapter proposes the practical fingerprinting localization technique that is able to solve the two major problems of the conventional probabilistic fingerprinting localization system. First, the suggested method proposes to use a regression approach to predict the RSS information for the whole testbed utilizing very little training data. It helps to reduce the workload of the painstaking offline procedure. Second, it suggests using clustering in order to reduce computational complexity. Hence, the regression and clustering approaches are combined in IPS to address the real issues of the conventional fingerprinting localization.

The GPR was implemented as a regression component and APC as a clustering component. GPR predicts not only mean RSS but also the corresponding variances that are helpful in the probabilistic approach of fingerprinting localization. In addition, APC is initialization independent and performs better clustering results. Hence, the APC algorithm was used along with a machine learning approach of fingerprinting localization.

The mathematical model of GPR and APC are elaborated and their respective results of RSS prediction and clustering are illustrated and analyzed in this chapter. The positioning result of the proposed method was compared with an existing probability based fingerprinting methods. Furthermore, the positioning methods were evaluated with different numbers of AP deployment environment.

Chapter 6

Conclusions

In this dissertation, some novel approaches are presented to improvise the conventional fingerprinting localization and solve the real issues of fingerprinting localization system. Since GNSS cannot perform at indoor, various alternative technologies have been used for IPS development. The BLE beacons are utilized to realize the proposed approaches owing to the fact that it consumes less energy and is embedded in many latest smartphones and tablets. Although the fingerprinting method has become a prime choice in the design of IPS owing to its good localization accuracy, its offline workload and online computational complexity grow with the size of the testbed area. The proposed methods address these problems and minimize the practical limits in realizing reliable IPS. In what follows, a summary of those approaches in the order they appear in the dissertation.

In Chapter 4, the process flow of the conventional Wk-NN fingerprinting localization is presented. The working procedure at the online and offline phases of the Wk-NN fingerprinting localization are illustrated in detail. The first proposed method is intended to reduce the offline workload of a typical Wk-NN fingerprinting approach. The suggested method takes advantage of WC localization and uses it twice in a pipeline to perform localization estimation. The major concept of this approach is to use less number of RPs with large space between them, yet yielding the localization result comparable to Wk-NN fingerprinting approach. For this, the method estimates the WC of the nearby APs and k-NN RPs (lightly populated RPs). Now, second WC estimation is performed utilizing the first WC estimation and the k-NN RPs. Hence, the proposed approach is able to minimize the number of RPs on the testbed that eventually reduces the offline workload. The experimental results with real field deployment at two different indoor environments show that the proposed method reduces the total number of RPs by 41.49% at the corridor and 49.23% at the computer lab.

The second proposed technique in Chapter 4 tries to address the problems at the online phase of the typical Wk-NN fingerprinting localization. The suggested method improves the localization accuracy, minimizes the computational cost, and reduces the physical size of the fingerprinting radio map database. To achieve this improvisation, it uses multiple features to define a fingerprint of an RP and employs RSS clustering. This approach also takes advantage of WC localization as in the first approach but uses it like a fingerprint. Hence, the fingerprint information at an RP contains the WC estimation from nearby APs and the ranks and signal strength of those APs. Note that the fingerprint information is generated from the signals from the nearby APs only. Hence, it helps to significantly reduce the physical size of the radio map database. On the online phase, coarse localization is performed to determine the cluster-head. Here, the online WC estimate is

compared with the stored one to decide the cluster-head. The proportional weight is given to the WC-based and RSS-based Wk-NN approaches for the final localization estimation. The observed experimental results show that the proposed method can reduce the database size by 69.7% in the testbed environment. The proposed method also outperforms the existing Wk-NN fingerprinting in terms of localization accuracy and computational time.

Finally, in Chapter 5, the third contribution of the dissertation is presented. The fingerprinting localization can also be realized using a probabilistic approach. Here, the mean RSS and its standard deviation of each at the RPs is stored in the radio map database. On the online phase, a probability distribution over the RPs is estimated employing the observed RSS and stored RSS and variance. The localization estimation by probabilistic-based fingerprinting localization is better than the deterministic-based or Wk-NN fingerprinting localization. However, the computational complexity increases in the probabilistic approach of fingerprinting. Therefore, using the probabilistic approach of fingerprinting localization helps to improvise the localization accuracy with an increased online computational cost. Furthermore, the major issue with the conventional fingerprinting localization is the effort seeking offline radio map construction phase. A novel positioning method is proposed that not only reduces the offline workload but also minimizes the computational complexity of the probabilistic-based fingerprinting method.

