# A Study on Wi-Fi-Based Indoor Positioning System Using Fingerprint Database Constructed with Estimated Reference Locations

**Submitted to**
**Graduate School of Information Science and Technology,**
**Hokkaido University**

**2020**

**Myat Hsu Aung**

# Abstract

Recently, positioning systems have been widely used in consumer devices. The current position information of people and objects may help to navigate themselves to their destination in outdoor environment and even in indoor environment. Positioning systems play an essential role not only in navigation and tacking but also in many location-based applications. Global positioning system (GPS) works well in outdoor environments. However, the indoor positioning systems still have the great challenge for such a platform because of signal blockage and attenuation and hence causing it large error in position and reduces the sensitivity below the threshold of GPS receivers.

Many researchers have developed indoor positioning systems using various techniques. A major difficulty in indoor localization systems is the trade-off between accuracy and their costs in terms of infrastructure deployment and calibration process. A system has higher performance also has higher cost. Indoor positioning systems need higher accuracy than outdoor positioning systems due to the relatively narrower spaces available in indoor environments. Depend on the applications, it requires the various accuracy. We investigate a Wi-Fi-based indoor positioning system. In our research work, we have been implemented a Wi-Fi-based indoor positioning system using fingerprint method. The fingerprint database is constructed with estimated reference locations. We further propose the construction of fingerprint database using two merging methods to improve the database management. The proposed approach can control accuracy for any location-based applications.

Wi-Fi-based indoor positioning system is a positioning system that is used to locate objects or devices using the information from the Wi-Fi access points. There are several approaches to estimate positions such as trilateration or multilateration, triangulation, and fingerprinting. In this thesis, implementations of Wi-Fi-based indoor positioning system using fingerprint method are investigated. The fingerprint method does not require installing additional hardware such as dedicated Wi-Fi de-

vices. In addition, the actual access point locations are not required. However, the site survey process is required to know the actual reference locations stored in the fingerprint database. This process requires quite a large cost since many data should be collected with their actual positions using precise indoor maps or floor plans. Motivated by this, we propose a Wi-Fi-based indoor positioning system using estimated reference locations. We are trying to develop a method to create reference location databases without precise reference locations. In our proposed approach, the database is constructed by gathering pairs of media access control (MAC) addresses and received signal strength indicator (RSSI) values using a reference device moving at a constant speed with a simple direction. Assuming a constant speed, the location of each reference location can be estimated from the velocity. To estimate the user's position, we propose a position estimation algorithm using Euclidean-like distance that finds the best match of the current RSSI values in the fingerprint database. Estimation accuracy evaluation results show that our proposed approach can estimate the user's position without any precise reference locations.

In the fingerprint method, the database construction is important for position estimation performance. We proposed a Wi-Fi-based indoor positioning system using a fingerprint method, whose database is constructed with estimated reference locations. The reference locations and their information, called data sets, are obtained by moving reference devices at a constant speed while gathering information of available access points. Each data set includes some errors due to such as the fluctuation of RSSI values, the device-specific Wi-Fi sensitivities, the access point installations, and removals. To tackle this problem, we propose two methods to merge data sets to construct a consistent database suppressing such undesired effects. One method is using the mean-shift clustering algorithm to get estimated reference locations from multiple data sets. The number of estimated reference locations in the database depends on the mean-shift parameter to calculate means in the mean-shift algorithm. The experimental results show that the average error can be reduced compared with the database construction using one data set as the database. The other method assumes that the intervals of reference locations in the database are constant and that the fingerprint for each reference location is calculated from multiple data sets. This constant interval is called a grid size, and the number of estimated reference locations depends on the grid size. When merging data sets using a grid size, RSSI values in a specific area are averaged so that it can be expected to reduce noises due to the RSSI value fluctuation. The performance can be controlled by changing the grid size depending on the applications. Through

experimental results, we reveal that our approach can achieve an accuracy of 80 %. In both approaches, a large error occurs in cases having a small number of reference locations. In the grid-based approach, the estimated position error reduces in cases having a large number of reference locations. The obtained results show that the grid-based approach is effective and performs comparatively better. We have created the Wi-Fi-based indoor positioning system that is the cost effective and flexible for any location-based application.

# Acknowledgments

I would like to express my gratitude to my supervisor, associate professor Hiroshi Tsutsui, for the patient guidance, much encouragement, helpful advice, and essential support for the success of my work provided throughout my research. I have learnt many things about research working from him, and how to manage risks and particularly how to explore new ideas. His guidance helped me in all the time of research.

I also thank professor Yoshikazu Miyanaga for his support and guidance though during my study program.

My sincerest gratitude also goes out professor Takeo Ohgane, professor Kunimasa Saitoh, and associate professor Matteo Convertino for taking time to examine the contents of this thesis and providing valuable suggestions and comments that improved my research and dissertation as the members of my dissertation committee.

I would like to thank the Scholarships by the ASEAN University Network/ Southeast Asia Engineering Education Development Network (AUN/SEED-Net) Project, JICA for giving me the opportunity to attend the doctoral degree course at Hokkaido University as a JICA scholarship student and providing financial assistance to conduct this research for three years.

I also thank the GI-CoRE GSB, Hokkaido University for fruitful discussions. This work is supported in parts by the Ministry of Internal Affairs and Communications for SCOPE Program (175003017).

During the four and a half years in which I have been doing my research, I am thankful to all the members of the Information and Communication Networks Laboratory for their help and accompany.

Finally yet importantly, I wish to thank my family for all the love and support, during the year of my PhD study. Without their support and encouragement, my dissertation would not have been possible.

# Publications

## Transactions

1. Ye Chan, Myat Hsu Aung, Pho Kaung, "Fuzzy Adaptive Fusion Filter for RSSI Siganl Processing in Indoor Navigation," University of Yangon Research Journal, Vol. 9, No. 2, pp. 407–412, 2019.

2. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "WiFi Fingerprint Based Indoor Positioning Systems Using Estimated Reference Locations," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E103-A, No. 12, Dec. 2020.

## Conference Paper with Referee

1. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "An accuracy evaluation of WiFi based indoor positioning system using estimated reference locations," Proceedings of International Workshop on SmartInfo-Media Systems in Asia (SISA), pp. 321–326, Sept. 2017.

2. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "Construction and Management of Fingerprint Database with Estimated Reference Locations for WiFi Indoor Positioning Systems," Proceedings of the 23rd Multi-conference on Systemics, Cybernetics and Informatics (WMSCI 2019), Vol. 2, pp. 7–10, Jul. 2019.

## Conference Paper without Referee

1. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "An Implementation of WiFi Based Indoor Positioning System Using Estimated Reference Locations," Proceeding of GI-CoRE GSQ, GSB & IGM Joint Symposium, Jul. 2017.

2. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "WiFi Indoor Positioning System Using Fingerprint Database Constructed by Mean-Shift Clustering with Estimated Reference Locations," Proceedings of the GSB Student Workshop, The 2nd GI-CoRE GSQ, GSB & IGM Joint Symposium, Aug. 2018.

3. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "An Accuracy Evaluation of Fingerprint Database Constructed by Mean-Shift Clustering for WiFi Indoor Positioning Systems," Proceedings of 2018 Winter International Symposium on Big-Data, Cybersecurity and IoT, Dec. 2018.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Positioning system is a system which provides the position information of objects or people. This system uses to navigate and track objects or people in outdoor environment and even in indoor environment. In ancient times, people used celestial elements such as sun, moon, and star to known the current location, and the way and direction for navigate themselves to their destination. They known their location based on natural landmarks, such as trees, mountains, and rivers. However, sometimes they lost their way when they cannot see sun, moon, and star. Later, people used tools like hourglass, map, compass, and among others for navigation. Drawing details landmarks, they kept and recorded the location information on the map. Using tools, the more accurate navigation can be achieved.

In the 18th century, the object can be detected and estimated the distance between two ships using radio wave. During the World War II, the radar technology was developed using radio wave. Radar provides the range of the object from the radar device by reflected radio signal from the object. That can be known the location of the object from the radar device.

In the 19th century, the first satellite was lunched. The global positioning system (GPS) was developed which used satellites. GPS was originally proposed for military service to navigate ship, aircraft, and automobile. The location of vehicles can be estimated using radio signal from at least three or more satellites by GPS receiver. The technology has been allowed for civilian use from the 1980s. The technology is rapidly developed and GPS receivers are installed and build in many devices like smartphones and mobile devices. The location of the objects or people is easy to known and more accurate for navigation and tracking systems.

Nowadays positioning systems have become indispensable not only for navi-

gation and tracking but also for various location-based applications such as health care system, security, manufacturing and location-based IoT system. For such applications, different level of accuracy is desired that depends on each application requirements. The global positioning system (GPS) is most popular and suitable technology for outdoor positioning system. However, the large error occurs in indoor environment because the GPS signal becomes weak when they penetrate the construction materials. Reliable and accuracy of indoor positioning system remains as one of the greatest challenges in the area of location-based application. Also, a flexible and low-cost indoor positioning system is highly demanded. There are multiple factors that have direct impact on the development of indoor positioning solutions such as environment, accuracy, application, and suitable technologies, among many others.

The indoor positioning systems (IPS) in public buildings such as shopping malls, stations, and airports is quite attractive not only for the general users but also the location-aware service providers. In case of buildings for specific users such as company buildings, universities, and schools, the IPS will help us to develop more efficient navigation systems and people tracking systems. For indoor positioning systems (IPS), RF signals from wireless devices such as Wi-Fi, Bluetooth, ultra-wide band (UWB), and radio frequency identification (RFID) devices can be utilized to estimate mobile device positions. The other type is non-RF based system that includes infrared, visible light communication, sound-based, vision-based, and inertial sensor. In Chapter 2, we review some existing indoor positioning systems and their positioning technologies. In Sect. 2.1, we briefly introduce radio frequency (RF) based positioning systems and non-RF based positioning systems. The advantages and disadvantages are presented for each system.

We focus on the most cost-effective method that is Wi-Fi-based IPS since the Wi-Fi coverage is getting higher due to the significantly increasing number of private or public Wi-Fi access points (APs) in metropolitan areas. Wi-Fi has been a default feature in most of the current smartphones, laptops and other portable devices. However, existing Wi-Fi network is originally developed for data communication rather than positioning purposes. The positioning techniques are required to use in positioning system using Wi-Fi network. The distance between the Wi-Fi access point and user's mobile device can be estimated using measurement signal such as received signal strength indicator (RSS), time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA) from the access point. In Sect. 2.2, we present the principle of such distance measurement techniques. The location

of user's mobile device can be estimated using the distance measurement between the access points and user's mobile device with some positioning methods such as trilateration, triangulation, and fingerprint which are described details in Sect. 2.3.

In this research, we focus on the fingerprinting approach since this approach does not require additional hardware installation such as dedicated beacon devices. The actual location of each access point does not required. The user positions are estimated by finding the best match of the current information of available APs in fingerprint database.

There are various research works in literature proposing Wi-Fi-based IPS with different approaches in fingerprint database creation and location estimation. In the traditional fingerprint method, the site survey process is required to know the actual reference locations in the coverage area. We need to store the information of APs reachable from a lot of known reference locations. Most of the research works are using the previously determined reference locations during their fingerprint database constructions. While there are many efforts to know the actual position of reference locations in a large area, this site survey process is time-consuming, labor-intensive, and vulnerable to environmental changes. Focusing on the difficulty of database creation, we are trying to develop a method to create the database of reference location information which does not require precise reference locations.

In Chapter 3, we proposed a Wi-Fi-based indoor positioning system using a fingerprint method, whose database is constructed with estimated reference locations. We created the fingerprint database by gathering pairs of media access control (MAC) addresses and received signal strength indicator (RSSI) at each estimate reference location using a reference device moving in a constant speed with simple direction. It can be estimated the location of each reference location. Pair of MAC address and RSSI values from all available APs from each estimated reference location is stored in fingerprint database. To estimate the user's location, we implemented the position estimation algorithm to find the best matching reference location which stored in the fingerprint database. In the position estimation algorithm, we used Euclidean-like distance to calculate the vector distance between user's location and reference locations. According to evaluation results, the user's location can be estimated without any precise reference locations. It reduce the administration cost to know the actual reference locations.

However, the accuracy is influenced by the number of reference locations stored in the database and the device-specific Wi-Fi sensitivity. When we use dense reference locations to achieve high resolution, the amount of data stored in database

increases. In this case, a high computational cost is required in the user's position estimation process. The fingerprint database is established using received signal strength indicator (RSSI) values from APs, with the mobile device scanning for RSSI values from all available APs. In a dense AP environment, not all APs may be observed in a single sampling time. In these instances, when the mobile device is unable to receive RSSI values from some APs, these values are classified as unavailable RSSI values. Furthermore, RSSI values fluctuate due to spatial and temporal variations of interference, hardware variations, environmental effects such as human presences, and resource collision caused by other devices utilizing the same service at the same time. This leads to an inaccurate estimation of the user's location by selecting an irrelevant reference location from the database. Another difficulty is that the database depends on the APs, which means new installations and replacements of APs have a significant impact on the estimation accuracy. Considering these issues, not only the database creation but also how to manage the database after the database creation is essential. It might be also possible that the user's input data is used for update the fingerprint data on the fly, utilizing the user's locality.

To mitigate the challenges caused by the fluctuation of RSSI values, multiple data sets are used to create a single database with higher fidelity. The multiple data sets are collected using different device and different time duration to considering the device sensitivity and APs installations and replacements. In Chapter 4, we created the fingerprint database with estimated reference locations from multiple data sets using two data merging methods such as the mean-shift clustering algorithm and data merging with constant interval. Using mean-shift clustering algorithm, the number of estimated reference locations depends on the thresholds parameter, a radius for the mean calculation to settle each reference location as the result of clustering. The large mean-shift parameter gives high computation time in each reference location calculation compare with small parameter. However, the large error occurs in the large mean-shift parameter. This proposed approach can be reduced the accuracy error due to the estimation accuracy evaluation results that compared with the database are constructed using one data set. Using this approach, the location of each reference locations in database is changed whenever updating database.

The second merging method involved the constant interval between the reference location using the grid size. The number of reference location in merged data set from multiple data sets depends on the grid size. When merging data sets, RSSI values are averaged in the window of each estimated reference location in the new data set to be stored in the fingerprint database. We showed the effectiveness of the

proposed approach through the experimental results, including the errors and accuracy for various grid sizes and window sizes. The large grid sizes give high accuracy but large error while the small grid sizes give a low error but low accuracy. This means that, in the proposed merging method, the performance can be controlled, changing the grid size depending on applications. Note that the window size $w$ can be properly set for each grid size $d$ regarding the experimental result. When $d$ is small, a normalized average error of 0.02 corresponding 2.4 m can be archived. Also, when $d$ is large, an accuracy of 80 % can be archived. Comparison of two merging method, updating and maintaining the fingerprint database require high computation cost using mean-shift clustering algorithm, although the mean-shift clustering algorithm is popular in data clustering. Finally, Chapter 5 summarizes and concludes this thesis.

# Chapter 2

# Related works

## 2.1 Positioning system

Positioning systems have become indispensable not only for navigation and tracking but also for various location-based applications. There are numbers of existing positioning systems which utilize a variety of sensing technologies and system architectures. These systems have varying characteristics, such as accuracy, scalability, cost, and technology. This chapter describes some positioning systems and their principles.

There are two main types of positioning technologies: radio frequency (RF) based system which used radio signal from transmitter device to know the distance from the receiver devices, and non-RF based system which utilized the different signals such as laser, light, acoustic signal, magnetic, and among others. RF based systems are such as Global positioning system (GPS) based, Wi-Fi-based, ultra-wide band based, radio frequency identification (RFID) based, and Bluetooth-based. The other type includes infrared-based, visible light based, sound-based, vision-based, and inertial sensor. In Sect. 2.1.1 and Sect. 2.1.2, we discuss their benefits and limitations in the positioning system, and review related existing positioning systems.

These technologies are used in position system that depend on the requirements of applications and environments. Among them, GPS system, Wi-Fi based positioning system, and inertial positioning system are used due to their specific characteristics. Wi-Fi coverage is getting higher due to the significantly increasing number of private or public Wi-Fi access points (APs) in metropolitan areas. GPS receiver

and some inertial sensor have been built in most of the current some vehicles, smartphones, laptops and other portable devices.

In Sect. 2.2, we introduce four typical distance measurements, namely time of arrival (TOA), time difference of arrival (TDOA), angle of arrival (AOA), and received signal strength (RSS), utilize to estimated distance from the transmitter to receiver. Using the measurement signal, the position of the user's device can be calculated using positioning algorithm. There are various positioning algorithms such as trilateration, triangulation, and fingerprinting that describe in Sect. 2.3. We present comparison of positioning algorithms with different measurement in each positioning system.

### 2.1.1 Radio frequency based positioning system

#### Global positioning system (GPS) based

The global positioning system (GPS) is a satellite-based radio positioning system with a nominal constellation of 24 satellites. It was originally intended for military navigation applications, later the system was made available for civilian use. GPS has been widely used for navigation and positioning, and various types of GPS receivers for different positioning accuracy have been available. In the general GPS, as shown in Fig. 2.1, GPS receiver needs to receive GPS signals from a least three satellites to employ a position calculation process to determine physical locations [1]. Using signals from four or more satellites, 3D position of user can be determined.

GPS is most popular and suitable technology for outdoor positioning system. Since GPS is a satellite based positioning system, it works well in outdoor environment but not well in indoors navigation. The large error occurs in indoor environment because the GPS signal becomes weak when they penetrate the construction materials. Indoor positioning systems need higher accuracy than outdoor localization systems due to the relatively narrower spaces available in indoor environments.

For indoor environment, it is required additional devices installation like GPS repeater [2–4]. In [3], positioning accuracy about 2–4 m is achieved using GPS repeater in indoor experiments. However, it is required to have infrastructure set-up which could amount to large initial deployment costs.

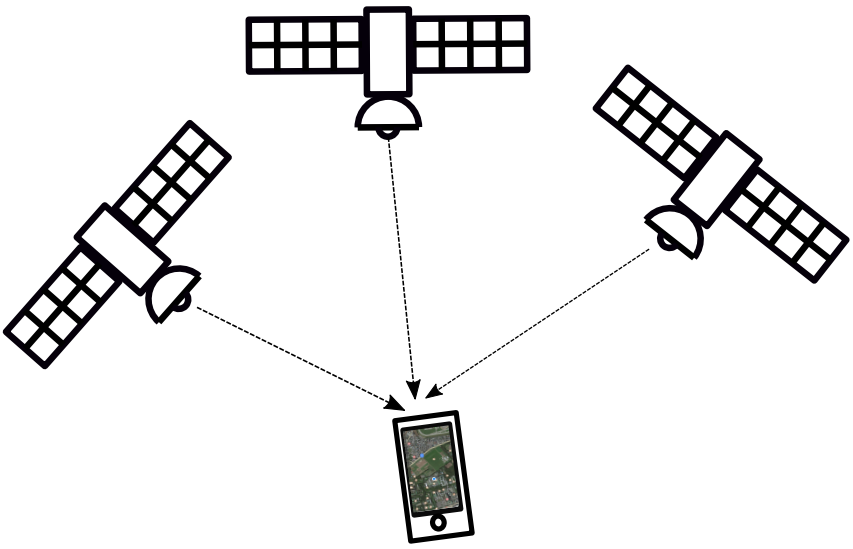


Figure 2.1: Global positioning system.

