On WLAN Fingerprint Indoor Positioning Systems Clustering, and Classification for Enhanced Performance

Al Glehawi

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Electrical and Computer Engineering Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/3698

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
ON WLAN FINGERPRINT INDOOR POSITIONS SYSTEMS-CLUSTERING, AND CLASSIFICATION FOR ENHANCED PERFORMANCE

by

Haider G. AL Glehawi

A thesis submitted to the Graduate College in partial fulfillment of the requirements for the degree of Master of Science in Engineering Electrical and Computer Engineering Western Michigan University August 2018

Thesis Committee:

Dr. Ikhlas M. Abdel-Qader, Ph.D., Chair
Dr. Azim Houshyar, Ph.D.
Dr. Janos L Grantner, Ph.D.
The most economic and affordable IPS are those incorporating existing infrastructure, such as the widely spread Wireless Local Area Network (WLAN). The Received Signal Strength (RSS) fingerprinting-based system is one of the most promising and powerful techniques so far to be used for indoor positioning. However, there are two challenges in using RSS based IPS; the first challenge is the variation of RSS to indoor multipath propagation, and the second is the high number of Access Points (APs) that are deployed in the region of interest. The first issue leads to degradation in the performance of RSS based IPS, while the second makes the computational cost too high and the positioning accuracy not adequate in real time.

WLAN dual bands to create a meaningful clustering structure using K-means and C-means clustering approaches, and the Euclidian distance similarity measure for classification. Also, to mitigate the dynamic nature of the RSS values, averaging consecutive values by using the moving average was implemented. The investigation of using K-means and fuzzy C-means with two AP selection techniques shows that the performance of fuzzy C-means clustering technique is better than that of K-means. The proposed system resulted in a percentile reduction of computational complexity due to the reduced number of AP used while achieving a 0.7m positioning accuracy.
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Ikhlas M. Abdel-Qader, for her guidance during this master thesis project. My thanks also go to Dr. Janos L Grantner and Dr. Azim Houshyar for giving me the opportunity to perform my master thesis project under their supervision. I specially would like to extend my gratitude Mr. Ahmed K. Abed, a PhD Candidate at the Digital Signal and Image Processing Laboratory (DISPLAY), Electrical and Computer Engineering at WMU for serving as my research mentor and for being very generous in both time and knowledge. Lastly but not least, my special thanks and gratitude to my family and my friends for their infinite support during this thesis work and my graduate studies.

Haider G. AL Glehawi
TABLE OF CONTENTS

ACKNOWLEDGMENTS...........................................................................................................ii
LIST OF TABLES...................................................................................................................v
LIST OF FIGURES ..............................................................................................................vi
CHAPTER
1 INTRODUCTION AND OVERVIEW.................................................................................1
  1.1 IPS Motivation..............................................................................................................1
  1.2 Problem Statement......................................................................................................2
  1.3 Approach Contribution...............................................................................................5
  1.4 Thesis Outline............................................................................................................6
2 BACKGROUND AND PERTINENT LITERATURE.........................................................7
  2.1 Wireless Communication Technology.........................................................................7
     2.1.1 Ultrasound Frequency (UF)................................................................................8
     2.1.2 Radio Frequency (RF)........................................................................................9
  2.2 Receive Signal Strength-Based Methods.........................................................................10
  2.3 Measurement Models for Positioning Time Based......................................................14
     2.3.1 Time of Arrival (ToA).........................................................................................15
     2.3.2 Time Difference of Arrival (TDoA).......................................................................16
     2.3.3 Angle of Arrival (AoA).......................................................................................17
  2.4 IPS Estimation Techniques...........................................................................................18
  2.5 Related Works............................................................................................................21
Table of Contents-Continued