The use of regression is proposed for prediction of RSS data at the locations without prior measurements. In particular, very little training data is used to train a GP model, which is able to predict the RSS and its variance at all the RPs on the testbed. For the computational cost, APC is used to extract the RSS clustering information as in Chapter 4. Hence, a combination of GPR and APC is proposed to solve the real issues that are cemented on the conventional fingerprinting localization. The proposed method is evaluated along with the typical GPR-based method and the Horus with APC through real field deployment. The experimental results show that the proposed method can minimize the offline workload and increase localization accuracy with a decreased computational cost.

Bibliography

[1] "GPS: The Global Positioning System," https://www.gps.gov/, accessed: 2019-04-04.

[2] "Information analytical centre of GLONASS and GPS controlling," https://www.glonass-iac.ru/en/, accessed: 2019-04-04.

[3] "European GNSS Service Centre — European GNSS Service Centre," https://www.gsc-europa.eu/, accessed: 2019-04-04.

[4] "BeiDou Navigation Satellite System," http://en.beidou.gov.cn/, accessed: 2019-04-04.

[5] M. Ibnkahla, *Wireless sensor networks: a cognitive perspective*. Crc Press, 2016.

[6] M. E. Rusli, M. Ali, N. Jamil, and M. M. Din, "An improved indoor positioning algorithm based on rssi-trilateration technique for internet of things (iot)," in *2016 International Conference on Computer and Communication Engineering (ICCCE)*. IEEE, 2016, pp. 72–77.

[7] S. Subedi, G.-R. Kwon, S. Shin, S.-s. Hwang, and J.-Y. Pyun, "Beacon based indoor positioning system using weighted centroid localization approach," in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2016, pp. 1016–1019.

[8] T. T. Brooks, H. H. Bakker, K. Mercer, and W. Page, "A review of position tracking methods," in *1st International conference on sensing technology*, 2005, pp. 54–59.

[9] S. P. Subramanian, J. Sommer, S. Schmitt, and W. Rosenstiel, "Ril—reliable rfid based indoor localization for pedestrians," in *2008 16th International Conference on Software, Telecommunications and Computer Networks*. IEEE, 2008, pp. 218–222.

[10] B.-S. Choi, J.-W. Lee, and J.-J. Lee, "Localization and map-building of mobile robot based on rfid sensor fusion system," in *2008 6th IEEE International Conference on Industrial Informatics*. IEEE, 2008, pp. 412–417.

[11] ——, "An improved localization system with rfid technology for a mobile robot," in *2008 34th Annual Conference of IEEE Industrial Electronics*. IEEE, 2008, pp. 3409–3413.

[12] R. Tesoriero, J. A. Gallud, M. D. Lozano, and V. M. Penichet, "Tracking autonomous entities using rfid technology," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 650–655, 2009.

[13] A. D. Koutsou, F. Seco, A. R. Jiménez, J. O. Roa, J. L. Ealo, C. Prieto, and J. Guevara, "Preliminary localization results with an rfid based indoor guiding system," in *2007 IEEE International Symposium on Intelligent Signal Processing*. IEEE, 2007, pp. 1–6.

[14] H. Chae and K. Han, "Combination of rfid and vision for mobile robot localization," in *2005 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE, 2005, pp. 75–80.

[15] D. Hepeng and S. Donglin, "Indoor location system using rfid and ultrasonic sensors," in *2008 8th International Symposium on Antennas, Propagation and EM Theory*. IEEE, 2008, pp. 1179–1181.

[16] C.-S. Wang and L.-C. Cheng, "Rfid & vision based indoor positioning and identification system," in *2011 IEEE 3rd International Conference on Communication Software and Networks*. IEEE, 2011, pp. 506–510.

[17] B. Ozdenizci, K. Ok, V. Coskun, and M. N. Aydin, "Development of an indoor navigation system using nfc technology," in *2011 Fourth International Conference on Information and Computing*. IEEE, 2011, pp. 11–14.

[18] B. Ozdenizci, V. Coskun, and K. Ok, "Nfc internal: An indoor navigation system," *Sensors*, vol. 15, no. 4, pp. 7571–7595, 2015.

[19] Y. Luo and C. L. Law, "Indoor positioning using uwb-ir signals in the presence of dense multipath with path overlapping," *IEEE Transactions on wireless communications*, vol. 11, no. 10, pp. 3734–3743, 2012.

[20] D.-H. Kim, G.-R. Kwon, J.-Y. Pyun, and J.-W. Kim, "Nlos identification in uwb channel for indoor positioning," in *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2018, pp. 1–4.