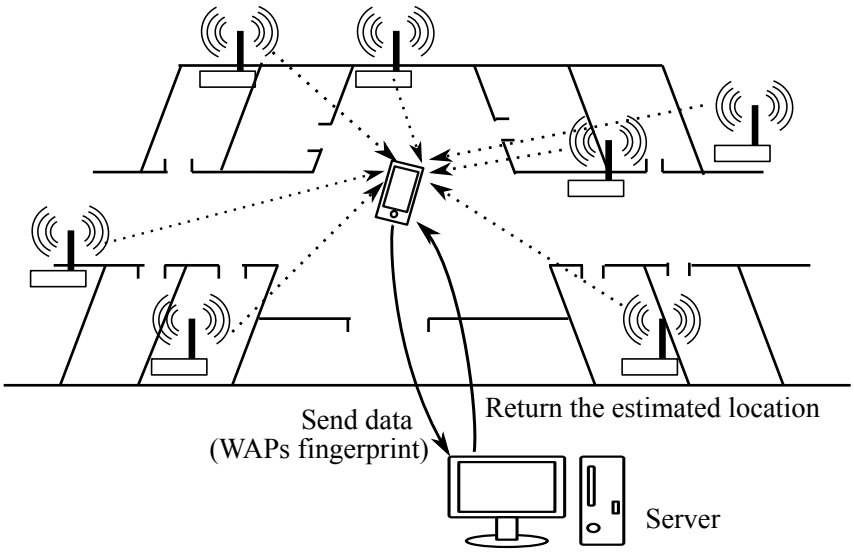



Figure 2.2: Wi-Fi-based positioning system.

### Wi-Fi-based

Wi-Fi is a technology that uses radio waves to connect to internet and communicate in local area network. Wi-Fi-based positioning system is used to determine the location of objects or devices using the information from the Wi-Fi access points. There are several positioning techniques that use Wi-Fi-based positioning system. Trilateration and triangulation techniques are used to know the position of user's device by the device's distance from access points. The distance between the Wi-Fi access points and user's device can be calculated according to TOA, TDOA, AOA, and RSS which present in Sect. 2.2. Another position techniques is fingerprinting that is the most cost-effective method since Wi-Fi coverage is getting wider due to the significantly increasing number of private and public Wi-Fi access points (APs) in the metropolitan area. The existing Wi-Fi access points can be used in positioning system without installation additional devices.

There are two different methods for implementing user's position computation process: self and remote positioning. In case of self positioning, the position is known by the user and the whole process operating on the user's device. In remote positioning, the position is determined at the server side using the information of access points which are collected from the user's device. The second method is typically used. In general Wi-Fi-based positioning systems, as shown in Fig. 2.2, a user sends data of access points reachable from the current position to the server, and receives the estimated user location information. There are various research works in literature proposing Wi-Fi-based IPS with different approaches in [5–13]. The accuracy of the Wi-Fi-based positioning systems varies from sub-meter to several meters for different algorithms and deployment densities.

### Radio frequency identification (RFID) based

RFID is an automatic identification method that relies on storing and remotely retrieving data using data-carrying devices known as RFID tags or transponders. The power required to operate device is transferred using a contactless technology from a data-capturing device and RFID reader. The basic communication between the reader and the transponder of an RFID system is based on radio frequency (RF) technology. The distance to be covered can be adjusted by using the frequency. An RFID reader generally acts as sender as well as receiver. It captures an encoded radio signal to interrogate the tag.

The RFID tags are primarily classified into active and passive tags. An active

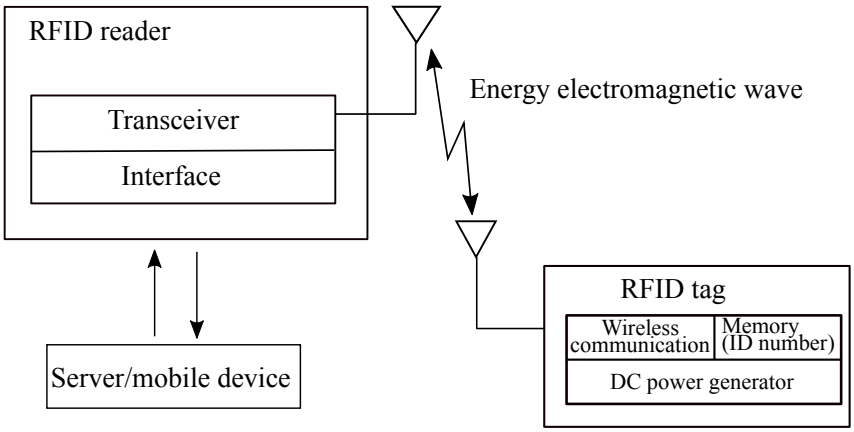



Figure 2.3: Simplified RFID system's architecture.

tag is a battery-powered transceiver and thus has a wider transmission range, hence reducing the number of tags required for an installation. A passive tag has a shorter transmission range because it does not use a battery; it gets its power from reader's signal before it can respond with information. The reader and tag, as well as a sever for position computation, are used in an RFID-based positioning system.

RFID positioning system can be classified as fixed-tag position and fixed reader or antenna position, in accordance with different roles of tags and readers/antennas. In the fixed-tag, the tags are deployed on fix position while the readers/antennas are usually attached to mobile objects. This is cost effective when the objects to be tracked are relatively large, few in numbers, and usually move in a 2D plane or on a certain route. In the fixed reader/antenna scheme, the readers/antennas and tags are placed in an opposite way to the fixed-tag. The readers/antennas are installed at fixed positions while the tags are attached to the items to be tracked. Many equipment are need for this positioning system. RFID positioning system is better than GPS positioning system for both indoor and outdoor environment. But the cost of the whole system is high [14, 15].

### Ultra-wide band based

Ultra-wide band (UWB) is a short-range high-speed radio technology for wireless communication with the ability to have a robust resistance to non-line of sight and multi-path effects. UWB can be used for positioning by utilizing the distance measurement of the signals to obtain the distance between the UWB transmitters and

receivers which is called active UWB-based positioning. We can classified into two UWB-based positioning system: active and passive. Passive UWB-based positioning system is a system that uses signal reflection, and not an attached UWB tag, to determine the position of a person or object by means of the principle of radar. When a person moves within a room with known positions of installed UWB transmitters and receivers, the body of the person reflects the signals emitted by the transmitters. The receivers sense the reflected signal, and the position of the person is estimated. The advantages of UWB-based indoor positioning system are long battery life for UWB tags, robust flexibility, high data rate, high penetrating power, low-power consumption and transmission, food positioning accuracy and performance, and multi-path effects. However, UWB is expensive to scale because of the need to deploy more sensors in a wide coverage area to improve performance. [16]

#### Bluetooth-based

Bluetooth is a short-range and low power technology that is used for peer to peer data communication. Bluetooth based positioning system is a system that locate and track objects and people inside a building by providing real-time position information of user's device using fixed Bluetooth sensors. Bluetooth devices that are within the range of the installed Bluetooth sensors are able to connect the user's device, thus receiving information from the Bluetooth sensors. The position of the device are computed using different positioning techniques such as ToA, AoA, and RSSI [17, 18]. Bluetooth based system mostly uses RSSI due to less complex. Bluetooth based positioning system is low cost and energy-efficient system. However, the many Bluetooth sensors are needed to install on a large scale due to short-range.

### 2.1.2 Non-radio frequency based positioning system

#### Infrared based

The system uses infrared signals to determine the position of object or people. Each location in a building covered with a network of IR sensors which detect these transmissions. IR can be used in two different ways; direct IR and diffuse IR. The direct IR uses a point-to-point ad-hoc data transmission standard designed for very low power communication. The diffuse IR has stronger signals than direct IR. Direct IR requires line of sight between devices. However, diffuse IR does not

require direct line of sight due to wide angle LEDs which emit signals in many directions. Angle of arrival (AoA) is mostly used in IR-based positioning system.

### Sound-based

There are two types of sound-based positioning system: acoustic signal and ultrasound. The acoustic signal-based positioning system uses sound sources (audio speaker) from acoustic signals to estimate the user position by capturing the microphone sensors. The traditional method used for acoustic-based localization has been the transmission of modulated acoustic signals, containing time stamps or other time related information, which are used by the microphone sensors for TOA estimation.

The ultrasound based system uses ultrasonic sensor. The ultrasound based positioning system also mainly relies on the TOA measurements of ultrasound signal and the sound velocity to calculate the distance between a transmitter and a receiver. Ultrasonic positioning system has a more limited range because of the loss of energy with the distance traveled. Also they are sensitive to ultrasonic ambient noise and have a low update rate.

The sound-based systems achieve high localization accuracy at room level. However, they need high cost implementation. The sound-based systems need of extra infrastructure such as sound sources or ultrasonic sensors, those are expensive to deploy and maintain. The system implementation on a large scale is the disadvantages in this system. It make the sound-based not a very popular technology for positioning system in large sale area. However, nowadays the acoustic signal based becomes popular in indoor positioning system using existing sound sources in the building [19]. But the main disadvantage is the line of sight.

### Vision-based

The vision-based positioning system is a system that determines the position of a person or an object by identifying a image that in with view, with camera and computer version based technology. The vision-based system uses camera's motion relative to a rigid scene to estimate the current position of the camera and to locate moving objects or people in the images [20–22]. This system is very effective in actual environment and low cost positioning system due to using low cost CMOS sensors. However, incorrect recognizing the object of interest in the input image leads to wroth positioning system.

### Visible light communication based

Visible light communication (VLC) is an emerging technology for high-speed data transfer that used visible light, modulated and emitted primarily by light-emitting diodes (LEDs). Visible light communication (VLC) based positioning system uses light sensors to measure the position and direction of the LED emitters for tracking and navigation purposes. The LEDs transmit the signal, which when collected by receiver can be used for localization. The visible light communication is a short-range free-space optical wireless communication technology that uses visible light to transmit data at 380–780 nm of wavelength. In VLC based system, it consists of a light source, image sensor, and a line of sight communication channel. Light sources are mounted on the ceiling of a room to transmit signals from their known position. The signals are received and demodulated by image sensor from an unknown position. The position of image sensor is then calculated using this information. For visible light, AoA is considered the most accurate localization technique. The advantage of visible light based localization is its wide scale proliferation. However, a fundamental limitation is that line of sight between the LED and the sensors is required for accurate localization.

VL positioning system is complex in design, especially when implemented over a wide coverage area. However, the use of LEDs for landscape architecture or illumination has attracted attention, and the LED industry is rapidly growing as LEDs have several advantages such as long life expectancy, high tolerance to humidity, low power consumption, and environmental friendliness. Therefore, we can expect that the existing fluorescent lamps will be replaced with LED lights in the indoor environment. Using existing LED lights, the indoor positioning systems based on LEDs have been proposed [23].

### IMU sensor based

Inertial measurement units (IMU) sensors based positioning system is system that measures the position, orientation, and velocity of the user's device using accelerometer (motion sensor) and gyroscopes (rotation sensors) for tracking and navigation, also known as an inertial navigation system (INS). In general, an inertial navigation system (INS) is used with the dead reckoning method using IMU sensors. A starting position of device is required. Since the sensors can only detect changes from one state to another, the starting position has to be set and known for tracking and navigation. Output of accelerometer and gyroscope measurements are 3D vectors

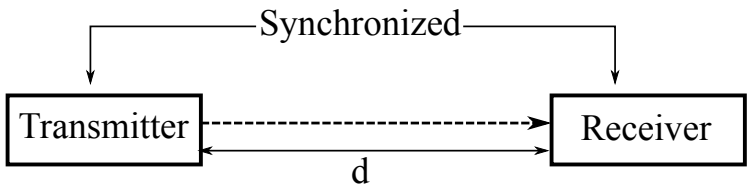



Figure 2.4: Principle of time of arrival (ToA).

of accelerations and angular velocities that are used to derive a positioning solution. This positioning solution includes the position (normally latitude, longitude), the direction (east, west, north, south), and the orientation (pitch, yaw, roll) [24]. To estimate device's position and orientation, double integration of acceleration measurement and single integration of angular velocity are required. Use of integration is the main cause for position error of INSs because of accumulation of constant errors (sensor drift, bad sensor calibration, etc.). Due to the integration a constant error results in error in position of the device. For positioning system, INS combines with other positioning techniques have been developed [25–29].

## 2.2 Distance estimation methods

### 2.2.1 Time of arrival (ToA)

Time of arrival (ToA) is the ranging based distance measurement method that is the time taken for electromagnetic wave to travel the distance. The transmitter transmits a signal at a time that is known to the receiver. The receiver's clock is somehow synchronized to the transmitter's clock as shown in Fig. 2.4. The distance between the transmitter and receiver can be calculated using

$$d = (t_a - t_s) \times c, \tag{2.1}$$

where $t_a$ is time of arrival and $t_s$ time take-of from the transmitter. $c$ is the speed of light. TOA is sometimes called time of flight (TOF). Knowing the distance from transmitters, the location of the receiver can be calculated.

TOA requires strict synchronization between transmitters and receivers and timestamps to be transmitted with the signal. The delay time can be occurred the error. The clock errors are usually less than 1 ms and varies slowly over time. In electromagnetic, time synchronization error is 0.3 m. In multipath indoor environments,

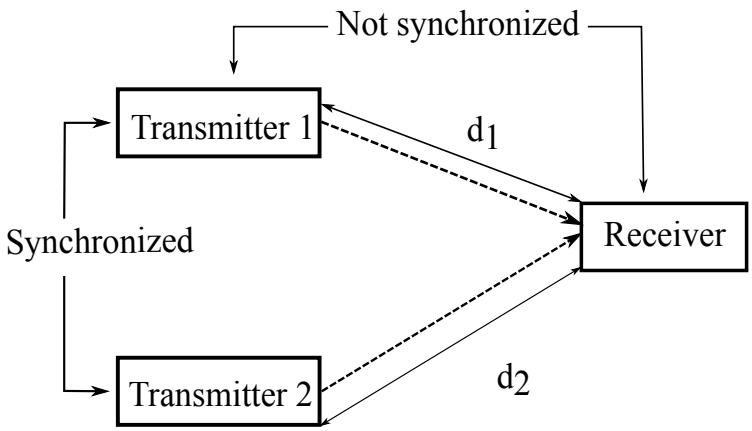



Figure 2.5: Principle of time difference of arrival (TDoA).

the obstacles deflect the emitted signals when the direct line of sight path between the transmitter and receiver is not available. Due to this effect, the location errors cannot be eliminated. The short time signal is better the performance, however, the noise can be occurred due to small signal power [30–33].

The satellite has very accurate rubidium clocks. These clocks are allowed to run freely in each satellite as it is difficult to synchronize 20 odd satellites. A quadratic polynomial expression is used to determine the correction to GPS Time. The parameters of correction are determined by the master control station and uploaded to the satellites. With the correction, the GPS time obtained from the SV time is accurate to about 3 ns. The error contribution from this source can be 1 m.

### 2.2.2 Time difference of arrival (TODA)

Time difference of arrival (TDOA) is also ranging-based distance measurement method. It also uses the time taken for signal to travel the distance between transmitter and receiver. However, TDOA is used to determine position of receiver based on the signal arrival time difference from multiple transmitters which are known the actual location. The principle of time difference of arrival is shown in Fig. 2.5. Using the signal from multiple transmitters, a difference in distances between two transmitters can be calculated by multiplying the speed of light $c$ and a difference in the arrival time from two transmitters as follows,

$$d_2 - d_1 = (t_{r_2} - t_{r_1}) \times c. \tag{2.2}$$

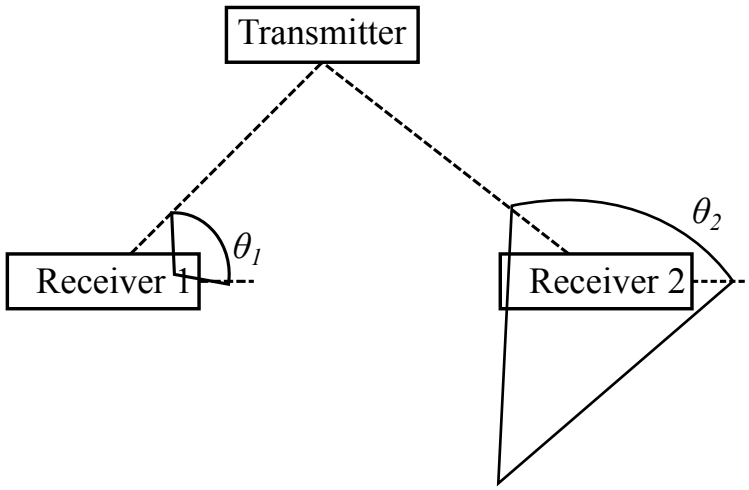



Figure 2.6: Principle of angle of arrival (AoA).

There are two different architectures that is active and passive system. In both architectures, there is a need for synchronization. The active system needs one transmitter and multiple receivers. The synchronization is required between receivers. In the passive system, there is one receiver and multiple transmitters. In case of GPS, there is the passive system. All the transmitters are synchronized closely. Some of the indoor locations [34–39], utilizes multiple receivers and single transmitter architecture.

### 2.2.3 Angle of arrival (AOA)

Angle of arrival (AOA) is the angle and distance calculated relative to two or multiple receivers through the intersection of direction lines between the receivers. The calculation of the angle and distance is used to estimate and determine the position of a transmitter, and the information is used for tracking or for navigation purposed [40, 41]. Figure 2.6 shows the principle of angle of arrival (AoA). AoA had been achieved high average accuracy when the environment is line of sight.

### 2.2.4 Received signal strength indicator (RSSI)

Received signal strength indicator (RSSI) is a measurement of the power present in a radio infrastructure, measured in decibel-milliwatts (dBm) or milliWatts (mW). As all electromagnetic waves have inverse-square relationship between received power

and distance as in following expression,

$$P_r \propto \frac{1}{d^2}, \tag{2.3}$$

where $P_r$ is the received power at a distance $d$ from transmitter.

Figure 2.7 shows the principle of RSSI values that related with distance between access point and mobile device. The RSSI values decrease as the signal propagates when the mobile device is apart from the access point. Using RSSI values and the signal propagation model along the line of sight, the distance can be calculated using following equation:

$$\text{RSSI} = -10n \log_{10} d + A. \tag{2.4}$$

Where $n$ is the propagation path loss exponent, $d$ is the distance between access point and mobile device, and $A$ is the RSSI value at one meter of distance from access point as reference RSSI value of access point. In other to determine the path loss exponent, the RSSI values within 10 m of the sink will be measured with a step size of 1 m and the root mean square (RMS) of the measured RSSI values will be calculated and then the best value for the path loss exponent $n$.

The path loss exponent is a significant factor which depends on the local environment and differs from channel to channel. Typical path loss exponents (PLE) are different for various environments. In the buildings, the PLE's range is 1.6 to 3.5 when the receiver and transmitter at the same floor, for multiple floors the PLE are between 2 and 6. However, attenuation due to obstacles in buildings expand the range of empirical PLE.

RSSI method is among the cheapest and easiest method to implement. However, it does not provide the best accuracy because of RSSI fluctuation. RSSI values fluctuate with various effect as follows:

1) Hardware: Orientation, direction, and type of antenna, and the device specific Wi-Fi sensitivity

2) Spatial: Distance between receiver (mobile device) and transmitter (access point)

3) Temporal: Time and period of measurement

4) Interference: resource collision caused by other devices utilizing the same service at the same time

5) Human: User's presence or absence, orientation, movement

6) Environment: Building types, materials

Using RSSI values in positioning system, filtering process is required to improve location accuracy. Various algorithms can be used to smooth the RSSI value and reduce error [42–44].

## 2.3 Location estimation methods

The location estimation methods specify how to calculate the position of objects and people. By knowing the distance to the transmitters using the distance estimation methods as mentions in Sect. 2.2, the location of objects and people can be estimated using location estimation methods such as trilateration, triangulation and fingerprint as shown in Fig. 2.8.

### 2.3.1 Trilateration

Trilateration is location estimation method which uses distance measurements relative to known position to determine the position of the object. At least three known position of transmitters are required to use trilateration method. Similarly, four or more known position is used to determine the position of unknown device which is called multilateration. Figures 2.9 and 2.10 show the overview of trilateration and multilateration algorithm, respectively. The location of the target unknown receiver is computed as the intersection point of the spheres center at each transmitter and with radius equal to the distance to the receiver. The distance from each transmitter to receiver is described as following equation:

$$(x_i - x_u)^2 + (y_i - y_u)^2 + (z_i - z_u)^2 = d_i^2 \tag{2.5}$$

where $(x_i, y_i, z_i)$ and $(x_u, y_u, z_u)$ are the known position of transmitter $i = 1, 2, \ldots, N$, and the target unknown position, respectively. $d_i$ is the distance between target and known position $i$. The distance between transmitter and receiver is calculated using TOA, TDOA, and RSSI. The position of receiver is estimated using trilateration algorithm with least square method.