CHAPTER

3 IPS DEPLOYMENT CHALLENGES ..............................................................22
  3.1 Localization Domains ..................................................................................22
  3.2 Control Influences .......................................................................................23
  3.3 Uncontrolled Influences ...............................................................................24
    3.3.1 RSS Variance ......................................................................................25
    3.3.2 Number of the Users per Cell ..............................................................25
    3-3-3 Indoor Environment and RSS Attenuation ...........................................26

4 SYSTEM DESIGN AND EXPERIMENT RESULTS .....................................28
  4.1 System Design .............................................................................................29
  4.2 Framework and Implementation .................................................................30
  4.3 Offline Phase (Training Phase) .................................................................33
    4.3.1 Clustering Method ...............................................................................37
    4.3.2 K-Means Clustering RSS-Based K-Means .........................................38
    4.3.3 C-Means (Fuzzy Method) Clustering RSS-Based Model ....................41
  4.4 Online Phase (Localization Phase) ............................................................44
    4.4.1 Sampling Verification Model ................................................................45
    4.4.2 Access Point Selection by Strongest and Random Methods ...............47
  4.5 Small-Scale Area .......................................................................................50

5 CONCLUSION ..................................................................................................53
  5.1 Summary and Contributions .......................................................................53
  5.2 Future Work ...............................................................................................54
LIST OF TABLES

2.1 Models for the Signal Strength-Based. ..............................................................19

4.1 The main error rate comparison between three different scalers.......................51
LIST OF FIGURES

2.1 The standard channels of ISM band. ...............................................................13
2.2 Access Point snapshot capture of VeiwNet scanner .................................................14
2.3 The concept of Time Difference of Arrival using one transmitter ................... 17
2.4 The Angle of Arrival (AoA) position method .....................................................18
3.1 The variation of RSS signal with time. ..............................................................26
4.1 The proposed frame work. .................................................................................28
4.2 Waldo Library at Western Michigan University with APs coverage area. ....29
4.3 Signal Deployment inside the building for the 5 GHz Band. ..............................32
4.4 Signal Deployment inside the building for the 2.4 GHz Band. .........................32
4.5 Reference points and clustering zones for the Waldo Library. .........................35
4.6 Fingerprint Position Estimation Process RSS based. ........................................36
4.7 The RSS fluctuating during the time measuring sample (0°, 90°, 180°, 270°) 36
4.8 The RSS averaging during the time measuring sample ....................................37
4.9 The clustering zones at Waldo Library. ............................................................38
4.10 The K-means clustering with delegation exemplar indicated .........................40
4.11 CDF plots of error K-means clustering model ..................................................41
4.12 The mutual RPs by using the Fuzzy Logic function .......................................42
4.13 Fuzzy Membership Value with RPs clustering distribution ..........................43
List of Figures - Continued

4.14 CDF plots of error by using C-means. .............................................................43
4.15 Estimate location process online and offline phases. .................................44
4.16 The Sampling Verification Model solution online detection. .......................46
4.17 CDF for the K-means vs. Sampling Verification Model ...............................47
4.18 APs variance overall of the area of interest .................................................48
4.19 CDF plots of error comparing Random vs. Strongest ...............................50
4.23 Scale 1x1m RPs small area clustering .......................................................51
4.24 CDF plots of error comparing 1x1m, 2x2m, 4x4m. .....................................52
CHAPTER 1

INTRODUCTION AND OVERVIEW

There is a wide range of techniques proposed for IPS and quite number of applications that needs IPS: ultrasound, Radio Frequency Identification (RFID), wireless sensors nodes, radio signal, and Bluetooth. Many of these technologies are of high quality and provide more accurate position that is used at the indoor environment. On the other hand, extra hardware or changes have to be added to a building’s infrastructure, which means more costs and efforts [3]. However, RSS already available since most of the indoor buildings have the Wi-Fi infrastructure.