[21] A. Waadt, S. Wang, C. Kocks, A. Burnic, D. Xu, G. H. Bruck, and P. Jung, "Positioning in multiband ofdm uwb utilizing received signal strength," in *2010 7th Workshop on Positioning, Navigation and Communication*. IEEE, 2010, pp. 308–312.

[22] S. Subedi, S. Das, and N. S. V. Shet, "Dynamic spectrum allocation in wireless sensor networks."

[23] K. Sohraby, D. Minoli, and T. Znati, *Wireless sensor networks: technology, protocols, and applications*. John Wiley & Sons, 2007.

[24] M. Siekkinen, M. Hiienkari, J. K. Nurminen, and J. Nieminen, "How low energy is bluetooth low energy? comparative measurements with zigbee/802.15. 4," in *2012 IEEE wireless communications and networking conference workshops (WCNCW)*. IEEE, 2012, pp. 232–237.

[25] E. Dahlgren and H. Mahmood, "Evaluation of indoor positioning based on bluetooth smart technology," *Master of Science Thesis in the Programme Computer Systems and Networks*, 2014.

[26] R. Faragher and R. Harle, "Location fingerprinting with bluetooth low energy beacons," *IEEE journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, 2015.

[27] A. Corbacho Salas, "Indoor positioning system based on bluetooth low energy," B.S. thesis, Universitat Politècnica de Catalunya, 2014.

[28] G. G. Anagnostopoulos and M. Deriaz, "Accuracy enhancements in indoor localization with the weighted average technique," *SENSORCOMM*, vol. 2014, pp. 112–116, 2014.

[29] B. Sklar, "Rayleigh fading channels in mobile digital communication systems. i. characterization," *IEEE Communications magazine*, vol. 35, no. 7, pp. 90–100, 1997.

[30] K. Kaemarungsi, "Design of indoor positioning systems based on location fingerprinting technique," Ph.D. dissertation, University of Pittsburgh, 2005.

[31] Z. Jianyong, L. Haiyong, C. Zili, and L. Zhaohui, "Rssi based bluetooth low energy indoor positioning," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2014, pp. 526–533.

[32] R. G. Brown and R. F. Meyer, "The fundamental theorem of exponential smoothing," *Operations Research*, vol. 9, no. 5, pp. 673–685, 1961.

[33] M. Heidari and K. Pahlavan, "Identification of the absence of direct path in toa-based indoor localization systems," *International Journal of Wireless Information Networks*, vol. 15, no. 3-4, pp. 117–127, 2008.

[34] C. Yang and H.-R. Shao, "Wifi-based indoor positioning," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 150–157, 2015.

[35] S. S. Chawathe, "Beacon placement for indoor localization using bluetooth," in *2008 11th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2008, pp. 980–985.

[36] ——, "Low-latency indoor localization using bluetooth beacons," in *2009 12th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2009, pp. 1–7.

[37] A. Montaser and O. Moselhi, "Rfid indoor location identification for construction projects," *Automation in Construction*, vol. 39, pp. 167–179, 2014.

[38] Y. Tian, "Practical indoor localization system using gsm fingerprints and embedded sensors," Ph.D. dissertation, Université Pierre et Marie Curie-Paris VI, 2015.

[39] H.-H. Liu and C. Liu, "Implementation of wi-fi signal sampling on an android smartphone for indoor positioning systems," *Sensors*, vol. 18, no. 1, p. 3, 2018.

[40] H. Zhu and T. Alsharari, "An improved rssi-based positioning method using sector transmission model and distance optimization technique," *International Journal of Distributed Sensor Networks*, vol. 11, no. 9, p. 587195, 2015.

[41] S.-H. Lee, I.-K. Lim, and J.-K. Lee, "Method for improving indoor positioning accuracy using extended kalman filter," *Mobile Information Systems*, vol. 2016, 2016.

[42] X. Huang, M. Barralet, and D. Sharma, "Accuracy of location identification with antenna polarization on rssi," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1. Citeseer, 2009.

[43] G. Li, E. Geng, Z. Ye, Y. Xu, J. Lin, and Y. Pang, "Indoor positioning algorithm based on the improved rssi distance model," *Sensors*, vol. 18, no. 9, p. 2820, 2018.

[44] M. Ayadi, S. Tabbane, and Z. Belhaj, "Calibration of propagation model for indoor tunisian environment," in *Proc. SETIT*, 2005, pp. 27–31.

[45] J. J. Treurniet, C. Sarkar, R. V. Prasad, and W. De Boer, "Energy consumption and latency in ble devices under mutual interference: An experimental study," in *2015 3rd International Conference on Future Internet of Things and Cloud*. IEEE, 2015, pp. 333–340.