Trilateration has been widely used in GPS positioning. GPS receivers use the trilateration method to determine the user's position, motion and direction. In indoor positioning using Wi-Fi, bluetooth and other sensing devices, the additional

devices installation is required. The location of those devices is needed to know for distance calculation.

### 2.3.2 Triangulation

Triangulation uses the geometric properties of triangles to estimate the position of a target object by computing angular measurements relative to two known reference points. In other hand, the position of the target object is found by the intersection of two pairs of angle direction lines, a method known as direction finding. Angle of arrival is used to compute the distance between direction lines or fixed points to locate the object. The position of the object is determined by calculating the position of transmitter based on the angle and distance relative to the reference points. Furthermore, when two or three reference points are used to determine position, it results in a simple and low-cost system.

### 2.3.3 Fingerprinting

Fingerprint method has been used for indoor positioning system to estimate the position of objects by finding the best match of the current information of unknown position in an fingerprint database. The fingerprint method consists of two phases: offline phase and online phase. In the offline phase, information from sensor or radio wave transmitter are collected at each specific location. Such locations are called reference locations. The collected data for each reference location is stored in a fingerprint database. This phase can be regarded as a database creation phase or training phase. In the online phase, the real-time position of the user is estimated using input set of current data that is collected by user's mobile device. A feature vector called fingerprint is generated from the current data and it is compared with those stored in the fingerprint database. The position is estimated based on the best matching vector in the database.

In fingerprint method, the site survey process is required to known the actual location of reference locations in the coverage ares. There are many efforts to know the actual position of reference locations in a large area, this site survey process is time-consuming, labor intensive, and vulnerable to environmental changes.

Compared with other techniques like trilatration and triangulation, this technique is more suitable for positioning system in indoor environment which is use existing infrastructure like Wi-Fi. Much research interest in Wi-Fi based indoor

positioning system using fingerprint method has developed focusing on various aspects [43, 45–49].

## 2.4 Conclusion

The advantages and disadvantages of aforementioned various positioning system used in indoor environment are compared in Table 2.1. There are needed to consider not only accuracy and performance but also installation costs, calibration time, power consumption, maintenance and life time [50–53]. The cost of IPS contains several parts: the cost of the infrastructure components, system installation and maintenance. In case of GPS, it does not suitable for indoor environment due to the additional devices installation that leads to high cost and complex implementation. The accuracy of RFID and IR based positioning systems are higher than other positioning systems. However, the cost of the whole system is increased because of these extra infrastructure to be installed.

Usually, there is a trade-off between the cost and the performance of indoor positioning system. A system has higher performance also has higher cost. Among various positioning systems, Wi-Fi-based IPS is more cost effective because there is no extra cost incurred due to reuse the existing infrastructures.

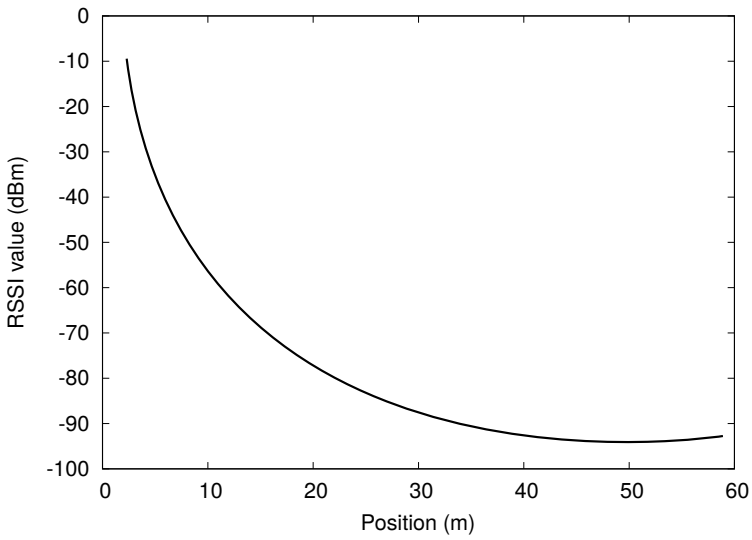



Figure 2.7: Principle of Received signal strength indicator.

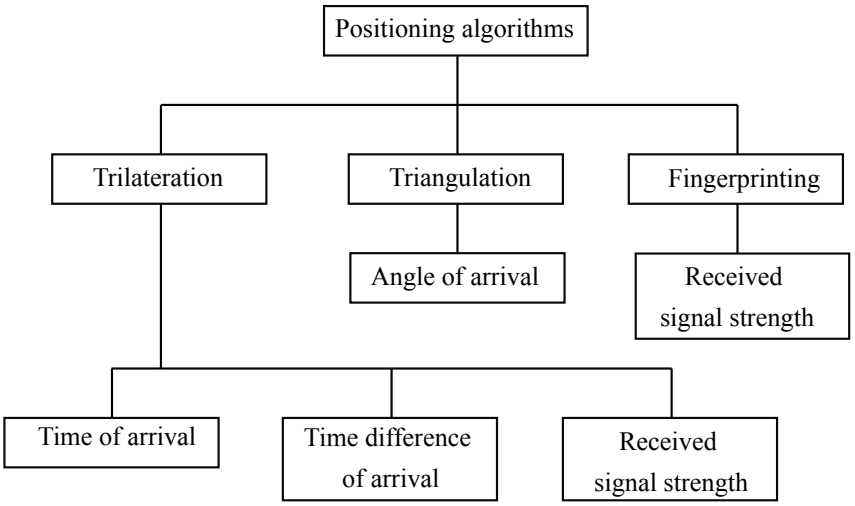



Figure 2.8: Positioning algorithms.

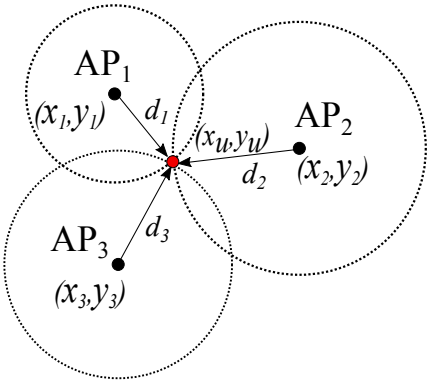



Figure 2.9: The overview of trilateration algorithm.

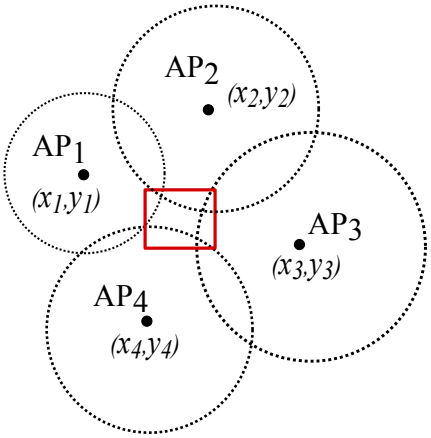



Figure 2.10: The overview of multilateration algorithm.

Table 2.1: Comparison of indoor positioning technologies.

| Technique | Advantages | Disadvantages |
|---|---|---|
| Wi-Fi | use existing infrastructure and no need LOS | inaccurate due to noise in case of changes in the environment |
| RFID | High accuracy, no need LOS between reader and tags | High installation cost |
| UWB | High accuracy and multipath | High cost of UWB equipment |
| Bluetooth | low energy consumption and low cost device | need many extra Bluetooth device to better accuracy |
| Infrared | suitable for small space | require LOS and need devices installation |
| Sound (ultrasonic) | comparatively less absorption | sensitive to environmental change, short range, and need devices installation |
| Sound (acoustic) | use existing sound source | sensitive to environmental change and short range |
| Vision | inexpensive due to low cost COMS sensor, effective in real-time | require LOS and need large amount of sensor on large scale area that leads high cost implementation |
| Visible light | use existing infrastructure, multhipath-free | high power consumption, and range is effected by environment change |
| IMU | no need additional hardware | require LOS and accumulate sensor drift error |

# Chapter 3

# Proposed Wi-Fi-based indoor positioning system using estimated reference locations

## 3.1 Introduction

Recently, positioning systems have been widely used in consumer devices. Users' current position information may help users to navigate themselves to their destination in outdoor environment and even in indoor environment. There are two main types of positioning technologies: radio frequency (RF) based system and non-RF based system. RF based systems are such as global positioning systems (GPS), Wi-Fi based positioning systems, radio frequency identification (RFID) based systems, ultra-wide band based systems, and Bluetooth based systems. The other type includes infrared systems, visible light communication systems, vision based systems, sound based systems and IMU sensor based system. In Chapter 2, these positioning systems are literately reviewed.

Among these technologies, the most popular positioning system is GPS. Since GPS is a satellite based positioning system, it works well in outdoor environment but not well in indoors navigation. The GPS signal is blocked by most construction materials and hence large error occurs in indoor positioning [8]. In other words, there are accuracy problems for absolute positioning in a specific indoor area [8, 11]. It should be also noted that indoor localization systems need higher accuracy than outdoor localization systems due to the relatively narrower spaces available in

indoor environments [7].

For indoor positioning systems (IPS), RF signals from wireless local area network (WLAN), Bluetooth, and cellular networks can be utilized to estimate the users' positions [45]. Among them, we focus on Wi-Fi based IPS since the Wi-Fi coverage is getting higher due to greatly increasing number of private or public Wi-Fi access points in metropolitan areas. The IPS in public buildings such as shopping malls, stations, and airports is quite attractive not only for the general users but also the location-aware service providers. In case of buildings for specific users such as company buildings, universities, and schools, the IPS will help us to develop more efficient navigation systems and people tracking systems.

There are a lot of issues to be tackled in IPS. Among them, we focus on the difficulty of database creation. The information of access points should be collected in advance as a form of database. In case of fingerprinting based IPS, the information of access points reachable from user's location is used for estimating the user's location by comparing it with the pre-stored data in the database. Such user's information is called fingerprint. As for database, we need store the information of access points reachable from a lot of reference points whose locations are known. This database creation requires quite large cost since many data should be collected with their actual positions using precise indoor map or floor plan.

Motivated by this, we propose a Wi-Fi based IPS using estimated reference locations in this paper. We are trying to develop a method to create database of reference point information which does not require precise reference point locations. Using the proposed system, it is expected that the cost of database creation can be dramatically reduced. In the proposed approach, the database is constructed by gathering pairs of media access control (MAC) addresses and received signal strength indicator (RSSI) values using reference devices moving at a constant speed in simple direction. Assuming a constant speed, the location of each reference point can be estimated from the velocity. In this chapter, we present an accuracy evaluation of the proposed system. Evaluation results show that user's locations can be roughly estimated.

## 3.2 Proposed approach

Wi-Fi fingerprint positioning systems estimate user locations based on fingerprint methods utilizing the RSSI values from reachable APs from the user in the coverage

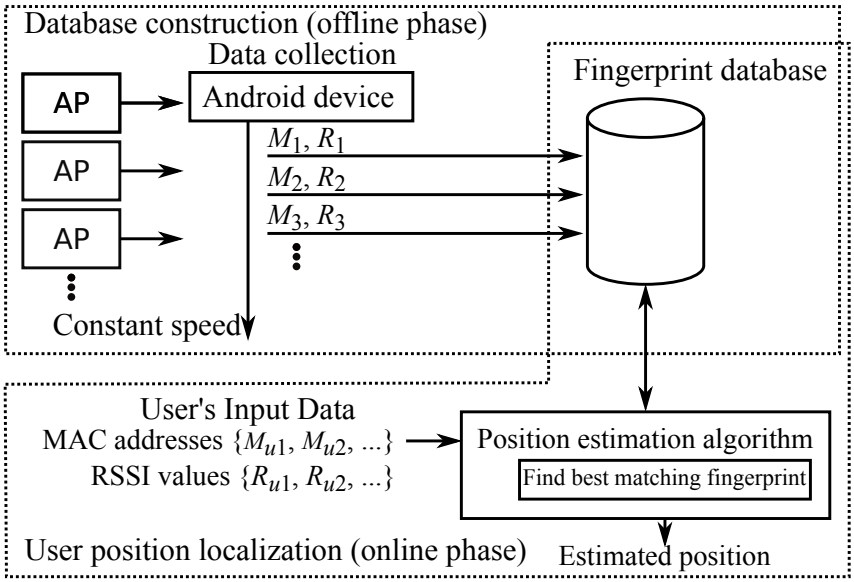



Figure 3.1: The overview of the proposed system.

area. Figure 3.1 shows the overview of the proposed system.

### 3.2.1 Database construction

In the database construction part, pairs of MAC addresses and RSSI values which can be obtained from reachable access points (APs) from each reference point are collected in the target area by using such as Wi-Fi scanning Android OS applications. These MAC-RSSI pairs are gathered every specific period such as one second with walking at a constant speed to estimate the actual location of each reference point. The absolute timestamps for each pair is also collected. The locations where MAC-RSSI pairs are sampled can be regarded as reference points. The actual locations can be estimated from the timestamps assuming that the samples are gathered with walking at a constant speed. As a result, a set of pairs of MAC addresses and RSSI values for each reference point is stored as a fingerprint in the database. Let the set of MAC addresses $M_i = \{M_{i1}, M_{i2}, \ldots, M_{iN_i}\}$ for reference point $i$, where the number of access points available at reference point $i$ is given by $N_i$. The RSSI values paired with $M_i$ is denoted by $R_i = \{R_{i1}, R_{i2}, \ldots, R_{iN_i}\}$. We denote the set of MAC-RSSI pairs for reference point $i$ by $P_i = \big(M_i, R_i\big)$. The list of $P_i$ is stored in the database as $P$.

### 3.2.2 Position localization

In the user position localization, the user's mobile device samples the MAC address and the RSSI value of each AP available from the current user's position. This sample can be denoted using the notations similar to $P_i = \big(M_i, R_i\big) \in P$ as follows,

- the set of MAC addresses at the user's position $u$:
  $M_u = \{M_{u1}, M_{u2}, \ldots, M_{uN_u}\}$,
  where $N_u$ is the number of available APs,
- the RSSI values paired with $M_u$:
  $R_u = \{R_{u1}, R_{u2}, \ldots, R_{uN_u}\}$, and,
- the set of MAC-RSSI pairs at the user's position $u$:
  $P_u = \big(M_u, R_u\big)$,

where $P_u$ is the input in the position estimation algorithm, and the algorithm estimates the user's location by using $P_u$ and $P$. Note that $P$ is the database.

The procedure of the user's position estimation is described as follows:

(1) Pick up reference locations, which include at least one MAC address of user's input and create the set of such reference points. This set is described by

$$S = \big\{i \big| M_u \cap M_i \neq \varnothing\big\}\,. \tag{3.1}$$

(2) Create the list of unique MAC addresses which is related to the input data.

$$M' = \bigcup_{i \in S} M_i \tag{3.2}$$

(3) Create a location vector, that is, a fingerprint for the user's input and reference locations selected in step (1). The location vector is given as the RSSI values list for the corresponding MAC address list created in step (2). If no RSSI value is available for a MAC address, the element is set $\varnothing$. That is,

$$V_u = \{V_{uj}\},\ V_{uj} = \begin{cases} R_{uj} & \text{if } M'_j \in M_u \\ \varnothing & \text{otherwise,} \end{cases} \tag{3.3}$$

where $M'_j \in M'$. Similarly, as for selected references points, the location vectors are created by

$$V_i = \{V_{ij}\},\ V_{ij} = \begin{cases} R_{ij} & \text{if } M'_j \in M_i \\ \varnothing & \text{otherwise.} \end{cases} \tag{3.4}$$

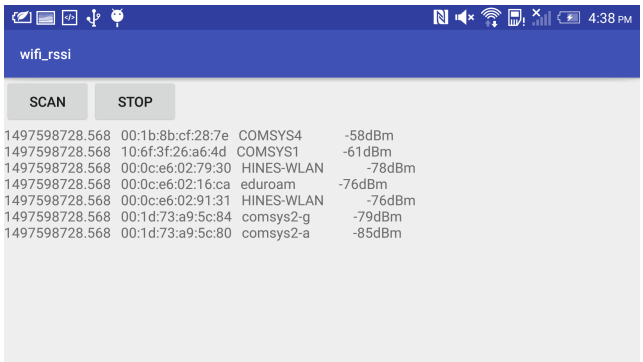



Figure 3.2: Screen shot of the Android application.

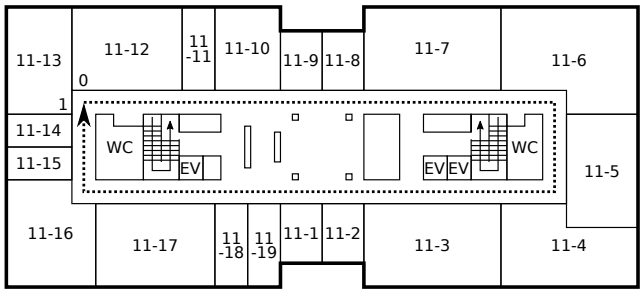



Figure 3.3: Floor map and the forward direction in the experiment. Normalized position 0 corresponds to the starting point while 1 to the end point. The total distance is approximately 120 m.

(4) Calculate vector distances $D_{ui}$ between user's input vector $V_u$ and all reference vectors $V_i$ using Euclidean-like distance as follows,

$$D_{ui} = \frac{1}{N_j^A} \sum_{\{j|M'_j \in M', V_{uj} \neq \varnothing, V_{ij} \neq \varnothing,\}} |V_{uj} - V_{ij}|^2 , \tag{3.5}$$

$$N_j = \sum_{\{j|M'_j \in M', V_{uj} \neq \varnothing, V_{ij} \neq \varnothing,\}} 1, \tag{3.6}$$

where $N_j$ is the number of accumulations and $A \geq 1$. This parameter $A$ controls the contribution of $N_j$, where larger $N_j$ gives a smaller distance.

(5) Output the position $\hat{u} = \text{argmin}_i \, D_{ui}$ as the estimated user's position.

Table 3.1: Reference Android devices.

| Device symbol | Model |
|---|---|
| A | HTC One (M7) |
| B | Nexus 7 (2013) |

Table 3.2: Different types of data set depends on the moving speed and direction .

| Data | Note |
|---|---|
| Slow 1 | Moving at a constant slow speed. |
| Slow 2 | Moving at almost the same speed as Slow 1. |
| Fast | Moving at a relatively fast speed. |
| Reverse | Moving in the reverse direction at almost the same speed as Slow 1 and 2. |

Table 3.3: Number of samples for each data set.

| Data set | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of reference locations | 228 | 202 | 167 | 196 | 242 | 206 | 171 | 200 | 203 | 188 | 152 | 164 | 206 | 191 | 154 | 167 |

## 3.3 Data collection

First, we developed an Android application which gathers MAC-RSSI pairs of available access points every specific period aiming database construction and location estimation algorithm development. A screenshot of this application is shown in Fig. 3.2. By using data gathered by this Android application, we can create a database and verify the developed algorithm.

Data was collected on 11th floor of Graduate School of Information and Science Technology Building, Hokkaido University, whose floor map is shown in Fig. 3.3. In this experiment, we used two Android devices. One is (A) HTC One (M7) and the other is (B) Nexus 7 (2013), as shown in Table 3.1. Data was collected simultaneously using these two devices. We need to collect at least two data sets assuming that one is for database creation (training) and the other is for position estimation (testing). We obtained four data sets for each device. These four types of data acquisition is summarized in Table 3.2. The data is gathered by moving along on the dashed line in Fig. 3.3 with two Android devices at a constant slow speed, a faster speed, and in the reverse direction at the slow speed. The total walking distance is

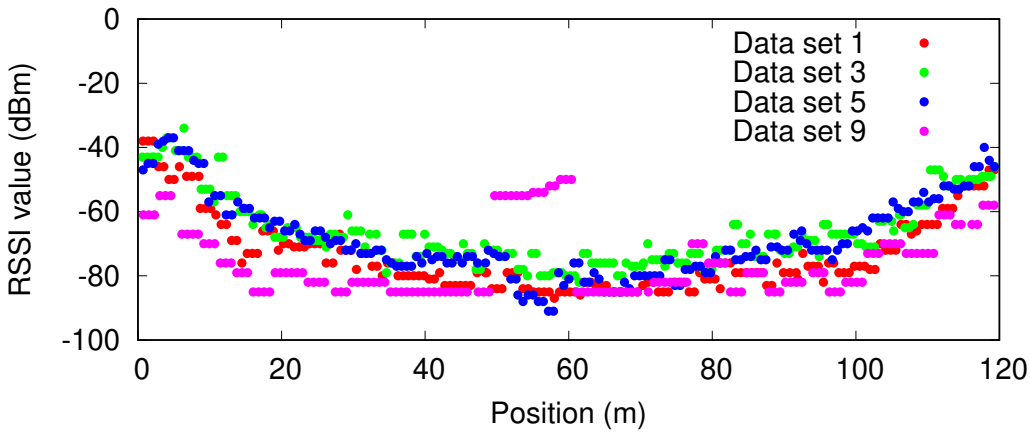



Figure 3.4: RSSI values of an AP (MAC address) in different data sets from different positions. The data set number is the same as in Table 4.1.

about 120 m. Note that the starting point and the end point is same.