1.1 IPS Motivation

Millions of people around the world visit vital areas, such as malls, high buildings, subways, and underground transportation at which the internet connections are lost. The GPS (Global Positioning System) has become an easily available technology, which is more flexible and easy to use in daily life. However, a GPS is still unable to address these drawbacks that could appear in the indoor environment. A GPS does not operate well inside buildings because the Line of Sight (LOS) challenges the connection between the satellites and receivers. This situation restricts the signals in the indoor environment, plus it costs more money if an external system is needed [1].

A wide range of new technological devices that are Wi-Fi-based is commonly used to maintain the connections among people. The internet and Wi-Fi have been taking on the crucial role in the IPS. Usually, every big recreation center is provided with Wi-Fi services, e.g., the malls, shopping centers, and the airports. Recently, the importance of
the localization and environment has been realized, and localization techniques can be used for many purposes, such as tracking objects and persons, statistics and measurement, and advertisement as well.

There are some barriers that might hamper the desired goal of this study. Two steps have to be counted: “where is the object located” and “what is the fastest way to find it?” To solve these problems, two actions were considered. First, a map of the library building provided all of the possible reference points and supplied the IPS. This part will be created at the offline phase. The offline phase is run by using the existing data of the Wi-Fi technology of the Receive Signal Strength (RSS).

At the second step, the IPS uses the online RSS, which had been found on a mobile. This signal will be used later to estimate the location. To determine the position for the user, RSS will be used and is collected at the offline phase and compared with the actual online RSS by using the modification system and fusion algorithms. Then the advanced and accurate results will be presented.

1.2 Problem Statement

Wi-Fi localization that is RSS based is unlicensed frequency and is used at the estimation position procedure. Because the work was done on the indoor environment, RSS has been affected by indoor objects. The IPS is sensitive to signal fading, attenuation, reflection, and environmental changes. There are some fundamental obstacles, e.g., cut off the line of the RSS, multipath fading, and severe fluctuating signal among the receiver, transmitter, and human body effects.
There are two types of these parameters that can affect the accuracy of the IPS measurements. The first type of parameters are controllable, such as walls, furniture, doors, and the access point itself. Then there are uncontrollable parameters, such as persons, moving objects, and a variety of cellphones and devices. These parameters could reduce the system performance and might present a high rate of error. RSS loses part of its energy during traveling time by hitting objects, and it is difficult to match the RSS at the radio map. In that case, computing step will make undetermined position and more difficult access into estimate range [2]. The different rate of the transmitter and receiver signal can have a negative impact on the IPS [3].

A fingerprinting approach is commonly used with an IPS, which means that Euclidian mathematics are determined by computing the minimum distance between the nearest RP and the user. The traditional Euclidian distance cannot offer a full resolution and fix the RSS diversion that comes from the indoor environment phenomenon. The measurements of RSS users might not match the radio map that has been previously built, which could reasonably lead to the wrong position when the online phase is activated in the system [4].

Many of the indoor position techniques have been developed based on the Wi-Fi theory and the RSS models. RSS technique becomes the most dominant because the Wi-Fi option is available on the rest of the users’ cellphones. The IPS uses the radio signal, which is sent from the APs, and estimates the location by using the fingerprinting and clustering methods [5,6]. Without precision and an accurate RSS vector, the measurement location in this study could be inaccurate enough to run through the propagation environment. For the indoor estimation sequence processing, collecting database for
fingerprinting will come first [7]. For the full range indoor space, adding extra clusters makes this model not feasible because of the adverse effects of the indoor environmental behavior and increase of the computing complexity. Whereas, removing a large number from the comparative and selective signals process cannot give an optimal location and render the APs ineffective inside of buildings [8,9].