[46] G. Welch, G. Bishop *et al.*, "An introduction to the kalman filter," 1995.

[47] V. Apte, Y. A. Powar *et al.*, "Improving the accuracy of wireless lan based location determination systems using kalman filter and multiple observers," in *IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006.*, vol. 1. IEEE, 2006, pp. 463–468.

[48] G. A. Terejanu, "Discrete kalman filter tutorial," *University at Buffalo, Department of Computer Science and Engineering, NY*, vol. 14260, 2013.

[49] G. Li, E. Geng, Z. Ye, Y. Xu, J. Lin, and Y. Pang, "Indoor positioning algorithm based on the improved rssi distance model," *Sensors*, vol. 18, no. 9, p. 2820, 2018.

[50] C. Zhou, J. Yuan, H. Liu, and J. Qiu, "Bluetooth indoor positioning based on rssi and kalman filter," *Wireless Personal Communications*, vol. 96, no. 3, pp. 4115–4130, 2017.

[51] J. Hu, D. Liu, Z. Yan, and H. Liu, "Experimental analysis on weight k-nearest neighbor indoor fingerprint positioning," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 891–897, 2019.

[52] P. Kriz, F. Maly, and T. Kozel, "Improving indoor localization using bluetooth low energy beacons," *Mobile Information Systems*, vol. 2016, 2016.

[53] 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.

[54] R. Ma, Q. Guo, C. Hu, and J. Xue, "An improved wifi indoor positioning algorithm by weighted fusion," *Sensors*, vol. 15, no. 9, pp. 21 824–21 843, 2015.

[55] C. Li, Z. Qiu, and C. Liu, "An improved weighted k-nearest neighbor algorithm for indoor positioning," *Wireless Personal Communications*, vol. 96, no. 2, pp. 2239–2251, 2017.

[56] S. Lee, B. Cho, B. Koo, S. Ryu, J. Choi, and S. Kim, "Kalman filter-based indoor position tracking with self-calibration for rss variation mitigation," *International Journal of Distributed Sensor Networks*, vol. 11, no. 8, p. 674635, 2015.

[57] P. Barsocchi, S. Chessa, A. Micheli, and C. Gallicchio, "Forecast-driven enhancement of received signal strength (rss)-based localization systems," *ISPRS International Journal of Geo-Information*, vol. 2, no. 4, pp. 978–995, 2013.

[58] X. Li, J. Wang, and C. Liu, "A bluetooth/pdr integration algorithm for an indoor positioning system," *Sensors*, vol. 15, no. 10, pp. 24 862–24 885, 2015.

[59] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern wlan fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1974–2002, 2017.

[60] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[61] E. Gokcay and J. C. Principe, "Information theoretic clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 2, pp. 158–171, 2002.

[62] G. Caso, L. De Nardis, and M.-G. Di Benedetto, "A mixed approach to similarity metric selection in affinity propagation-based wifi fingerprinting indoor positioning," *Sensors*, vol. 15, no. 11, pp. 27 692–27 720, 2015.

[63] M. Mézard, "Passing messages between disciplines," *Science*, vol. 301, no. 5640, pp. 1685–1686, 2003.

[64] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.

[65] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[66] P. Sadhukhan, "Performance analysis of clustering-based fingerprinting localization systems," *Wireless Networks*, pp. 1–14, 2018.

[67] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piché, "A comparative survey of wlan location fingerprinting methods," in *2009 6th workshop on positioning, navigation and communication*. IEEE, 2009, pp. 243–251.

[68] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *Ieee Infocom 2004*, vol. 2. IEEE, 2004, pp. 1012–1022.

[69] J. Luo and L. Fu, "A smartphone indoor localization algorithm based on wlan location fingerprinting with feature extraction and clustering," *Sensors*, vol. 17, no. 6, p. 1339, 2017.

[70] "Oracle Database Software Downloads — Oracle Technology Network — Oracle," https://www.oracle.com/technetwork/database/enterprise-edition/downloads/index.html, accessed: 2019-04-04.

[71] "NSDate- Foundation—Apple Developer Documentation," https://developer.apple.com/documentation/foundation/nsdate?language=objc, accessed: 2019-04-04.

[72] P. Danziger, "Big o notation," *Source internet: http://www. scs. ryerson. ca/~ mth110/Handouts/PD/bigO. pdf, Retrieve: April*, 2010.