As for device (A), 42 access points are detected while 47 access points with device (B). The number of available access points depends on the device's Wi-Fi sensitivity. The RSSI values are collected from all available access points for about 228 reference points which is obtained with the slow speed in every about one second. As for the fast speed data, about 171 reference points are collected. Note that the number of available reference points depends on the target area size or the walking distance. We collected data sets on different two days, considering the effect of newly installed and/or replaced APs. Finally, we obtained a total of 16 data sets, as shown in Tables 3.2 and 3.3 as columns.

The numbers of detected APs were 92 and 93 in total for Day 1 and Day 2, respectively. Some APs seem replaced, resulting in a total of 144 APs were detected in 16 data sets. In addition to the AP variation, the numbers of samples were different from each other data set due to the speed differences, as shown in Table 3.3. Furthermore, RSSI values themselves fluctuate due to various environmental effects. This fluctuation can be confirmed from Fig. 3.4, which shows the RSSI values of an AP (MAC address) in different data sets.

## 3.4 Experimental results

We implemented the proposed Wi-Fi-based indoor positioning system using the fingerprint method. The fingerprint database is created using data sets in Table 3.3. As mentioned in Sect. 3.3, the absolute timestamps for each reference point is also collected. The numbers of samples are different with each other due to speed change. Therefore, timestamp normalization is required. Assuming that the number of samples is $N_k$ in an obtained data set and that the timestamp of each sample is given by $t_k$, $k = 0, 1, \ldots, N-1$, the normalized timestamp is given by,

$$L_k = \frac{t_k - t_0}{t_{N_k-1} - t_0}, k = 0, 1, \ldots, N_k - 1. \tag{3.7}$$

Note that $L_0 = 0$, $L_{N_k-1} = 1$, and 1 corresponds to 120 m. When an obtained data set is used for a database creation or user's input, $L_k$ is regarded as the estimated reference point location, or the estimated user's ground truth location, denoted by $L_k^{\mathrm{T}}$, respectively. The user's position is estimated by the Wi-Fi based positioning approach based on these estimated reference points. However, due to the difference of sampling points, the Wi-Fi based estimation always includes error. Therefore, this error due to sampling can be regarded as the lower limit of the error.

By using the normalized timestamps for reference points, the estimated position error for user location $k$ is given by,

$$E_k = \min\left( \left| L_k^{\mathrm{T}} - \hat{L}_k^{\mathrm{T}} \right|, 1 - \left| L_k^{\mathrm{T}} - \hat{L}_k^{\mathrm{T}} \right| \right), \tag{3.8}$$

where $L_k^{\mathrm{T}}$ is the estimated ground truth, and $\hat{L}_k^{\mathrm{T}}$ is the estimated location using fingerprinting method. Here, we are using the assumption that the starting point and the end point is same. Note that, hence, the maximum error is not greater than 0.5.

To estimate the user's location, we need to define vector distances $D_{ui}$ between user's input vector $V_u$ and reference vector $V_i$. As mentioned in Sect. 3.2.2, $D_{ui}$ is calculated using Euclidean-like distance. In this experiment, we set 3 to $A$ in Eq. (3.5). This parameter $A$ controls the contribution of $N_j$ defined in Eq. (3.6), where larger $N_j$ gives a smaller distance. Also, thresholding based on $N_j$ is utilized to remove noise. If $N_j < 10$, we set a large value to the distance.

Figures 3.5, 3.9, 3.7, and 3.11 show the normalized errors using same data set for database creation and user's position estimation for devices (A) and (B),

respectively. Note that, in these figures, the lower limit errors due to the difference of sampling points are also included.

When the same data set is used for both database creation and user's position estimation, it is obvious that the error is small such as under 0.024 in the normalized error, which corresponds to about 3 m, as shown in Figs. 3.5, 3.7, 3.9, and 3.11. As for device (A), in Figs. 3.5 (a) to (d) and 3.9 (a) to (d), it can be seen that the error is larger in cases of the fast moving speed and reverse moving direction compared to the slow moving speed cases. As for device (B), in Figs. 3.7 (a) to (d) and 3.11 (a) to (d), the error is relatively small in all cases while some large errors can be found in Fig. 3.7 (a). This error in Fig. 3.7 (a) is because $P_i$ for some neighboring $i$ values are identical.

Figures 3.6, 3.8, 3.10, and 3.12 show the normalization errors using different data sets for datatbase creation and user's position estimation. When the different data sets are used for database creation and user's position estimation, larger error occurs compared to the cases when same data set is used, as shown in Figs. 3.6, 3.8, 3.10, and 3.12. Large errors occur in Figs. 3.6 (e), 3.8 (d), 3.10 (c), and 3.12 (c) where the maximum error is 0.21, 0.38, 0.485, and 0.33 in the normalized error, which corresponds to 25 m, 45 m, 39 m, and 58 m, respectively.

Figures 3.13 to 3.16 show the errors in cases when the data set of one device is used for database creation, the other device is used for user's position estimation. In Figs. 3.13 (a), (c), and (d), device (A)'s Slow 1 data set is used for database creation while the data sets of device (B) are used as user's input. In Figs. 3.13 (e), (g), and (h), device (B)'s Slow 1 data set is used for database creation while the data sets of device (A) are used as user's input. Large errors can be found when the database is created using device (B)'s data set. It seems that such tendency depends on the device specific Wi-Fi sensitivity.

In Table 3.4, finally, we summarize the maximum normalized error for each case. Apart from these results shown in this paper, we found that the error increases when the database is created by using data set with the small sampling number.

## 3.5 Conclusion

A Wi-Fi based indoor positioning system using estimated reference locations is proposed. In the proposed approach, the fingerprint database is constructed by gathering MAC-RSSI pairs using a reference device moving in a constant speed with sim-

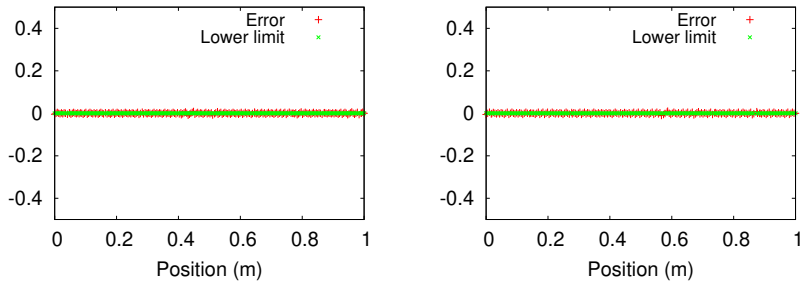
Table 3.4: Evaluation results of database creation (Slow 1, Device A, Day 1) using for different testing data sets.

| Database | Normalized error | Testing data sets, day 1 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Device A | | | | Device B | | | |
| | | Slow1 | Slow2 | Fast | Reverse | Slow1 | Slow2 | Fast | Reverse |
| Slow1 (Device A), day 1 | Avg | 0.003 | 0.022 | 0.028 | 0.045 | 0.045 | 0.048 | 0.065 | 0.029 |
| | Max | 0.009 | 0.073 | 0.104 | 0.197 | 0.011 | 0.119 | 0.485 | 0.108 |

| Database | Normalized error | Testing data sets, day 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Device A | | | | Device B | | | |
| | | Slow1 | Slow2 | Fast | Reverse | Slow1 | Slow2 | Fast | Reverse |
| Slow1 (Device A), day 1 | Avg | 0.040 | 0.052 | 0.053 | 0.069 | 0.034 | 0.072 | 0.026 | 0.078 |
| | Max | 0.143 | 0.450 | 0.465 | 0.455 | 0.169 | 0.488 | 0.088 | 0.486 |

ple direction. Assuming a constant speed, the position of each reference location can be estimated from the velocity to reduce administration cost while knowing the actual position of each reference location. Estimation accuracy evaluation results show that the proposed approach can estimate user's location without any precise reference point locations.

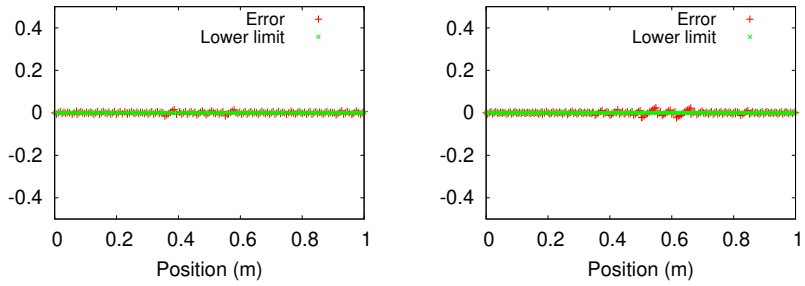
The accuracy is influenced by the number of reference locations stored in the database and the device-specific Wi-Fi sensitivity. New installation and replacements of APs also have a significant impact on the estimation accuracy. Considering these issues, not only the databases creation but also how to manage the database after the database creation is essential. It might be also possible that the user's input data is used for update the fingerprint data on the fly, utilizing the user's locality.




(a) DB: Slow 1, input: Slow 1




(b) DB: Slow 2, input: Slow 2

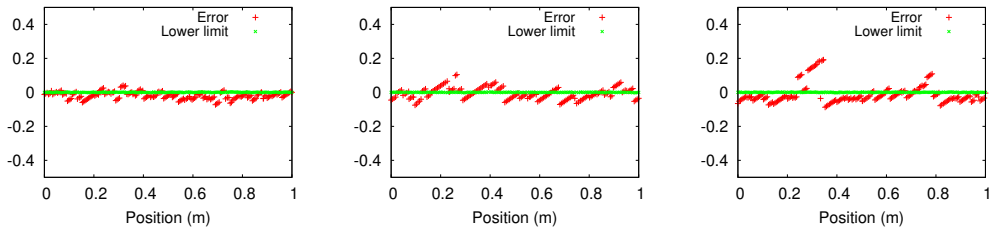



(c) DB: Fast, input: Fast

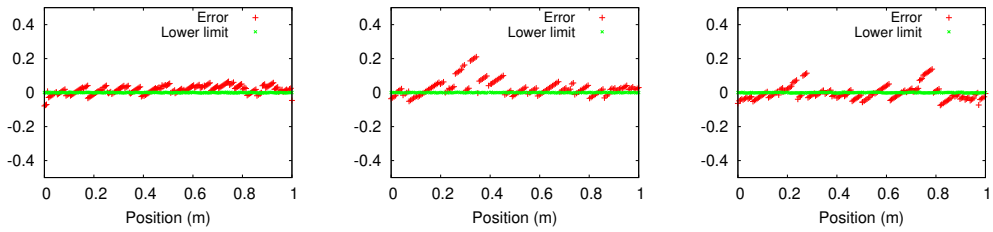



(d) DB: Reverse, input: Reverse

Figure 3.5: Normalized error of each estimated position using same data set for database creation and testing input data (Device (A)).

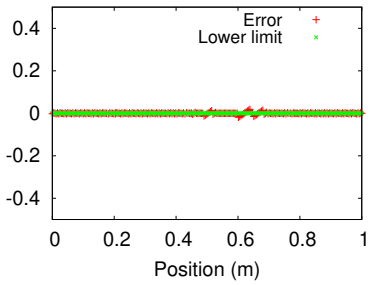



(a) DB: Slow 1, input: Slow 2 (b) DB: Slow 1, input: Fast (c) DB: Slow 1, input: Reverse

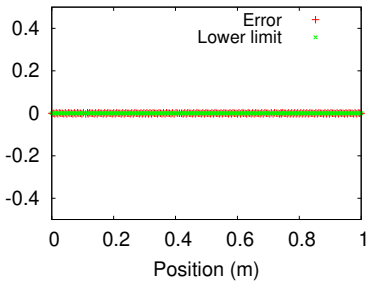



(d) DB: Slow 2, input: Slow 1 (e) DB: Slow 2, input: Fast (f) DB: Slow 2, input: Reverse

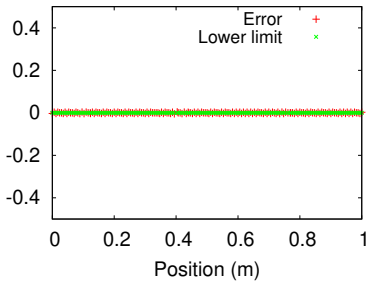
Figure 3.6: Normalized error of each estimated position using different data set for database creation and testing input data (Device (A)).



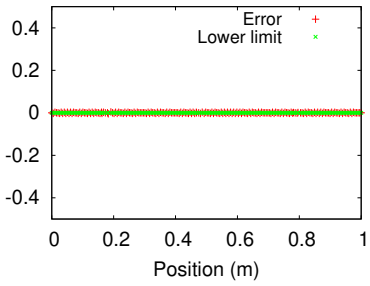

(a) DB: Slow1, input: Slow 1




(b) DB: Slow1, input: Slow 2

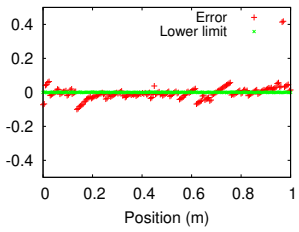



(c) DB: Slow1, input: Fast

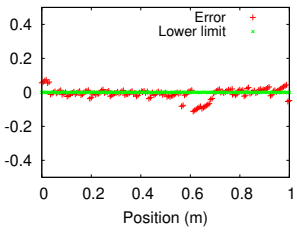



(d) DB: Slow1, input: Reverse

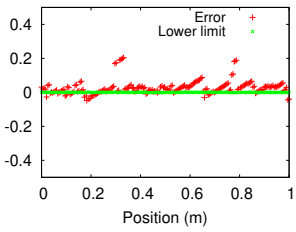
Figure 3.7: Normalized error of each estimated position using same data set for database creation and testing input data (Device (B)).



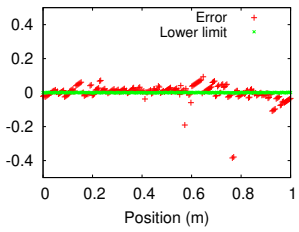

(a) DB: Slow1, input: Slow 2



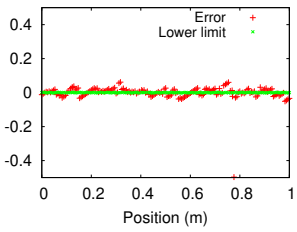

(b) DB: Slow1, input: Fast



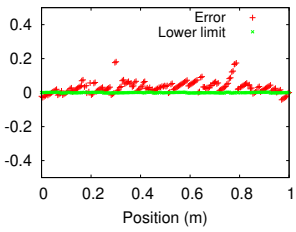

(c) DB: Slow1, input: Reverse




(d) DB: Slow 2, input: Slow 1

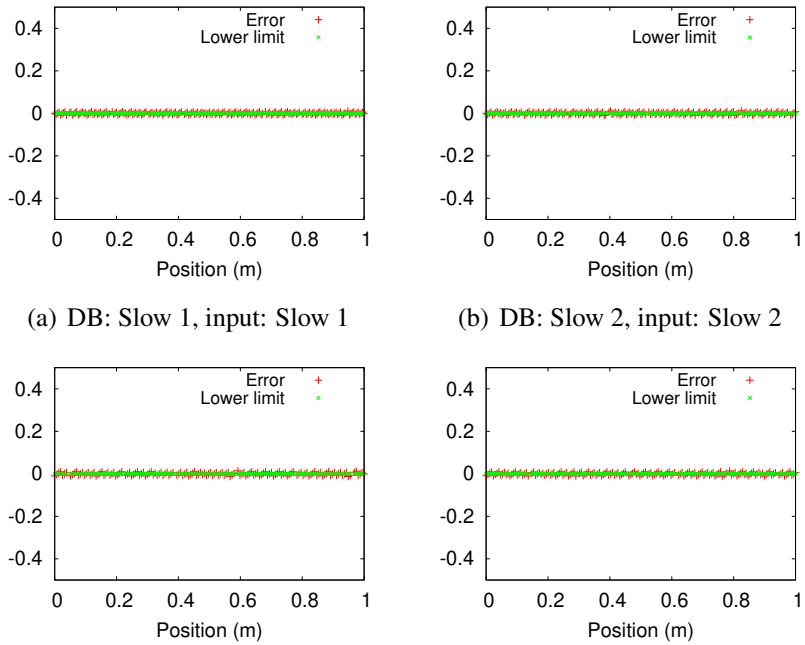



(e) DB: Slow 2, input: Fast




(f) DB: Slow 2, input: Reverse

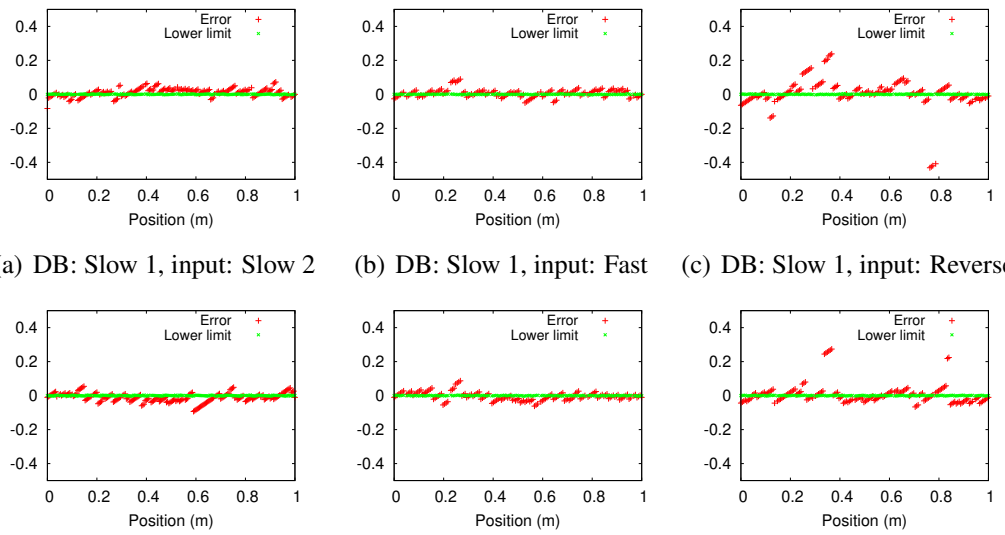
Figure 3.8: Normalized error of each estimated position using different data set for database creation and testing input data (Device (B)).




(a) DB: Slow 1, input: Slow 1 (b) DB: Slow 2, input: Slow 2

(c) DB: Fast, input: Fast (d) DB: Reverse, input: Reverse

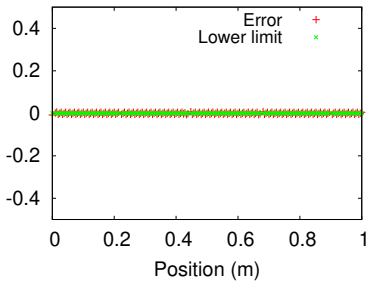
Figure 3.9: Normalized error of each estimated position using same data set for database creation and testing input data (Device (A), Day 2).