The Wi-Fi unlicensed frequency was used in this study and faced challenges in an indoor environment, hence, the radio wave is an informative signal. It is necessary to add all of these essential RSSs to the IPS algorithm. In other words, there are a lot of the APs signals that appear one time only and sometimes the APs signals are highly effective when considered in the collocation processes. Thus, it is essential to create a balance in the RSS estimation process between removing all of the unnecessary RSS and selecting the essential one from the radio map. With this work, an indoor environment localization system is proposed from the planning until the side of the implementation. Using a hybrid model does not always work more actively at the IPS and using multiple models at the same time to discriminate and classify can ignore some substantial information. These barriers might cause a miscalculation of the estimation process and a high risk of getting a wrong address. Therefore, we used the multiple models process to solve these matters [10,11]. Any improvements made with the matching rate and the sorting and selection process may improve the system efficiency.
1.3 Approach Contribution

The IPS would be much more efficient if the WLAN dual bands is utilized as previously proposed by Ahmed Abd, [12]. The probabilistic method was used by utilizing a K-means clustering approach to determine the overall minimum error rate of distance; then the benefit was increased by using c-means clustering (Fuzzy Method), which activates the membership value for each cluster. This experimental work may reduce the matching area, which is decreasing the IPS error rate. In using clustering groups, we aim to overcome many drawbacks that could affect the RSS deployment and reduce the time consumption and the effort. A clustering model enhances the estimation position time and relieves the computational complexity.

Building a radio map is one of the main challenges to the IPS. When there is a big data radio grid, finding associated RP locations for estimating the user’s position is difficult. Choosing optimal RSS from the training phase is key for smooth estimation location purposes. Strongest and Random Aps is a filtering model that aims to boost the APs selection process at the online time sequences. Not all the APs provide substation information for finding location steps. On the other hand, ignoring essential data APs from counting selection negatively influences the IPS. Therefore, Strongest & Random Selection models create a balance that may help the IPS choose the more efficient and active APs from the wide range of them.

The Verified Modification Approach, which is considered one of the common mathematical solutions, raises the IPS, especially when at the minimizing the fading and RSS fluctuating effect. When we face a fluctuating RSS at online detection, it’s difficult
to match the RSS online phase to the closest one at the offline phase, and the not actual RSS could lead us to the wrong position. Therefore, we have added another modification model to resolve the fading issue and to improve our system’s performance. The Sampling Verification Model supports the online phase, and, instead of only one sample RSS taken at the online time, results in a more efficient IPS, while reducing RSS variance effects and fading.

1.4 Thesis Outline

Based on the associated problems diagnosed at the IPS, this work will be presented as follows: Chapter 2 highlights background and related works for wireless communication technologies, the common techniques and the algorithms used to in IPS measurements. Additionally, the IPS fingerprint is based on the Wi-Fi, RSS, and the most current models. Chapter 3 presents the main localization challenges and classifies them into two types—controllable and uncontrollable parameters. Also, Chapter 3 discusses the effects of the IPS. Chapter 4 describes the experimental implementation of tools, software, time schedule, and all of the estimation location steps starting from collecting a data radio map up to the step of finding (X,Y) location. This chapter also presents all of the used models and positioning algorithms. Chapter 5 analyzes the final results. Finally, Chapter 6 describes the conclusion and possible future works.
CHAPTER 2

BACKGROUND AND PERTINENT LITERATURE

This chapter gives a brief description about the fundamental operation methods of the IPS and concentrates on the concepts related for this project. In this chapter, the importance of the wireless communication technology for IPS is explained and the different technologies used for positioning estimation are demonstrated. The Receive Signal Strength-based (RSS) Method is discussed and the reasons of efficiency increasing is illustrated in comparison with other models. In addition, this chapter provides a brief discussion about measurement models for positioning time and demonstrates IPS estimation techniques and RSS-based methods. Finally, we will present some related works that discuss possible solutions to the issues encountered in this project.