[73] S. Wang, P. Yang, and H. Sun, "Fingerprinting acoustic localization indoor based on cluster analysis and iterative interpolation," *Applied Sciences*, vol. 8, no. 10, p. 1862, 2018.

[74] B. Wang, S. Zhou, W. Liu, and Y. Mo, "Indoor localization based on curve fitting and location search using received signal strength," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 572–582, 2015.

[75] 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, pp. 205–218.

[76] S. Yiu and K. Yang, "Gaussian process assisted fingerprinting localization," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 683–690, 2016.

[77] P. Mirowski, T. K. Ho, S. Yi, and M. MacDonald, "Signalslam: Simultaneous localization and mapping with mixed wifi, bluetooth, lte and magnetic signals," in *International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 2013, pp. 1–10.

[78] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit *et al.*, "Fastslam: A factored solution to the simultaneous localization and mapping problem," *Aaai/iaai*, vol. 593598, 2002.

[79] L.-W. Yeh, M.-H. Hsu, H.-Y. Huang, and Y.-C. Tseng, "Design and implementation of a self-guided indoor robot based on a two-tier localization architecture," *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 271–281, 2012.

[80] 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, pp. 293–304.

[81] R. Liu, C. Yuen, T.-N. Do, and U.-X. Tan, "Fusing similarity-based sequence and dead reckoning for indoor positioning without training," *IEEE Sensors Journal*, vol. 17, no. 13, pp. 4197–4207, 2017.

[82] K. Liu, Z. Meng, and C.-M. Own, "Gaussian process regression plus method for localization reliability improvement," *Sensors*, vol. 16, no. 8, p. 1193, 2016.

[83] S. Kumar, R. M. Hegde, and N. Trigoni, "Gaussian process regression for fingerprinting based localization," *Ad Hoc Networks*, vol. 51, pp. 1–10, 2016.

[84] A. A. Y. A. Rehim and M. Amin, "Horus: A wlan-based indoor location determination system," Ph.D. dissertation, 2004.

[85] P. A. Karegar, "Wireless fingerprinting indoor positioning using affinity propagation clustering methods," *Wireless Networks*, vol. 24, no. 8, pp. 2825–2833, 2018.

[86] Z. Tian, X. Tang, M. Zhou, and Z. Tan, "Fingerprint indoor positioning algorithm based on affinity propagation clustering," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 272, 2013.

[87] C. RasmussenandC, "Williams, gaussian processes for machine learning," 2006.

[88] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.

[89] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[90] L. Martino, V. Laparra, and G. Camps-Valls, "Probabilistic cross-validation estimators for gaussian process regression," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 823–827.

[91] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[92] C. Zhu, R. Byrd, P. Lu, and J. Nocedal, "L_bfgs_b fortran subroutines for large scale bound constrained optimization, l bfgs b_ fortran sub routines for large scale bound constrained optimization," in *Tech. Report Department of Electrical Engineering and Computer Science*. Northwestern University, 1995.

[93] A. Yu-Hong Dai, "perfect example for the bfgs method," *Mathematical Programming, Springer-Verlag*, vol. 138, no. 1-2, pp. 501–530, 2013.

[94] H. Zang, F. Baccelli, and J. Bolot, "Bayesian inference for localization in cellular networks," in *2010 Proceedings IEEE INFOCOM*. IEEE, 2010, pp. 1–9.

Acknowledgement

Foremost, I would like to express my sincere gratitude to my advisor, Prof. Jae-Young Pyun for his motivation, invaluable support, supervision, personal guidance, and useful suggestions throughout the course of my research. I have been extremely lucky to have an advisor who cared about my work, and who responded to my questions and queries so promptly. I would also like to thank all the Professors of Information and Communication Engineering Department from whom I have learned a great deal of knowledge.

Besides my advisor, I would like to thank the thesis committee members: Prof. Goo-Rak-Kwon, Prof. Bumshik Lee, Prof. Euisung Kang, Sunchon National University, and Prof. Park Sang Hyun, Sunchon National University, for their insightful comments and suggestions. Likewise, I would like to thank all anonymous reviewers at different leading international conferences and reputed journals for their time and efforts they spend in giving valuable feedbacks for the manuscripts that I have submitted for possible publications. Their suggestions have certainly raised the quality of those manuscripts, which are the foundation of this dissertation.

I thank my fellow lab mates of Wireless and Mobile Communication System Lab for their support and for all the fun we have had in the last five years. In addition, I thank my friends at Chosun University.

My sincere gratitude is reserved for Sarala, my wife, for her continued support and encouragement. I am as ever, especially indebted to my parents and family for their love and support in every moment of my life.