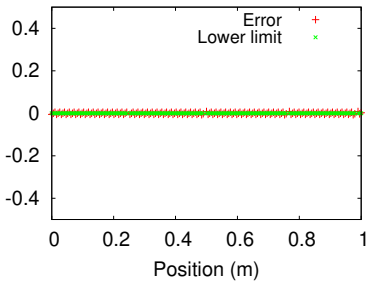

(a) DB: Slow 1, input: Slow 2 (b) DB: Slow 1, input: Fast (c) DB: Slow 1, input: Reverse

(d) DB: Slow 2, input: Slow 1 (e) DB: Slow 2, input: Fast (f) DB: Slow 2, input: Reverse

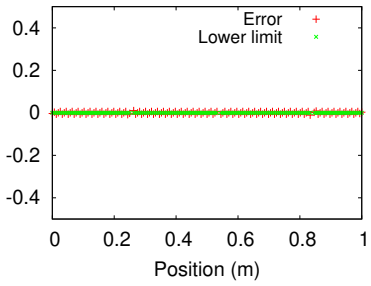
Figure 3.10: Normalized error of each estimated position using different data set for database creation and testing input data (Device (A), Day 2).



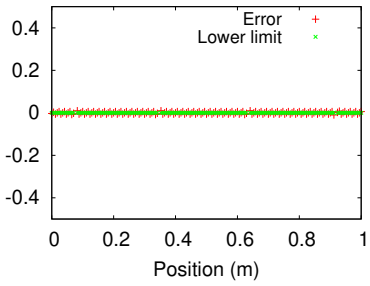

(a) DB: Slow1, input: Slow 1




(b) DB: Slow2, input: Slow 2

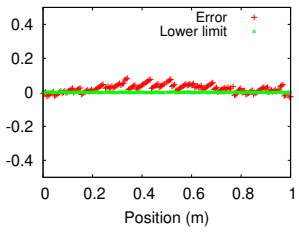



(c) DB: Fast, input: Fast

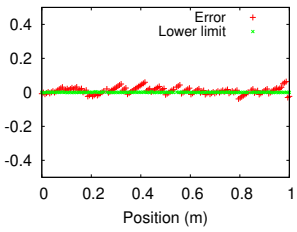



(d) DB: Reverse, input: Reverse

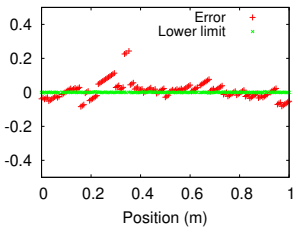
Figure 3.11: Normalized error of each estimated position using same data set for database creation and testing input data (Device (B), Day 2).



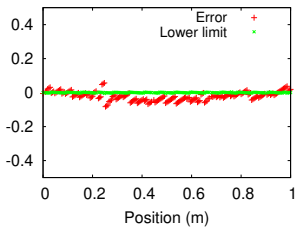

(a) DB: Slow1, input: Slow 2



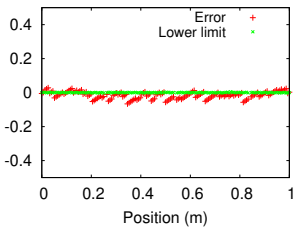

(b) DB: Slow1, input: Fast



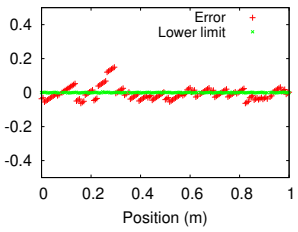

(c) DB: Slow1, input: Reverse




(d) DB: Slow 2, input: Slow 1

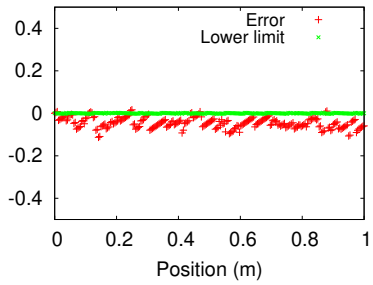



(e) DB: Slow 2, input: Fast

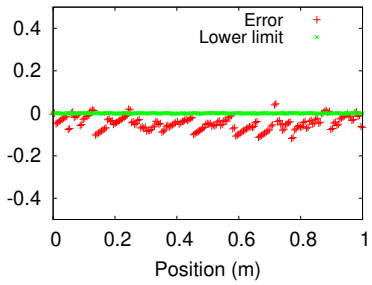



(f) DB: Slow 2, input: Reverse

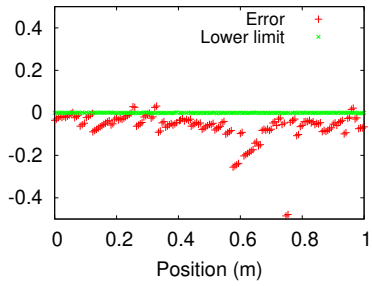
Figure 3.12: Normalized error of each estimated position using different data set for database creation and testing input data (Device (B), Day 2).



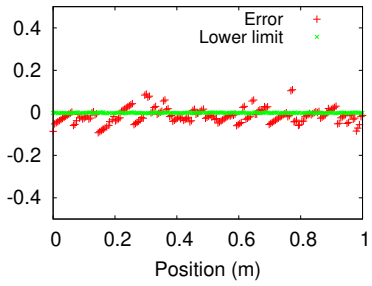

(a) DB: (A) Slow1, input: (B) Slow 1



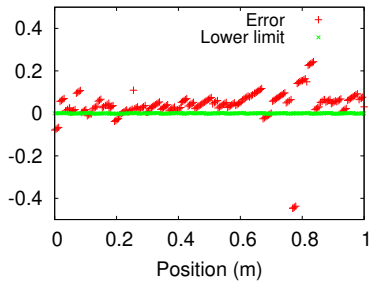

(b) DB: (A) Slow1, input: (B) Slow 2



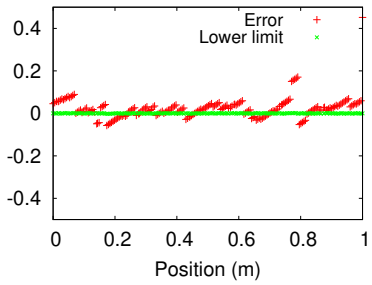

(c) DB: (A) Slow1, input: (B) Fast



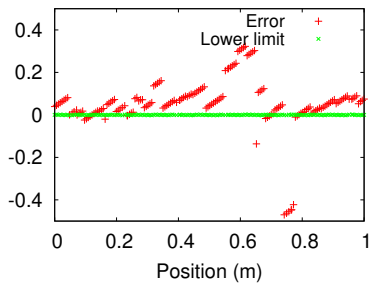

(d) DB: (A) Slow1, input: (B) Reverse



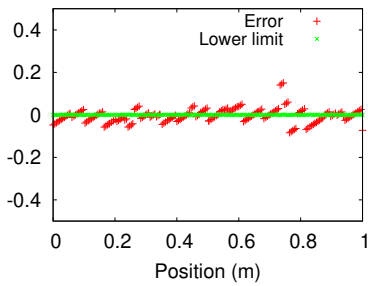

(e) DB: (B) Slow1, input: (A) Slow 1




(f) DB: (B) Slow1, input: (A) Slow 2

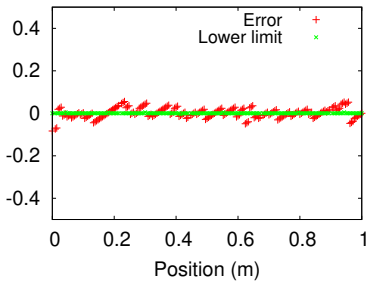



(g) DB: (B) Slow1, input: (A) Fast

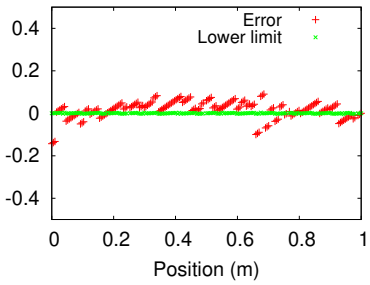



(h) DB: (B) Slow1, input: (A) Reverse

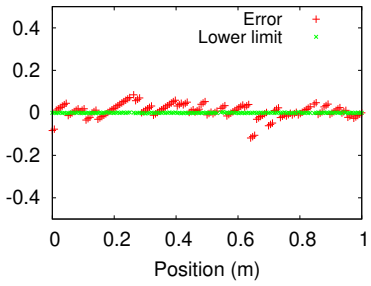
Figure 3.13: Normalized error of each estimated position (Different device data in same day (Day 1)).



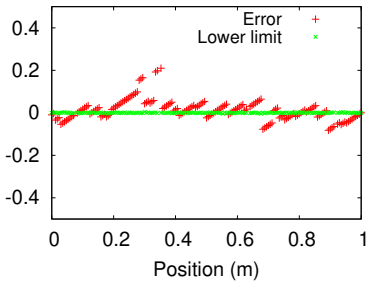

(a) DB: (A) Slow1, input: (B) Slow 1



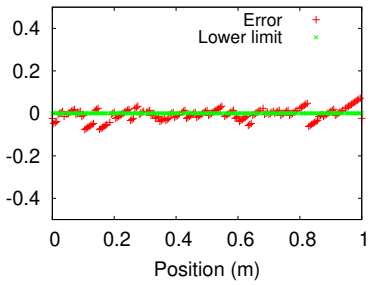

(b) DB: (A) Slow1, input: (B) Slow 2



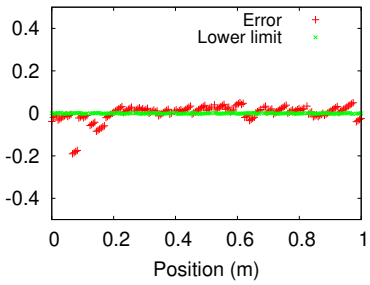

(c) DB: (A) Slow1, input: (B) Fast



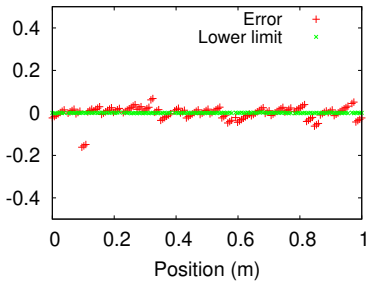

(d) DB: (A) Slow1, input: (B) Reverse



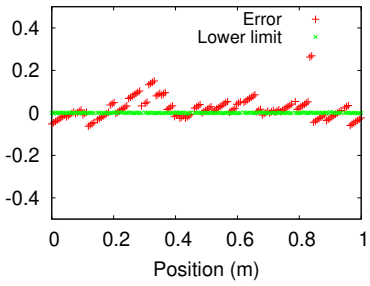

(e) DB: (B) Slow1, input: (A) Slow 1




(f) DB: (B) Slow1, input: (A) Slow 2

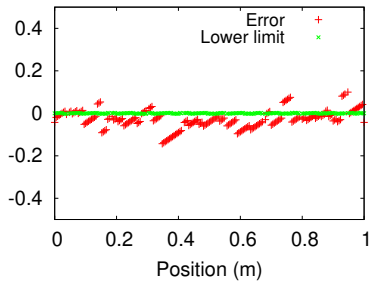



(g) DB: (B) Slow1, input: (A) Fast

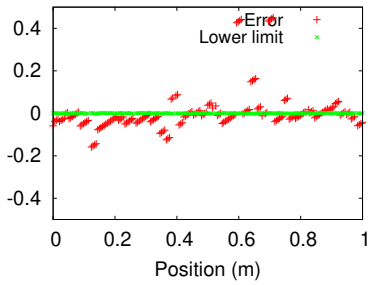



(h) DB: (B) Slow1, input: (A) Reverse

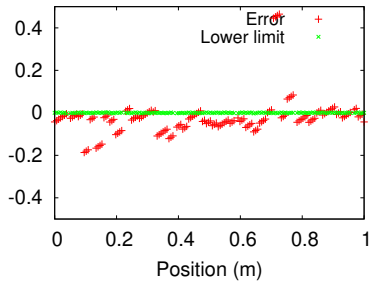
Figure 3.14: Normalized error of each estimated position (Different device data in same day (Day 2)).



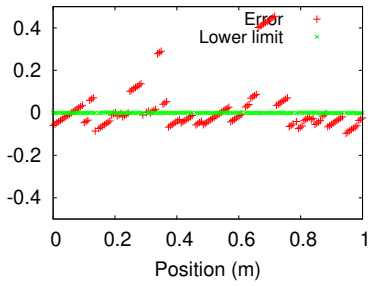

(a) DB: (D1) Slow1, input: (D2) Slow 1



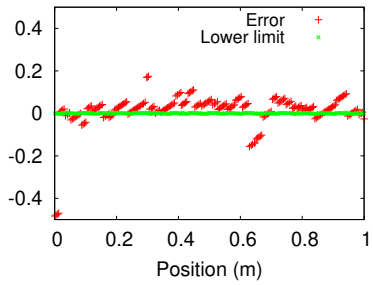

(b) DB: (D1) Slow1, input: (D2) Slow 2



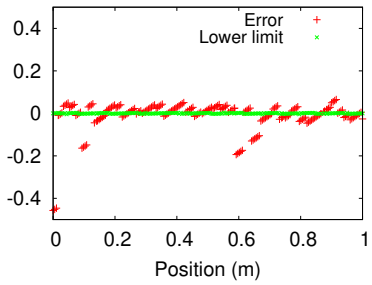

(c) DB: (D1) Slow1, input: (D2) Fast



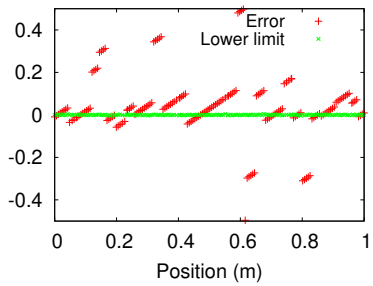

(d) DB: (D1) Slow1, input: (D2) Reverse



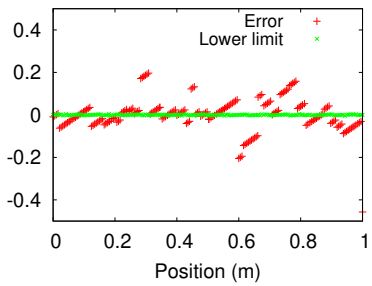

(e) DB: (D2) Slow1, input: (D1) Slow 1




(f) DB: (D2) Slow1, input: (D1) Slow 2

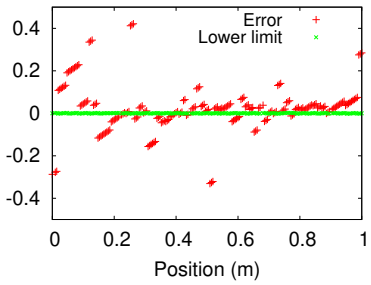



(g) DB: (D2) Slow1, input: (D1) Fast

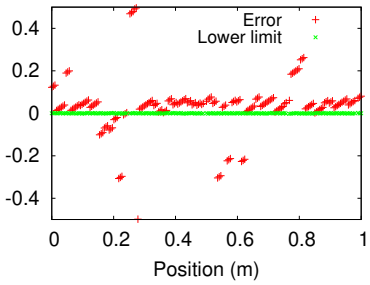



(h) DB: (D2) Slow1, input: (D1) Reverse

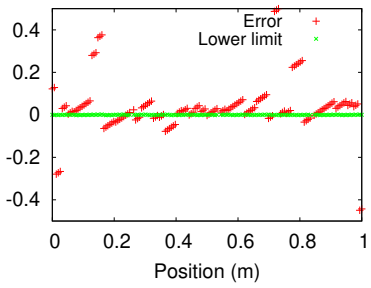
Figure 3.15: Normalized error of each estimated position (same device data in different day (Device A)).



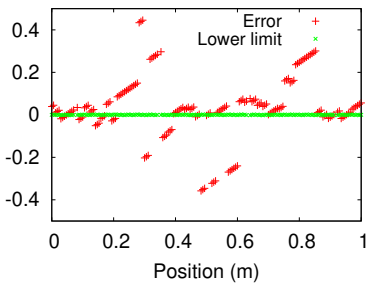

(a) DB: (D1) Slow1, input: (D2) Slow 1



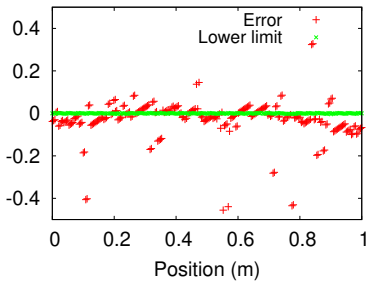

(b) DB: (D1) Slow1, input: (D2) Slow 2



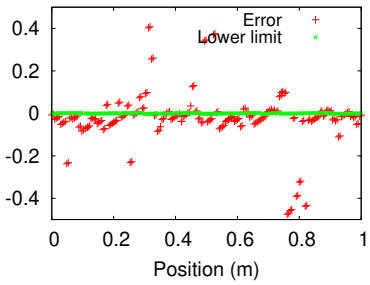

(c) DB: (D1) Slow1, input: (D2) Fast



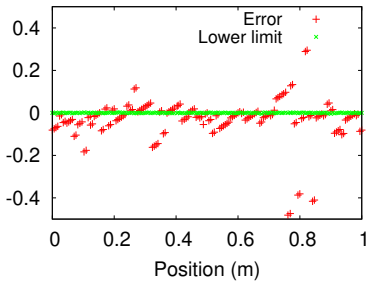

(d) DB: (D1) Slow1, input: (D2) Reverse



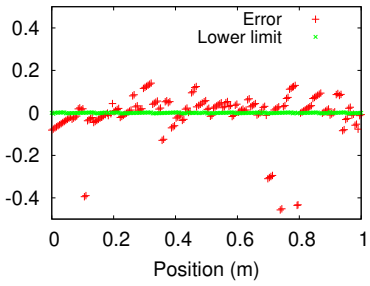

(e) DB: (D2) Slow1, input: (D1) Slow 1




(f) DB: (D2) Slow1, input: (D1) Slow 2




(g) DB: (D2) Slow1, input: (D1) Fast




(h) DB: (D2) Slow1, input: (D1) Reverse

Figure 3.16: Normalized error of each estimated position (same device data in different day (Device B)).

# Chapter 4

# Database construction

## 4.1 Introduction

Positioning systems play an essential role in many location-based applications for both indoor and outdoor environments. For the indoor positioning system (IPS), Wi-Fi-based IPS is the most cost-effective method since Wi-Fi coverage is getting wider due to the significantly increasing number of private and public Wi-Fi access points (APs) in the metropolitan area. Visible light [23] and sound-based [19] positioning systems are also popular with the high accuracy and usage of the available facilities inside the building. However, they all need calibration in order to determine to know the device's position in the coverage area. This calibration process has high administration costs. In the case of Wi-Fi-based IPS using the fingerprint method [9], AP locations are not required. However, we need to gather information of APs from various known reference locations stored in the fingerprint database.

In fingerprint based IPS, information of APs reachable from the user's location is used to estimate the user's location by comparing it with the pre-stored data in the database. In our previously proposed method in Chapter 3, the database was created with estimated reference locations, collecting information of all reachable APs using reference devices moving at a relatively constant speed in a specified direction. The location of each reference location can be estimated using the velocity. Thereby, the user's location can be estimated without any precise reference locations. However, the accuracy is influenced by the number of reference locations stored in the database and the device-specific Wi-Fi sensitivity.

The fingerprint database is established using received signal strength indicator

(RSSI) values from APs, with the mobile device scanning for RSSI values from all available APs. In a dense AP environment, not all APs may be observed in a single sampling time [54]. In these instances, when the mobile device is unable to receive RSSI values from some APs, these values are classified as unavailable RSSI values. Furthermore, RSSI values fluctuate due to various environmental effects such as hardware variations and resource collision caused by other devices utilizing the same service at the same time. This leads to an inaccurate estimation of the user's location by selecting an irrelevant reference location from the database. Another difficulty is that the database depends on the APs, which means new installations and replacements of APs have a significant impact on the estimation accuracy. Considering these issues, not only the database creation but also how to manage the database after the database creation is essential [55]. It might be also possible that the user's input data is used for update the fingerprint data on the fly, utilizing the user's locality.

To mitigate the challenges caused by the fluctuation of RSSI values, multiple data sets are used to create a single database with higher fidelity. The performance depends not only on RSSI values but also on the distance between reference locations. When we use dense reference locations to achieve high resolution, the amount of data stored in database increases. In this case, a high computational cost is required in the user's position estimation process [56].