2.1 Wireless Communication Technology

The mean core of the indoor position is based on the wireless technology, which provides substantial data links required for position destination. The optimal positioning solution with no additional hardware relies on a wireless network field for navigation. Thus, the most efficient solution requires no hardware costs to build a smart grid and feasible system allowance to meet the standard requirements. The IPs can be classified based on the characteristics of the wireless communication and data login to estimate the location. The most likely models will be discussed, including Ultrasound Frequency, Radio Frequency, Bluetooth, and WLAN models.
2.1.1 Ultrasound Frequency (UF)

This technique operates on an ultrasound frequency model, for which it is utilized for sending/receiving data signals for localization purposes. The principle operation of ultrasound is explained as follows: There are two ultrasound sensors transmitting the message in two axes (X, Y). The ultrasound sensor sends the wave signal and checks the traveling time for the sending/receiving process in two axes. The signal between the sensor and the object helps to estimate the location. One of the advantages of this technique is using frequency that cannot be detected by the human body, which might be the right way to avoid the interference problem [12]. This technology works in an indoor environment; UF signals are commonly used together with the Radio Frequency (RF) to earn a full advantage of the IPS. Also, this technique is of low cost compared to other technologies. But there is a fundamental challenge for UF. The challenge is that interference happens at different frequencies, which could affect the IPS performance and efficiency; in addition, extra hardware might be needed to build an area grid [13].

2.1.2 Radio Frequency (RF)

This model is based on the radio frequency band that monitors the objects or the assets. One positive outcome using this approach is that the radio signal can deploy more easily through walls and objects compared with other techniques. The principle operation of the RF is composed of two stations; first is the transmit/receive signal station. Second is the mobile station, which has a microchip. The microchip has an antenna on it called Radio Frequency Identification (RFID). This technique uses medium-range frequency around 1MHz. There’s another model with a full-range frequency band between 2GHz
and 5GHz. The microchip can read and identify using the wave signal reader and should be in contact with the object at the first station.

Thus, extra devices have to be added to the buildings and to the destination objects. This technique is expensive compared to the other models and is usually used for commercial applications [14] [15].

2.1.3 Bluetooth

Bluetooth technology operates in low-band frequency 2.4 GHz. Around 1 Mbps can be transmitted using Bluetooth for a range of 2-15 meters in distance. This model becomes more useful if it is embedded with the majority of the user’s cellphone and devices. Additionally, this type of wireless communication is advantageous in terms of minimum power consumption. It is available in smartphones and mini-equipment [16]. Each Bluetooth device has a unique identity (ID). Thus, this ID can be utilized for location and position purposes. There are many applications that operate with this approach, but Bluetooth still does not require its own components in buildings. This means that more expenses and extra maintenance efforts are required [17]. A new technology called Magnometers has been explored to minimize the distance error. This approach recognizes the unstable signals at the area of interest and reduces the effects of the magnetic propagation. Also, it provides an accurate destination. One of the current projects is conducted using this method, which is called Feldmann. This method uses the triangulation model with proposed accuracy within 2.08 meters; however, this result is only obtained in small areas. In fact, this signal is highly sensitive to obstacles, which leads to more fluctuation in radio signal [18].
2.2 Receive Signal Strength-Based Methods

RSS is considered the most dominantly used at the IPS. Wi-Fi access is available within most of the vital areas. This point is a good support for the indoor localization and identification system [19]. Generally, many recreation centers, such as shopping centers, malls, hospitals, and airports, provide Wi-Fi services. Using IPS based on RSS techniques is more affordable if it is compared to others models. No additional hardware is needed in the indoor environment for the position and location purposes if a Wi-Fi model is used [20]. The standard operation indoor frequency of WLAN is 2.4GHz, and the updated operation radio frequency is 5GHz. WLAN becomes the most popular model used in laptops and most of the cellphones provided with WLAN receivers. One of the positive points is that the coverage area and range of deployment for this model is between 50-100 meters [11]. These features and characteristics make WLAN more feasible and applicable at the IPS [17]. This model will be the basement technology of the current study.