Motivate by these challenges, we propose a Wi-Fi-based indoor positioning system using a fingerprint method with a novel data merging method to construct a consistent database. Multiple data sets are merged to get simple data set using two proposed merging methods. The first is the mean-shift clustering algorithm. The number of estimated reference locations depends on the thresholds parameter to calculate means in the mean-shift algorithm. In second proposed merging method, we use the grid size and window size to get merged data set from multiple data sets. The second proposed approach assumes that the intervals of reference locations in the database are constant and that the fingerprint for each reference location is calculated from multiple data sets. When merging data sets, RSSI values in a specific small area are averaged so that it can be expected to reduce noises due to the RSSI value fluctuation. We present a performance evaluation of our proposed system utilizing databases which are constructed under different conditions. We also use the user's location estimation algorithm from Sect. 3.2.2 to show the performance.

## 4.2 Related work

Wi-Fi-based indoor positioning system (IPS) is a positioning system that is used to locate objects or devices using the information from the Wi-Fi APs. RSSI is one of the most widely used cues for indoor positioning. The distance between the mobile device and each AP can be estimated using RSSI values. The mobile device position can be known based on trilateration or multilateration method using a measurement of the distances from three or more APs. In this case, however, the actual location of APs is required to be known. Another method is the fingerprinting, that does not require to install additional hardware such as dedicated wireless devices. In addition, the actual AP locations are not required. The user's position is estimated by finding the best match of the current RSSI values in an RSSI fingerprint database.

There are various research works in literature proposing Wi-Fi-based IPS with different approaches in fingerprint database creation and location estimation. In the traditional fingerprint method, the site survey process is required to know the actual reference locations in the coverage area. Most of the research works are using the previously determined reference locations during their fingerprint database constructions. While there are many efforts to know the actual position of reference locations in a large area, this site survey process is time-consuming, labor-intensive, and vulnerable to environmental changes.

To tackle this problem, many researchers have proposed techniques to reduce the required efforts. One approach is crowdsourcing [57], which collects the information of reference locations by various users from various devices without knowing the actual locations. In several research works [58, 59], the reference locations are estimated using the crowdsourcing approach to reduce the labor cost during the database creation process. Even in this case, however, the location assignment in the database using data from multiple users is still challenging.

In [59], the amount of data in the database is reduced. In this proposal, a large amount of reference locations are stored in the database by narrowing the distance between the reference locations. In this case, the user's position estimation process requires high computational cost. In our work, the mean-shift clustering algorithm [60] is used to estimate reference locations. The amount of reference locations can be controlled using a mean-shift parameter. In this case, the difficulty is updating and maintaining the database since the mean-shift requires high computational cost.

An additional issue is related to the RSSI value fluctuation, which leads to an inaccurate user position estimation. The effect of RSSI value fluctuation can be

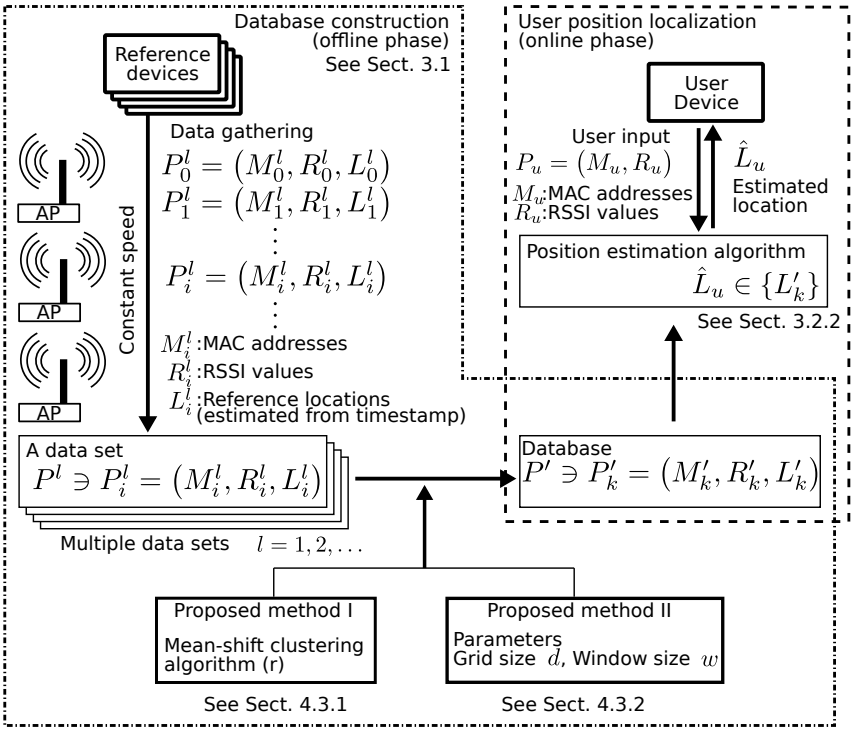



Figure 4.1: Overview of the proposed system.

reduced by repeating data collection in each reference location [52]. In the case of the crowdsourcing method, however, it is difficult to get RSSI values repeatedly in the same reference location from multiple users. This paper aims to provide a flexible and low-cost indoor positioning system that brings the advantage of estimated reference locations using the information of APs from multiple devices, and that does not need precise location information of APs. The proposed approach has the capability to use the information from multiple users.

## 4.3 Proposed approaches

In this section, we present our proposed approach. As described in Sect. 3.2.1, the set of MAC-RSSI pairs for reference location $i$ is collected by mobile devices as $P_i$, and the list of $P_i$ is stored in the database as $P$. In this work, we define *a data set* as $P^l$ similarly to $P$. The fingerprint database is created using multiple data sets collected at different times with different devices. The multiple data sets $P^l$ are

merged to realize a single, more consistent data set $P'$ to be stored in the database.

The number of reference locations is different with each data set. The actual location of reference location $i$ in each data set is denoted by $L_i^l$. $L_i^l$ is estimated from the timestamps. Similarly to the definitions in the previous section, we define some sets as follows,

- the set of MAC addresses at reference location $L_i^l$:
  $M_i^l = \{M_{i1}^l, M_{i2}^l, \ldots, M_{iN_i^l}^l\}$,
  where $N_i^l$ is the number of available APs,

- the RSSI values paired with $M_i^l$:
  $R_i^l = \{R_{i1}^l, R_{i2}^l, \ldots, R_{iN_i^l}^l\}$, and,

- the set of MAC-RSSI pairs at reference location $L_i^l$:
  $P_i^l = \left(M_i^l, R_i^l, L_i^l\right)$,

where list $P_i^l$ corresponds to a data set $P^l$, and $P_i^l \in P^l$. The merged data set $P'$ form data sets $P^l$ is stored in the database. We used two proposed approaches to obtain a merged data set such as mean-shift clustering algorithm and grid size.

### 4.3.1 Proposed fingerprint database construction using mean-shift clustering

In this section, we present our proposed approach to estimate the reference locations using mean-shift clustering algorithm for database creation. In order to obtain a merged data set $P'$ from data sets $P^l$, we need to assign an mean-shift parameter ($r$) to obtain the reference location $k$, where $P_k' = \left(M_k', R_k', L_k'\right) \in P'$. The number of estimated reference locations in the database depends on the mean-shift parameter ($r$), radius for the mean calculation to settle each reference location as the result of clustering.

The procedure of merging multiple data sets to get simple data set to be stored in server as fingerprint database as follows:

(1) For each $P_i^l$ from data sets $P^l$, we selected the locations $j$ which distance from location $i$ in parameter ($r$) as

$$P_j^l = \{j | L_i^l - r \leq L_i^l \leq L_i^l + r\}. \tag{4.1}$$

(2) The mean value $\bar{L}_i$ for each location from data set is calculated as

$$\bar{L}_i = \frac{\sum_{\{j|r \leq |L_i - L_j| \leq r\}} L_j}{\sum_{\{j|r \leq |L_i - L_j| \leq r\}} 1}. \tag{4.2}$$

The location i moved to $\bar{L}_i$ as $L'_i$ for $P'_i$.

(3) In order to obtain $R'_i$ for each location $L'_i$, the list of unique MAC addresses is created from the selected locations whose distance from location i is less than equal parameter $r$ that describes as

$$M'_k = \bigcup_l \bigcup_{i \in P_i^l} M_i^l. \tag{4.3}$$

(4) For each $L'_k$, using the above unique MAC addresses $M'_k$, $R'_{kj'} \in R'_k$ is calculated by averaging all RSSI values as follows,

$$R'_{kj'} = \frac{1}{\sum_{\{i,l,j|M_{ij}^l = M'_{kj'} \in M'_k\}} 1} \sum_{\{i,l,j|M_{ij}^l = M'_{kj'} \in M'_k\}} R_{ij}^l \tag{4.4}$$

A set of MAC-RSSI pairs $(M'_k, R'_k)$ for reference location $L'_k$ can be obtained, resulting in a new data set $P'$, where $P'_k = \left(M'_k, R'_k, L'_k\right)$.

(5) Repeats calculation step from 2 to 4 until $L'_k$ converges.

(6) Finally, a new data set $P'$ is stored in sever as fingerprint database.

### 4.3.2 Proposed fingerprint database construction using grid size and window size

In this section, we present our proposed approach to estimate the reference locations using grid size and window to obtain a merged data set $P'$ from data sets $P^l$ for database creation. We need to assign an actual location $L_k$ of reference location $k$, where $P'_k = \left(M'_k, R'_k, L'_k\right) \in P'$. In this paper, we assume a constant interval distance between every neighboring reference locations in $P'$. We call this constant interval *a grid size* $d$ and the actual location of the reference location $P'_k$ is denoted by,

$$L'_k = L'_{k-1} + d. \tag{4.5}$$

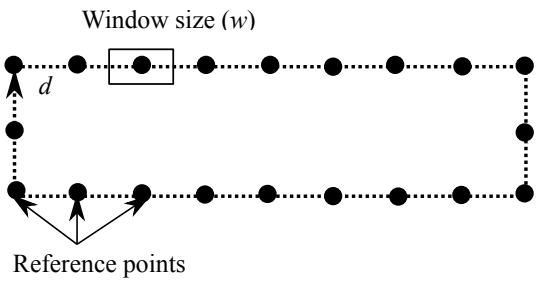



Figure 4.2: Relationship between reference points, grid size $d$, and a window whose size is $w$.

In order to obtain $M'_k$ and $R'_k$ for each $L'_k$ from data sets $P^l$, *a window*, whose size we call *window size* $w$, is used as shown in Fig. 4.2. In this paper, we set window size $w$ proportional to the grid size $d$ as the following

$$w = Bd, \tag{4.6}$$

where $B$ is a scale factor. We select locations from each data set $P^l$ using the window for each $L'_k$ by

$$S^l_k = \{i \big| L'_k - \frac{w}{2} \leq L^l_i \leq L'_k + \frac{w}{2}\}. \tag{4.7}$$

From the selected locations $S^l_k$, the list of unique MAC addresses related to $L'_k$ is created as

$$M'_k = \bigcup_l \bigcup_{i \in S^l_k} M^l_i. \tag{4.8}$$

For each $L'_k$, using the above unique MAC addresses $S^l_k$, $R'_{kj'} \in R'_k$ is calculated by averaging all RSSI values in the window over all data sets $P^l_i = \left(M^l_i, R^l_i, L^l_i\right)$ as follows,

$$R'_{kj'} = \frac{1}{\sum_{\{i,l,j|M^l_{ij}=M'_{kj'} \in M'_k\}} 1} \sum_{\{i,l,j|M^l_{ij}=M'_{kj'} \in M'_k\}} R^l_{ij} \tag{4.9}$$

Finally, a set of MAC-RSSI pairs $(M'_k, R'_k)$ for reference location $L'_k$ can be obtained, resulting in a merged data set $P'$, where $P'_k = \left(M'_k, R'_k, L'_k\right)$. To estimate the user location, this $P'$ is used in the same way to the case of $P$ in described in Sect. 3.2.2.

Table 4.1: Data sets for the fingerprint database construction.

| Condition | Data sets used to create the database | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Day 1 | | | | | | | | Day 2 | | | | | | | |
| | Device A | | | | Device B | | | | Device A | | | | Device B | | | |
| | Slow1 | Slow2 | Fast | Reverse | Slow1 | Slow2 | Fast | Reverse | Slow1 | Slow2 | Fast | Reverse | Slow1 | Slow2 | Fast | Reverse |
| Cond. A | 1 | **2** | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Cond. B | 1 | □ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Cond. C | 1 | **2** | 3 | 4 | | | | | 9 | 10 | 11 | 12 | | | | |
| Cond. D | | □ | | | 5 | 6 | 7 | 8 | | | | | 13 | 14 | 15 | 16 |

Here, we assume that data set 2 is used as the test input data as an example.
Cond. A: uses all data sets including the test input data set,
Cond. B: uses all data sets excluding the test input data set,
Cond. C: uses data sets from the same device used for test, and
Cond. D: uses data sets from the other device used for test.

## 4.4 Experimental results and discussion

We implemented the proposed Wi-Fi-based indoor positioning system using the fingerprint method. The fingerprint database is created using multiple data sets collected at different times with different devices. Since the numbers of samples are different from each other data set due to the speed differences, as shown in Table 3.3, a normalization is required. Assume that the timestamp of each sample is given by $t_i$, $i = 0, 1, \ldots, N^l - 1$, where the number of samples is $N^l$, the normalized timestamp is given by,

$$L_i^l = \frac{\left(t_i - t_0\right)}{\left(t_{N^l-1} - t_0\right)}, i = 0, 1, \ldots, N^l - 1. \tag{4.10}$$

Note that $L_0^l = 0$, $L_{N^l-1}^l = 1$, and 1 corresponds to 120 m as shown in Fig. 3.3. $L_i^l$ can be regarded as the normalized location for each data set $P_i^l = \left(M_i^l, R_i^l, L_i^l\right)$. When an obtained data set is used for the user's testing input, the estimated user's ground truth location is denoted by $L_u^{\mathrm{T}}$ for user location $u$.

By merging multiple data sets using the mean-shift clustering and the grid size approaches described in Sect. 4.3.1 and 4.3.2, the merged data set $P'$ can be obtained where $L'_k$ is regarded as an estimated reference location. Note that $P'_k = \left(M'_k, R'_k, L'_k\right) \in P'$, $k = 0, 1, \ldots, N' - 1$, where $N'$ is the number of estimated reference locations in the merged data set $P'$, and that $N'$ depends on the mean-shift parameter $r$ in mean-shift approach and depends on the grid size $d$ in the constant interval using grid size approach. Hereafter, we use normalized values for the mean-shift parameter $r$, the grid size $d$, the window size $w$, and estimated position error.

In the experiment for mean-shift approach, we set several parameter $r$, from

0.001 to 0.1, to calculate the mean value $\bar{L}_i^l$ for each reference location $P_i^l$ in data sets $P^l$. The location $L_i^l$ moved to $\bar{L}_i^l$ and calculated again the mean value for each location for each reference location The RSSI values for $\bar{L}_i^l$, we calculated by averaging all RSSI values of $L_i^l$, $L_i^l - r \leq L_i^l \leq L_i^l + r$. When $\bar{L}_i^l$ converges, $\bar{L}_i^l$ is $L'_k$ for estimated reference location $P'_k$.

In the gird size approach, we use several $N'$ values, $10 \leq N' \leq 500$, from each of which, the grid size $d$ can be calculated as follow,

$$d = \frac{1}{N'}. \tag{4.11}$$

To calculate RSSI values for $L'_k$, that is $R'_k$, we use the window whose size is $w = Bd$, as described in Sect. 4.3.2. In the experiment, we set 0.1, 0.5, 1.0, 1.5, and 2.0 to $B$. The grid size $d$ controls the window size $w$, where a smaller $d$ gives a smaller $w$. We believe that such window size $w$ proportional to the grid size $d$ is reasonable, and that this makes the experiment easy since the value of $w$ can be controlled by setting $d$.

If $B < 1$, the window size $w$ is smaller than $d$. If $B < 1$ and if $d$ is relatively small, the set of MAC addresses related to $L'_k$, which is $M'_k$ defined in Eqs. (4.7) and (4.8), is sometimes empty. In the experiment, such cases are excluded. Note that a larger window size gives a higher smoothing effect on averaging, which may reduce noises in RSSI values.

By using the normalized timestamps for reference locations, the estimated position error for user location $u$ is given by,

$$E_u = \min\left( \left| L_u^{\mathrm{T}} - \hat{L}_u^{\mathrm{T}} \right|, 1 - \left| L_u^{\mathrm{T}} - \hat{L}_u^{\mathrm{T}} \right| \right), \tag{4.12}$$

where $L_u^{\mathrm{T}}$ is the estimated ground truth, and $\hat{L}_u^{\mathrm{T}}$ is the estimated location using the fingerprinting method. Since the estimated location $\hat{L}_u^{\mathrm{T}}$ is always aligned on the grid, that is $\hat{L}_u^{\mathrm{T}} \in \{L'_k\}$, the estimation of the proposed approach always includes an error. Hereafter, we call such unavoidable error *the lower limit error*. Also, we used the assumption that the starting point and the end point is the same. Note that, hence, the maximum error is not greater than 0.5.

To estimate the user's location, we need to define vector distances $D_{ui}$ between user's input vector $V_u$ and reference vector $V_i$. Hereafter, we use the variable $i$ both for $P$ and $P'$, considering the consistency to the description in Sect. 3.2.2. As mentioned in Sect. 3.2.2, $D_{ui}$ is calculated using Euclidean-like distance. In this

experiment, we set 3 to $A$ in Eq. (3.5). This parameter $A$ controls the contribution of $N_j$ defined in Eq. (3.6), where larger $N_j$ gives a smaller distance. Also, thresholding based on $N_j$ is utilized to remove noise. If $N_j < 10$, we set a large value to the distance.

We created four types of databases to test the fidelity of the data sets. We denote these types as Conditions A, B, C, and D as listed below,

Cond. A uses all data sets including the test input data set,

Cond. B uses all data sets excluding the test input data set,

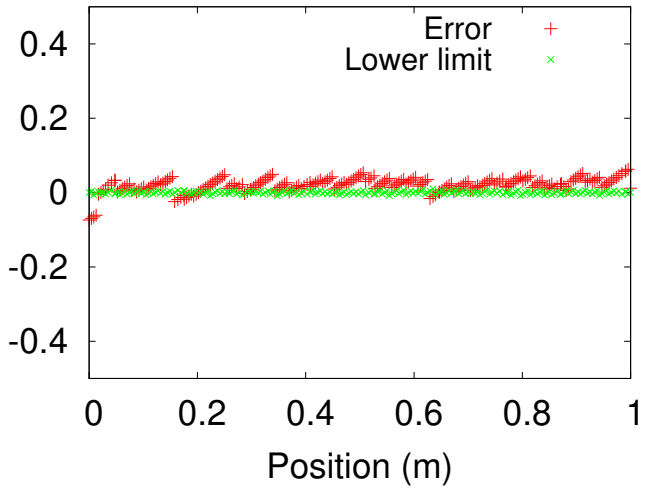
Cond. C uses data sets from the same device used for the test, and

Cond. D uses data sets from the other device used for the test.

In each condition, we created databases for each grid size $d$ with different window sizes $w$. Table 4.1 illustrates which data sets are used for database construction for each condition, where data set 2 is assumed as the test input data set as an example. We use 16 input data sets for each condition to get averaged errors in the following part in this section. In the case of Cond. A, 16 input data sets are used with a single database. In the case of Cond. B, each input data set is used with the database constructed without the input data set, resulting in 16 databases are used in total. Similarly, as for Conds. C and D, two device-specific databases are constructed and used for testing.

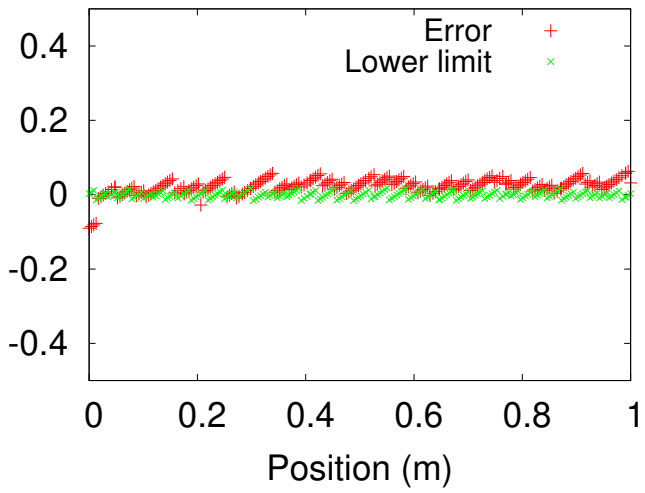
To see the effect of the mean-shift parameter $r$ and the grid size $d$, Figs. 4.3 to 4.6 and Figs. 4.7 to 4.10 show the normalized errors of each estimated position for $r = 0.005$, 0.01, 0.05, and 0.1, and $d = 0.005$, 0.01, 0.05, and 0.1, respectively. To get this result, Cond. A and input data set 2 are used. In Figs. 4.7 to 4.10, we set the window size $w = d$. Note that, in this figure, the lower limit errors due to the discrete reference locations are also included. Both the estimation error and the lower limit error depend on the mean-shift parameter $r$ and the grid size $d$. Large errors occur in the case of a large $r$ and $d$, where the maximum error is 0.49 and 0.46, respectively. In Figs. 4.7 to 4.8, the average error is 0.029, 0.003, 0.071, and 0.089, each of which corresponds to 3.48 m, 3.60 m, 8.52 m, and 10.68 m, respectively. In Figs. 4.7 to 4.8, the average error is 0.021, 0.024, 0.034, and 0.054, each of which corresponds to 2.52 m, 2.88 m, 4.08 m, and 6.48 m, respectively.

To evaluate the proposed methods, we compared the performance using the mean-shift clustering algorithm with constant interval using grid size. Figures 4.11



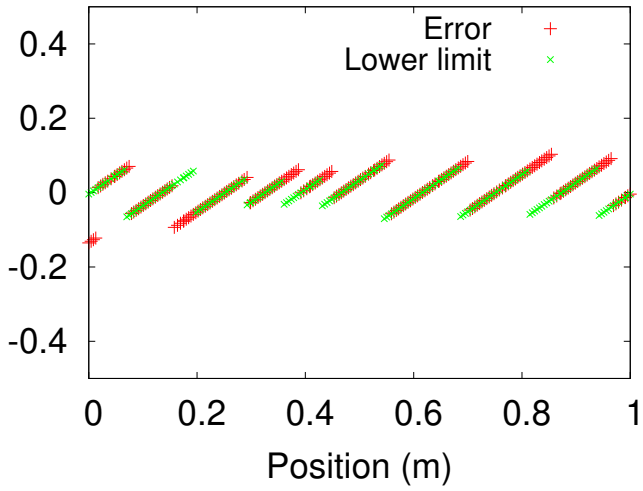

$r = 0.005$ (0.6 m), $N' = 81$
Average error: 0.029 (3.48 m)

Figure 4.3: Normalized error of each estimated position. Cond. B and input data set 2 are used.



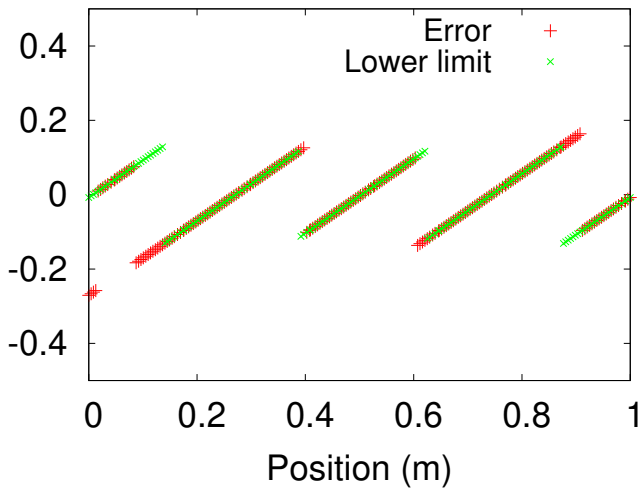

$r = 0.01$ (1.2 m), $N' = 38$
Average error: 0.003 (3.6 m)

Figure 4.4: Normalized error of each estimated position. Cond. B and input data set 2 are used.



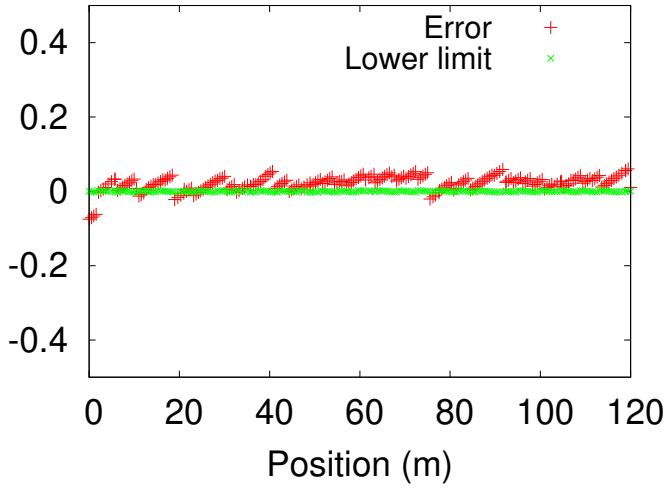

$r = 0.05$ (6 m), $N' = 8$
Average error: 0.071 (8.52 m)

Figure 4.5: Normalized error of each estimated position. Cond. B and input data set 2 are used.



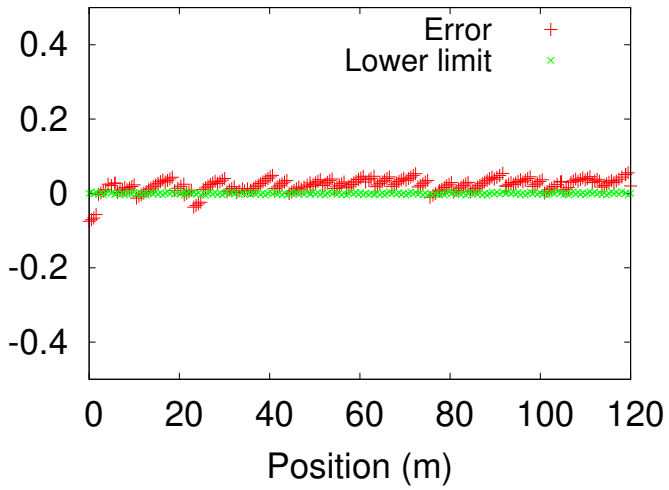

$r = 0.1$ (12 m), $N' = 4$
Average error: 0.089 (10.68 m)

Figure 4.6: Normalized error of each estimated position. Cond. B and input data set 2 are used.



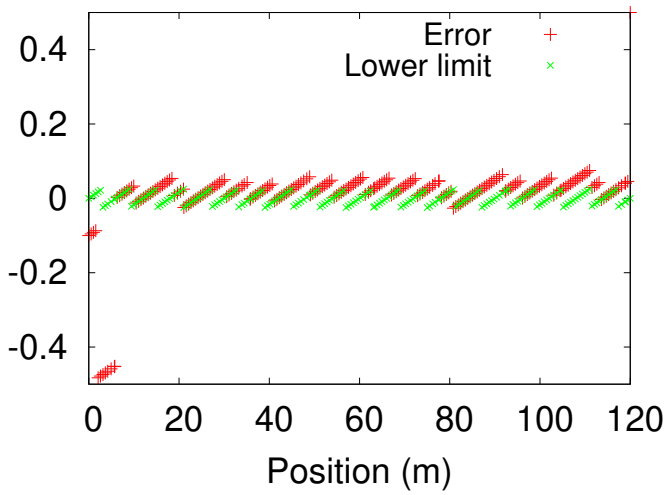

$d = 0.005$ (0.6 m), $N' = 200$
Average error: 0.021 (2.52 m)

Figure 4.7: Normalized error of each estimated position. Cond. A and input data set 1 are used. The window size is given by $w = d$.



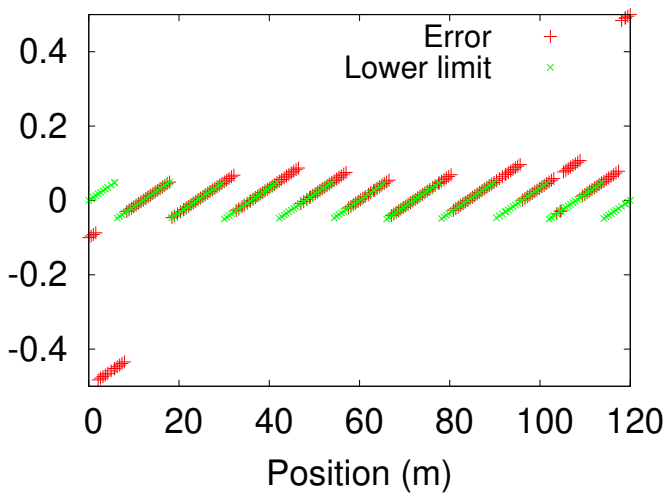

$d = 0.01$ (1.2 m), $N' = 100$
Average error: 0.024 (2.88 m)

Figure 4.8: Normalized error of each estimated position. Cond. A and input data set 1 are used. The window size is given by $w = d$.




$d = 0.05$ (6 m), $N' = 20$
Average error: 0.034 (4.08 m)

Figure 4.9: Normalized error of each estimated position. Cond. A and input data set 1 are used. The window size is given by $w = d$.




$d = 0.1$ (12 m), $N' = 10$
Average error: 0.054 (6.48 m)

Figure 4.10: Normalized error of each estimated position. Cond. A and input data set 1 are used. The window size is given by $w = d$.

and 4.12 present the average errors for each condition of database creation using the mean-shift clustering algorithm and using grid size, respectively. The average error for each graph is calculated over all 16 test inputs, as mentioned before. From these figures, it can be said that the grid size approach's performance is superior to the mean-shift approach. Note that, in the case of the mean-shift approach, similarly to the grid size approach, the number of estimated reference locations in the database depends on the mean-shift parameter, a radius for the mean calculation to settle each reference location as the result of clustering.

In Fig. 4.12, we compared the errors using different window sizes. The small grid size $d$ corresponds to a large number of estimated reference locations $N' = 1/d$, as shown in Eq. (4.11). Similar two approaches, a larger error occurs in cases having a small number of reference locations. Also, it can be seen that, in our proposed approach, the estimated position error reduces in cases having a large number of reference locations. The RSSI values for each estimated reference location, that is $R'_k$ for location $L'_k$, depend on the window size. When the window size is too small in the case of small grid sizes, the noise cannot be eliminated. In Fig. 4.12, we can find several $d$ values where the error in $w = 0.1d$ is larger than that in $w = 0.5d$. In the case of large window size, the window overlaps between the neighboring reference locations. We expected that a larger window size gives a higher smoothing effect on averaging, which may reduce noises in RSSI values. However, a larger window size gives large errors, as we can see in Fig. 4.12. In Fig. 4.12, the larger errors occur in large grid sizes, especially when the window size is larger than $1.0d$. This implies that, in this experiment, the window overlap causes estimation error since an overlapped window includes some reference locations from the neighboring reference locations.

According to Fig. 4.12, it can be seen that small grid sizes give smaller errors than large grid sizes. However, some large errors occasionally occur in some small grid sizes. This is due to the high sensitivity to noise caused by the effect of unavailable RSSI values in small grid size coupled with a small window size. When the grid size $d$ is less than 0.02 coupled with the smallest window size of $w = 0.1d$, the number of the selected locations $S^l_k$ (Eq. (4.7)) is not enough to create a stable $P'_k = \left(M'_k, R'_k, L'_k\right)$.

For further discussion about the window size, in Fig. 4.13, we show the error distribution with various grid sizes and window sizes $w = 0.1d$ to $2d$ in Condition B. The error distribution depends on both the grid size and the window size. As mentioned in the former part of this section, $M'_k$ becomes empty when both the

window size and the grid size are small. Such cases are excluded in the figure. From this figure, we can see that the error distribution for each grid size has a peak whose value is relatively small in both cases, $w = 0.1d$ and $2d$. The case of $w = 2d$ gives a sharp error distribution when the grid size is small, compared with the case of $w = 0.1d$. However, when the grid size is large, the error distribution of $w = 2d$ is not good, while the case of $w = 0.1d$ gives a relatively stable performance. This is because the absolute window size becomes large when the grid size is large since we use the relationship of $w = Bd$.

As mentioned in the former part of this section, the lower limit error due to the estimated location alignment to the grid is unavoidable. As we can see from Figs. 4.7 to 4.10, the larger the grid size is, the larger the error becomes. However, it can also be seen from Figs. 4.7 to 4.10 and 4.12 that the difference between the error and its lower limit is relatively small when the grid size is large. Figure 4.14 shows the average estimation error from the lower limit for each condition. From this figure, we can see that the grid sizes from 0.06 to 0.08 give better performance with window sizes not greater than $1.0d$. Note that Fig. 4.14 is obtained from Fig. 4.12, just subtracting the lower limit error.

Considering that the average error becomes close to its lower limit when the grid size is large, we define a new performance metric named *accuracy*. We assume that the estimation is correct when $|E_u| < d/2$, where the estimated user location belongs to the non-overlapped area centered by the grid. This assumption is reasonable for some applications where a precise position is not required. Using this assumption, we define accuracy as follows,

$$\text{Accuracy} = \frac{\text{No. of correctly estimated locations}}{\text{No. of test input locations}}. \tag{4.13}$$

Figure 4.15 shows the accuracy for each condition with various grid sizes and window sizes. From this figure, we can see that the large grid sizes and the small window sizes give high accuracy while the small grid sizes give low accuracy. The proposed approach can archive an accuracy of 80 %. When the grid size is small, the number of reference locations is relatively large. This large number of reference locations results in a high likelihood of selecting incorrect reference locations from the database when estimating the user location.

Finally, as for the database construction conditions, according to Figs. 4.12, 4.14, and 4.15, Cond. C, which is created from data sets of the same device for the

testing, gives the best performance. Also, Cond. A, which is created from all data sets, gives better performance than Conds. B and C. Comparing the average errors of Conds. A and B, the performance is slightly better when the testing input is included in the database. From these results, we can conclude that the fingerprint database should include data sets from a device assumed to be used in the real application. Note that our approach can archive an average error of 0.02 corresponding 2.4 m when the grid size is small in Cond. A, B, and C.

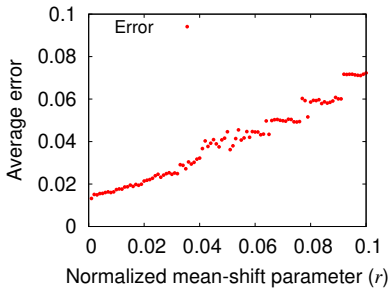
## 4.5 Conclusion

In this paper, we proposed a Wi-Fi-based indoor positioning system using a fingerprint method with a novel merging approach to construct a consistent database from multiple data sets. Since each data set is acquired by moving reference devices at a constant speed, the reference locations in the data set can be estimated using the velocity without any precise reference location information. We successfully investigated two merging methods to improved database management in our proposed system.

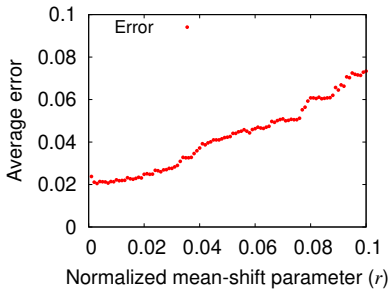
The first merging method is the mean-shift clustering algorithm is used to get merged reference locations form the multiple data sets. The number of estimated reference locations and location of each estimated reference locations in merged data set depend on the mean-shift parameter, a radius for the mean calculation to settle each reference location as the result of clustering. The merged data set is stored in sever as database. The large mean-shift parameter gives high computation time in each reference location calculation compare with small parameter. However, the large error occurs in the large mean-shift parameter. This proposed approach can be reduced the accuracy error due to the estimation accuracy evaluation results that compared with the database are constructed using one data set.

The second merging method involved the number of reference locations and noise reduction in RSSI value which depend on the grid size and the window size. When merging data sets, RSSI values are averaged in the window of each estimated reference location in the new database. We assumed a constant interval named grid size between neighboring reference locations. The number of reference location in merged data set from multiple data sets depends on the grid size. We showed the effectiveness of the proposed approach through the experimental results, including the errors and accuracy for various grid sizes and window sizes. The large grid sizes

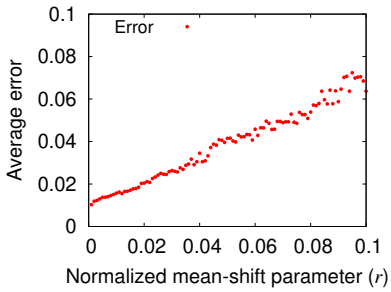
give high accuracy but large error while the small grid sizes give a low error but low accuracy. This means that, in the proposed merging method, the performance can be controlled, changing the grid size depending on applications. Note that the window size $w$ can be properly set for each grid size $d$ regarding the experimental result. When $d$ is small, a normalized average error of 0.02 corresponding 2.4 m can be archived. Also, when $d$ is large, an accuracy of 80 % can be archived. Comparison of two merging method, updating and maintaining the fingerprint database require high computation cost using mean-shift clustering algorithm, although the mean-shift clustering algorithm is popular in data clustering.



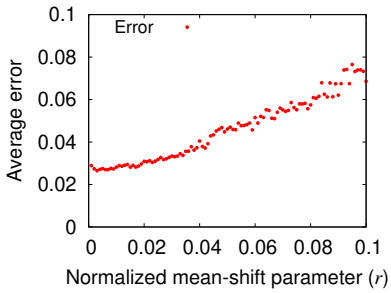

(a) Condition A.




(b) Condition B.

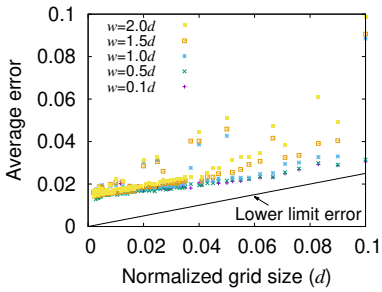



(c) Condition C.

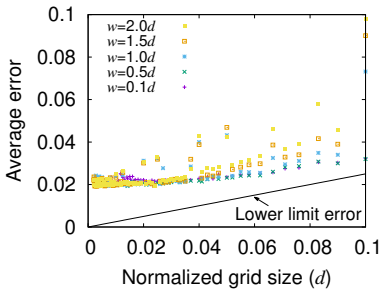



(d) Condition D.

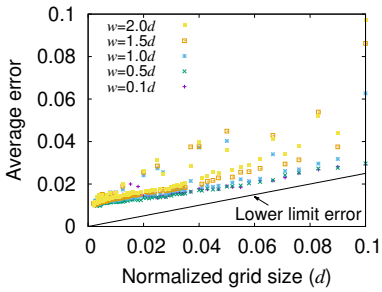
**Fig. 4.11** Average error of estimated position for each database creation (using mean-shift clustering).



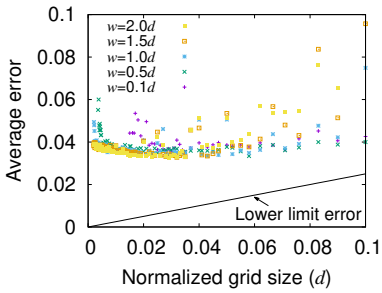

(a) Condition A.




(b) Condition B.

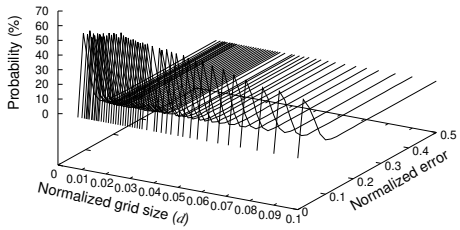



(c) Condition C.

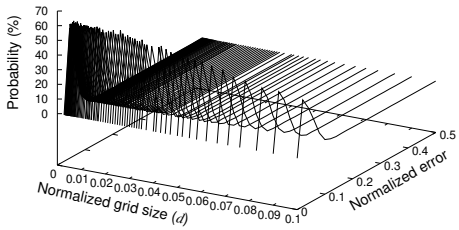



(d) Condition D.