The new Wi-Fi features play a functional role in reducing computational complexity and predicting estimation of position error rate. As previously discussed, there are many indoor parameter burdens that might affect the collecting process, such as multipath fading, redundancy, interference, fading, and reflection [21]. All of these drawbacks might lead to an uncertain position. Therefore, utilizing a new feature with 802.1n dual frequencies might allow for a reduction of the effects of these barriers’ radio propagation on the signal. Admittedly, the indoor environment RSS might also be impacted by other parameters; for example, infrastructure and furniture for the building itself, or even by the diversity of the user’s devices and brand manufacturers of the
cellphones [5,6]. Due to the indoor RSS behavior, the signal is affected by multipath fading and can, therefore, determine the RSS as in the following approach:

\[ P_r(d) = P_t - P_L(d) \]  \(2.1\)

Where:

\(P_r(d):\) represent the received power with estimate distance,

\(P_t:\) is the transmission power,

Where \(P_L(d)\) is the path loss at distance \(d\) as:

\[ P_L(d) = P_L(d_0) + 10\pi \log \frac{d}{d_0} + X_\sigma \]  \(2.2\)

Where:

\(P_L(d_0):\) the path loss for the reference distance \(d_0\),

\(\pi:\) is an exponent rely on the signal deployment at the indoor environment,

\(X_\sigma:\) is a variable representing for blindly applicable & uncertain for the model.

Studying RSS properties and signal behavior statistics helps to understand the mechanism of IPS adaptation and makes users’ requirements optimal [22]. Without precision and an accurate RSS vector, the measurement location in the present study could be inaccurate enough to run through the wave propagation environment. For the indoor estimation sequence processing, collecting database for fingerprinting should come first [7].

At the indoor environment approach, Wi-Fi and the RSS could be influenced by multipath fading and other propagation impacts, e.g., reflection [23]. Fluctuating
phenomenon is another challenge that could be faced by the APs signal, which could happen around the center position. Because of the irregular structure of some buildings, the radio propagation range period negatively influences any application operating in the coverage area, and that will lead to a wrong destination. There are many ways to correct the signal distortion by conducting the refine spectral signal at the measurement zone. Fading is considered another significant drawback of the IPS. There are two types of fading effects—large scale and small scale. At the essential point, the dwindle signal is abstracted by an environmental structure, for instance, walls, doors, and furniture. Usually, this type of RSS changing has a log-normally deploying and it is a way of collecting the path loss as a feature of the distance. At the small-scale fading, RSS is affected by multipath fading; the multipath fading can be analyzed with the Rayleigh distribution model if there is no line of sight (NLOS). If there’s line of sight, it can be modeled by the Rician distribution [24]. In the indoor environment, there is an attenuation in the radio signal that makes the signal weak. To fix this issue, it is recommended that an RSS be chosen that is less effected by the attenuation to obtain the more accurate position. One more advantage considered for the 2.4 GHz is working within three non-overlapping channels and mostly (1,6,11) for the 2.4 GHz standard and 36 for the 5 GHz standard. This feature gives the signal a smoother deployment [9]. Figure 2.1 shows the non-overlapping channels for the 2.4 GHz.
The brand new 802.11n model comes with more available technology that is then utilized toward the IPS. This advantage is called Service Set Identifier (SSID), which can distinguish the signal of the Access Point to the numbers of SSID. Each SSID has its own representative Virtual Local Area Network (VLAN). At the same time, the numbers of the SSID is kept up to five.

When more than five are added, this can exceed the traffic management of the access point and impact the performance of the network overall [25]. This authentication protocol is used at the university campus for this study. Figure 2.2 shows an example of the SSIDs segregation at the Waldo Library Building.
2.3 Measurement Models for Positioning Time Based

The primary purpose of any model estimation position is to find the most accurate destination distance in the least amount of time. Many of these models depend on the distance between the object (person or asset) and the nearest known reference point. Mathematically, the estimation collecting process can be modeled as follows:

Let’s say (S) is the indoor space that is looking for our object from the limited and unlimited divert locations, and these locations give us a number of the reference points (P), in a time set \((t_1, t_2, \ldots)\) and the number of the destination points \((D)\), might come from the (P); we can explicate the following model for the continuation of the number destination per timing:

\[
s = f(D, P)
\]
Where \( \hat{Q}_1, \hat{Q}_2, \hat{Q}_3, \ldots, \hat{Q}_n \), each \( \hat{Q} \) has different properties and belongs to the (S). We can depict \( f \) as an indoor localization function, and it is the map that gives us the number of the points and the real position of these points \( Q_i = (Q_1, Q_2, Q_3, \ldots, Q_n) \); we can collect the error rate with the following model:

\[
E = \frac{1}{n} \sum_{i=1}^{n} |\hat{Q}_i - Q_i|
\]

Destinations are not always maintained invariably with period \( D \), and the indoor localization is not necessary to track the object at the real-time movement. Thus, collecting the average of the positions is considered satisfied for the IPS with a minimum bitrate error.

### 2.3.1 Time of Arrival (ToA)

The other type of IPS classifier is a time-based model or Time of Arrival (ToA). The Wi-Fi signal, as any radio signal deployed inside of the building, is based on the characteristics and the properties of the indoor environment building. Thus, the signal takes time to travel from the source until arriving at the user. The time between the transmitter and receiver is accounted for, and it can be used to estimate the position. With this model, if the target can be reached, the distance can be determined by using the speed of the light equation algorithm, and the estimation distance can be addressed with the equation below:

\[
d = C \times t
\]

Where

\( C \) : the speed of the light \( C = 3 \times 10^8 m/s \)
\( t \): the traveling time

The distance can be collected from multiplying the ToA for the RSS and the radio signal velocity. The attenuation of the emitted signal strength is then determined. This technique is called Laureation estimates \([11]\). Usually, two RPs are sufficient to determine the position, but adding one more RP helps to avoid the obscuring.

There are inaccurate measurements that using the ToA model might bring about; for example, the actual signals measurements could be small compared to the high noise of the measurements accuracy because of the environment-behavior, e.g. the reflection, scattering, or multipath for the RSS. However, this issue for the ToA could be worse if there’s no-LoS IPS or if more accuracy is needed in the indoor environment. The required distance is too short compared with the GPS method.

2.3.2 Time Difference of Arrival (TDoA)

Usually, at the time-based algorithm localization objects, the traveling time is used to estimate the distance. However, sometimes sending time is uncertain, and Time Difference of Arrival (TDoA) is used to address this issue. Thus, the difference in the signal traveling time counted from all receivers to them to collect the distance. The computed time difference ignores the need for the transmission time. An accurate result can be obtained by using a precise measurements device that is synchronized between the receiver and the transmitter. However, in the case of TDoA, synchronization is not required between the transmitter and receiver. Synchronization is only needed between the clocks of the receivers because the estimated location depends on the difference in time/distance \([26]\). Figure 2.3 shows the TDoA model.
2.3.3 Angle of Arrival (AoA)

When there is more than one antenna at the transmitter system, it is not considered the optimum way to find the right direction. Thus, IPS projects have begun to use Angle of Arrival (AoA) as a more appropriate method. With this method, the path of the deployed signal at the multi-access transmitter can be seen. By observing the RSS direction, there’s a difference in time between each individual antenna. Figure 4 shows the estimated angle for the object close to the reference device and (D) is the distance between the two antennae. At the AoA, at least two reference points are needed to find the direction of the target as show as in figure 2.5. This technique does not need to synchronize between the references and the negative thing as the specific device used to
measure the angles [27]. The newest AoA technology is used with a new model called Ubisense. Network sensors with up to 10 users and one-meter transmission can be built by using Time Division Multiple Access (TDMA). This technique can provide a 3D location by receiving signals with only two Ubisensor [28].