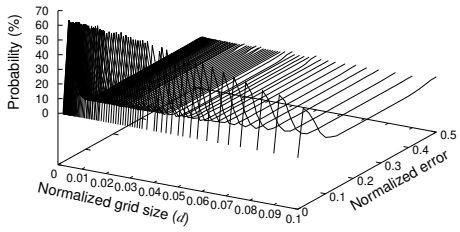
**Fig. 4.12** Average error of estimated position for each database creation (using grid size).



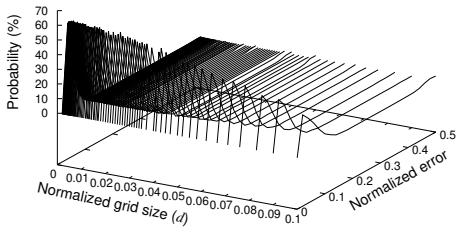

(a) $w = 0.1d$



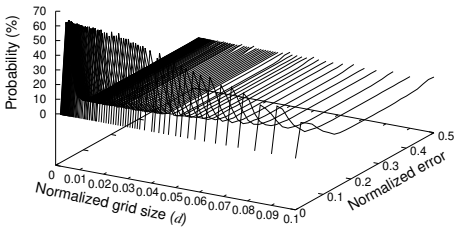

(b) $w = 0.5d$




(c) $w = d$

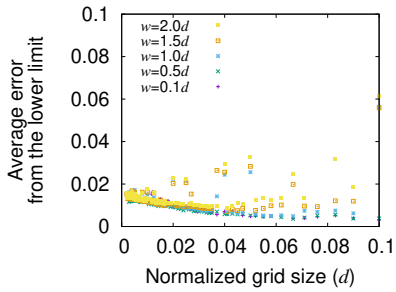



(d) $w = 1.5d$

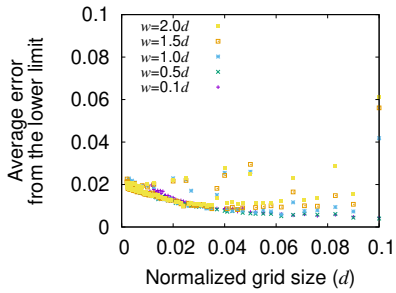



(e) $w = 2d$

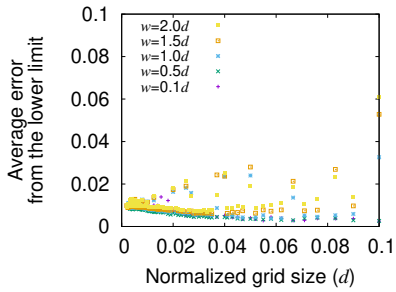
Figure 4.13: Error distribution of estimated position for each database creation.



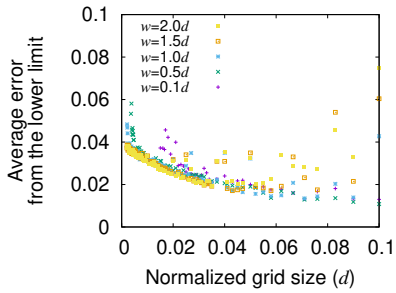

(a) Condition A.




(b) Condition B.

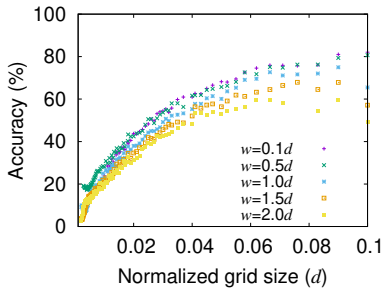



(c) Condition C.

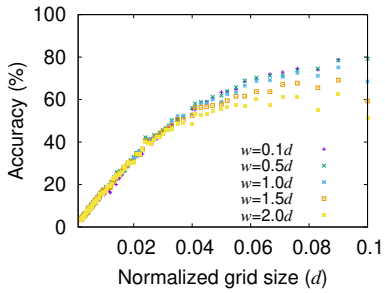



(d) Condition D.

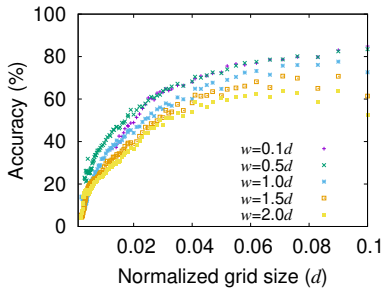
**Fig. 4.14** Average error from the lower limit for each database creation.



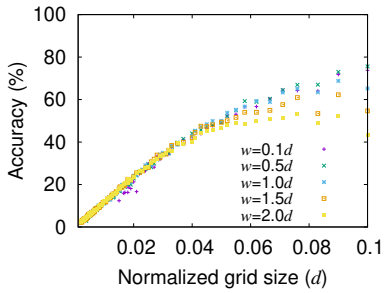

(a) Condition A.




(b) Condition B.




(c) Condition C.




(d) Condition D.

**Fig. 4.15** Accuracy of estimated position for each database creation.

# Chapter 5

# Conclusion

We propose a Wi-Fi-based positioning system using fingerprint method for indoor environment. In the proposed method, the fingerprint database is constructed with estimated reference locations. We managed to reduce the administrative cost to know the actual reference locations. Pairs of MAC-RSSI are gathered using a reference device moving in a constant speed with simple direction to estimate the reference locations. We successfully implemented a positioning system based on Wi-Fi fingerprint method with estimated reference location. The proposed approach is observed that the proposed approach can estimate user's location without any precise reference point locations.

The accuracy is influenced by the number of reference locations stored in the database and the device-specific Wi-Fi sensitivity. New installation and replacements of APs also have a significant impact on the estimation accuracy. Considering these issues, not only the databases creation but also how to manage the database after the database creation is essential. It might be also possible that the user's input data is used for update the fingerprint data on the fly, utilizing the user's locality. We propose a novel merging approach to construct a consistent database from multiple data sets. Since each data set is acquired by moving reference devices at a constant speed, the reference locations in the data set can be estimated using the velocity without any precise reference location information. We successfully investigated two merging methods to improved database management in our proposed system.

The first merging method is the mean-shift clustering algorithm is used to get merged reference locations form the multiple data sets. The number of estimated reference locations and location of each estimated reference locations in merged data set depend on the mean-shift parameter, a radius for the mean calculation to

settle each reference location as the result of clustering. The merged data set is stored in sever as database. The large mean-shift parameter give large error and high computation cost. However, this proposed approach can be reduced the accuracy error due to the estimation accuracy evaluation results that compared with the database are constructed using one data set.

The second merging method involved the number of reference locations and noise reduction in RSSI value which depend on the grid size and the window size. When merging data sets, RSSI values are averaged in the window of each estimated reference location in the new database. We assumed a constant interval named grid size between neighboring reference locations. The number of reference location in merged data set from multiple data sets depends on the grid size. We showed the effectiveness of the proposed approach through the experimental results, including the errors and accuracy for various grid sizes and window sizes. The large grid sizes give high accuracy but large error while the small grid sizes give a low error but low accuracy. This means that, in the proposed merging method, the performance can be controlled, changing the grid size depending on applications. Note that the window size $w$ can be properly set for each grid size $d$ regarding the experimental result. When $d$ is small, a normalized average error of 0.02 corresponding 2.4 m can be archived. Also, when $d$ is large, an accuracy of 80 % can be archived. Comparison of two merging method, updating and maintaining the fingerprint database require high computation cost using mean-shift clustering algorithm, although the mean-shift clustering algorithm is popular in data clustering.

In conclusion, the objective of our research to investigate and develop a Wi-Fi-based indoor positioning using fingerprint method with estimated reference location for cost effective and flexible for any location based applications have been completed successfully.

# Bibliography


[1] S. Panzieri, F. Pascucci, and G. Ulivi, "An outdoor navigation system using GPS and inertial platform," *IEEE/ASME Transactions on Mechatronics*, vol. 7, no. 2, pp. 134–142, Jun. 2002.

[2] N. Jardak and N. Samama, "Indoor positioning based on GPS-repeaters: Performance enhancement using an open code loop architecture," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 347–359, 2009.

[3] G. I. Jee, J. H. Choi, and S. C. Bu, "Indoor positioning using TDOA measurements from switching GPS repeater," in *Proc. The 17th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2004)*, Sep. 2004, pp. 1970–1976.

[4] K. Özsoy, A. Bozkurt, and I. Tekin, "2D indoor positioning system using GPS signals," in *Proc. 2010 International Conference on Indoor Positioning and Indoor Navigation*, 2010, pp. 1–6.

[5] H. Fonseca, A. Nogueira, and P. Salvador, "Localization system for wireless networks," in *Proc. 2013 IEEE International Conference on Communications Workshops (ICC)*, 2013, pp. 988–993.

[6] Y. Chapre, P. Mohapatra, S. Jha, and A. Seneviratne, "Received signal strength indicator and its analysis in a typical WLAN system (short paper)," in *Proc. 38th Annual IEEE Conference on Local Computer Networks*, Oct. 2013, pp. 304–307.

[7] T. Garcia-Valverde, A. Garcia-Sola, H. Hagras, J. A. Dooley, B. Callaghan, and J. A. Botia, "A fuzzy logic-based system for indoor localization using

WiFi in ambient intelligent enviornments," *IEEE Transactions on Fuzzy System*, vol. 21, no. 4, pp. 702–718, Aug. 2013.

[8] M. E. Rusni, M. Al, N. Jami, and M. M. Din, "An improved indoor positioning algorithm based on RSSI-trilateration technique for internet of things (IOT)," in *Proc. 2016 International Conference on Computer and communication Engineering*, Oct. 2016, pp. 72–76.

[9] V. Honkavirta, T. Perälä, S. Ali-Löytty, and R. Piché, "A comparative survey of WLAN location fingerprinting methods," in *Proc. Positioning, Navigation and Communication (WPNC)*, Mar. 2009, pp. 243–251.

[10] B. Molina, E. Olivares, C. E. Palau, and M. Esteve, "A multimodal fingerprint-based indoor positioning system for airports," *IEEE Access*, vol. 6, pp. 10 092–10 106, Jan. 2018.

[11] W. K. Zegeye, S. B. Amsalu, Y. Astatke, and F. Moazzami, "WiFi RSS fingerprinting indoor localization for mobile devices," in *Proc. UEMCON*, Oct. 2016, pp. 1–6.

[12] A. Bahillo, S. Mazuelas, J. Prieto, P. Fernández, R. M. Lorenzo, and E. J. Abril, "Hybrid RSS-RTT localization scheme for wireless networks," in *Proc. 2010 International Conference on Indoor Positioning and Indoor Navigation*, 2010, pp. 1–7.

[13] 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 *Proc. International Conference on Computer and Communication Engineering*, Jul. 2016, pp. 72–76.

[14] D. A. Savochkin, "Simple approach for passive RFID-based trilateration without offline training stage," in *Proc. 2014 IEEE RFID Technology and Applications Conference (RFID-TA)*, Sep. 2014, pp. 159–164.

[15] C. Huang, L. Lee, C. C. Ho, L. Wu, and Z. Lai, "Real-time RFID indoor positioning system based on Kalman-filter drift removal and heron-bilateration location estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 3, pp. 728–739, 2015.

[16] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, May 2016.

[17] H. Jun-Ho and S. Kyungryong, "An indoor location-based control system using bluetooth beacons for IoT systems," *Sensors*, vol. 17, no. 12, 2017.

[18] M. V. Bramha, J. Gan, N. Ghodpage, A. Nawale, and G. Bahe, "Indoor positioning system using magnetic positioning and BLE beacons," *International Research Journal of Engineering and Technology*, vol. 4, no. 3, pp. 1031–1035, Mar. 2017.

[19] H. Murakami, M. Nakamura, S. Yamasaki, H. Hashizume, and M. Sugimoto, "Smartphone localization using active-passive acoustic sensing," in *Proc. 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sep. 2018, pp. 206–212.

[20] K. Nakazawa and T. Oshita, "Position control of vision based robot hand-fusion of 2-D image and discrete range image by multi-agent," in *Proc. 1994 IEEE International Conference on MFI '94. Multisensor Fusion and Integration for Intelligent Systems*, 1994, pp. 353–359.

[21] J. Oh, J. Park, S. Park, and S. Y. Ko, "Image-based guidance system for intravascular microrobot: Fiducial marker-based registration using biplanar fluoroscopic images CTA images," in *Proc. 2015 15th International Conference on Control, Automation and Systems (ICCAS)*, 2015, pp. 919–922.

[22] O. Pribula and J. Fischer, "Real time precise position measurement based on low-cost CMOS image sensor: DSP implementation and sub-pixel measurement precision verification," in *2011 18th International Conference on Systems, Signals and Image Processing*, 2011, pp. 1–4.

[23] Y. Zhuang, Q. Wang, M. Shi, P. Cao, L. Qi, and J. Yang, "Low-power centimeter-level localization for indoor mobile robots based on ensemble Kalman smoother using received signal strength," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6513–6522, Aug. 2019.

[24] S. Ayub, A. Bahraminasab, and B. Honary, "A sensor fusion method for smart phone orientation estimation," in *Proc. 13th Annual Post Graduate*

*Symposium on the Convergence of Telecommunications, Networking and Broadcasting*, Jun. 2012.

[25] D. K. Shaeffer, "MEMS inertial sensors: A tutorial overview," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 100–109, 2013.

[26] H. Carlsson, I. Skog, and J. Jaldén, "On-the-fly geometric calibration of inertial sensor arrays," in *Proc. 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2017, pp. 1–6.

[27] C.-W. Tan and S. Park, "Design of accelerometer-based inertial navigation systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, pp. 2520–2530, 2005.

[28] W. Kang and Y. Han, "SmartPDR: Smartphone-based pedestrian dead reckoning for indoor localization," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2906–2916, 2015.

[29] M. Jin, B. Koo, S. Lee, C. Park, M. J. Lee, and S. Kim, "IMU-assisted nearest neighbor selection for real-time WiFi fingerprinting positioning," in *Proc. 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 745–748.

[30] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, 2005.

[31] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[32] I. Guvenc and C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Communications Surveys Tutorials*, vol. 11, no. 3, pp. 107–124, 2009.

[33] X. Li and K. Pahlavan, "Super-resolution TOA estimation with diversity for indoor geolocation," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 224–234, 2004.

[34] S. Kim and J.-W. Chong, "An efficient TDOA-based localization algorithm without synchronization between base stations," *International Journal of Distributed Sensor Networks*, vol. 11, no. 5, 2014.

[35] N. Podevijn, D. Plets, J. Trogh, L. Martens, P. Suanet, K. Hendrikse, and W. Joseph, "TDoA-based outdoor positioning with tracking algorithm in a public LoRa network," *Wireless Communications and Mobile Computing*, 2018.

[36] D. Trong-Hop and Y. Myungsik, "TDOA-based indoor positioning using visible light," *Photonic Network Communications*, vol. 27, pp. 80–88, Apr. 2014.

[37] D. Pan and Y. Yu, "Indoor position system based on improved TDOA algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 585, pp. 1–5, Aug. 2019.

[38] S. Jung, S. Hann, and C. Park, "TDOA-based optical wireless indoor localization using LED ceiling lamps," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1592–1597, 2011.

[39] G. Fuxiang and Z. Xiaoguo, "Hybrid GPS/TDOA location algorithm in non-line-of-sight environment," in *Proc. 2012 2nd International Conference on Computer Science and Network Technology*, 2012, pp. 2033–2036.

[40] S. Yang, H. Kim, Y. Son, and S. Han, "Three-dimensional visible light indoor localization using AOA and RSS with multiple optical receivers," *Journal of Lightwave Technology*, vol. 32, no. 14, pp. 2480–2485, 2014.

[41] R. B. Ertel and J. H. Reed, "Angle and time of arrival statistics for circular and elliptical scattering models," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1829–1840, 1999.

[42] C. Feng, W. S. A. Au, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 1983–1994, Dec. 2012.

[43] R. Zhang, W. Xia, Z. Jia, and K. Shen, "The indoor localization method based on the integration of RSSI and inertial sensor," in *Proc. Global Conference on Consumer Electronics (GCCE)*, Oct. 2014, pp. 333–336.

[44] Nguyen Dinh-Van, F. Nashashibi, Nguyen Thanh-Huong, and E. Castelli, "Indoor intelligent vehicle localization using WiFi received signal strength indicator," in *Proc. 2017 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, 2017, pp. 33–36.

[45] R. Yasmine and L. Pei, "Indoor fingerprinting algorithm for room level accuracy with dynamic database," in *Proc. International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS)*, Nov. 2016, pp. 113–121.

[46] B. Dong, T. Burgess, H.-B. Neuner, and S. Fercher, "Neural network based radio fingerprint similarity measure," in *Proc. 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sep. 2018, pp. 1–8.

[47] D. V. Le, N. Meratnia, and P. J. M. Havinga, "Unsupervied deep feature learning to reduce the collection of fingerprint for indoor localization using deep belief networks," Sep. 2018, pp. 24–27.

[48] G. Retscher, "Fusion of location fingerprinting and trilateration based on the example of differential Wi-Fi positioning," Sep. 2017, pp. 377–384.

[49] S.-S. Jan, S.-J. Yeh, and Y.-W. Liu, "Received signal strength database interpolation by kriging for a Wi-Fi indoor positioning system," *Sensors (Basel, Switzerland)*, vol. 15, pp. 21 377–21 393, Sep. 2015.

[50] A. M. Hossain and W.-S. Soh, "A survey of calibration-free indoor positioning systems," *The Journal of Computer Communications*, vol. 66, pp. 1–13, Jul. 2015.

[51] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2568–
2599, Apr. 2019.

[52] B. Dawes and K.-W. Chin, "A comparison of deterministic and probabilistic methods for indoor localization," *The Journal of Systems and Software*, vol. 84, no. 12, 2011.

[53] S. Adler, S. Schmitt, K. Wolter, and M. Kyas, "A survey of experimental evaluation in indoor localization research," in *Proc. 2015 International*

*Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015, pp. 1–10.

[54] S. Jeon, J.-P. Jeong, Y.-J. Suh, C. Yu, and D. Han, "Selective AP probing for indoor positioning in a large and AP-dense environment," *The Journal of Network and Computer Applications*, vol. 99, 2017.

[55] T. Gallagher, B. Li, A. G. Dempster, and C. Rizos, "Database updating through user feedback in fingerprint-based Wi-Fi location systems," in *Proc. 2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, 2010, pp. 1–8.

[56] M. A. Youssef, A. Agrawala, and A. Udaya Shankar, "WLAN location determination via clustering and probability distributions," in *Proc. the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).*, 2003, pp. 143–150.

[57] Y. Gu, C. Zhou, A. Wieser, and Z. Zhou, "WiFi based trajectory alignment, calibration and crowdsourced site survey using smart phones and foot-mounted IMUs," in *Proc. 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2017, pp. 1–6.

[58] P. Wilk, J. Karciarz, and J. Swiatek, "Indoor radio map maintenance by automatic annotation of crowdsourced Wi-Fi fingerprints," in *Proc. 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct. 2015, pp. 1–8.

[59] K. Kaji and N. Kawaguchi, "Design and implementation of WiFi indoor localization based on gaussian mixture model and particle filter," in *Proc. 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov. 2012, pp. 1–9.

[60] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

# Copyright notice

1. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "WiFi Fingerprint Based Indoor Positioning Systems Using Estimated Reference Locations," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E103-A, No. 12, Dec. 2020.

2. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "An accuracy evaluation of WiFi based indoor positioning system using estimated reference locations," Proceedings of International Workshop on SmartInfo-Media Systems in Asia (SISA), pp. 321–326, Sept. 2017.

3. Myat Hsu Aung, Hiroshi Tsutsui, Yoshikazu Miyanaga, "Construction and Management of Fingerprint Database with Estimated Reference Locations for WiFi Indoor Positioning Systems," Proceedings of the 23rd Multi-conference on Systemics, Cybernetics and Informatics (WMSCI 2019), Vol. 2, pp. 7–10, Jul. 2019.