![Diagram of Angle of Arrival (AoA) position method](image)

**Figure 2.4 The Angle of Arrival (AoA) position method**

2.4 IPS Estimation Techniques

A solid classification scheme is presented from the perspective of the characteristics and measurement methods. In order to present the review of different approaches, some of them have used the signal properties, which are received by the user cellphone, to collect the estimated location. The relation between the RSS and the distance is exponential. There are some methods that can solve localization distance based on RSS, such as proximity, range-based and fingerprinting methods. Table (1) lists the difference between the types of the RSS-based methods.
Table 2.1: Models for the signal strength-based.

<table>
<thead>
<tr>
<th>Method</th>
<th>Principle Operation</th>
<th>Location</th>
<th>Accuracy</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximity</td>
<td>There are two types of the estimate the location: first, physical contact, such as</td>
<td>No need to provide the</td>
<td>Less accuracy than other techniques, because it produces the estimate</td>
<td>Wi-Fi, Bluetooth,</td>
</tr>
<tr>
<td></td>
<td>touch sensors, capacity sensors. Second, other methods detect by a coverage area</td>
<td>coordinate grid and the</td>
<td>area closest to the AP and it changes due to the distance of the user</td>
<td>RFID,</td>
</tr>
<tr>
<td></td>
<td>network connection, such as Bluetooth and RFID cards</td>
<td>offline phase</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Calibration/Environment Impact</th>
<th>Technology</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range based methods</td>
<td>This method relies on the indoor parameters; such as environment structure and RSS coverage area for the object, the physical structure and properties of the Environment.</td>
<td>Need calibration, any change in the indoor environment might affect the accuracy, and it depends on the method used. This technique is highly affected by noise factors.</td>
<td>Wi-Fi, IR,</td>
<td></td>
</tr>
<tr>
<td>Fingerprinting</td>
<td>This method consists of two phases. The training phase, which has the reference points with all the RSS records at the coverage area and the online phase which compares the mobile user RSS with the adjacent RPs signal</td>
<td>More accurate and the error distance (1.2m - 2m). Adding more RPs will increase the accuracy and reduce the distance error.</td>
<td>Wi-Fi, Bluetooth, and Magnetic field</td>
<td></td>
</tr>
</tbody>
</table>
2.5 Related Works

The main difference between this study and other related works is that the experimental application of a dual-bands Wi-Fi-based localization system is the main focus. Whereas some of the related systems give applicable solutions to the position problem only from a one-sided perspective. The approach in this thesis presents desired models and provides a comprehensive practical study of the IPS. In the following lines, a couple of IPS model example solutions are provided.

Shih-Hau Fang et al. have proposed an IPS model solution to fix severe fluctuation in RSS, and they have presented advanced positioning architecture to mitigate the disturbance effects. Their models enhance the feature of the RSS, which is applied by IEEE 802.11 MAC software. Also, they can model the time domain by a convolution operation in the time domain and then transform to the collective random variable in the algorithm spectrum domain. Then the convolution step becomes a linear and separable operation to support fixing multipath fading at the IPS [3].

Finally, Pei Jiang et al. has proposed a fingerprint localization based on important access points (IAP). This adaptive method is guided to choose the high APs signal and is considered IAP. The highest signal comes from the IPS, selected from the database, then the distance is calculated by using similarity degree of the AP repetition of the fingerprint. This method has developed a matching algorithm that can reduce the range of the fingerprint matching [15].
CHAPTER 3

IPS DEPLOYMENT CHALLENGES

In this chapter, Indoor Localization challenges will be discussed. In this project, the finding indoor locations process has been improved in both time and distance directions. The RSS collected as a samples of stationary user at the same position for long time period to increase the accuracy, but, because of the user is moving, a few of the RSS samples are computed at the same location. The effects of changing parameters will be discussed, as well as the impact on the IPS performance.

3.1 Localization Domains

An IPS is influenced by a number of elements that could affect the system performance and efficiency. Accuracy refers to the distance between the actual position and the estimated one. There are controlled conditions that rely on the indoor environment’s statistical distribution. Coverage connection is playing a major role in the IPS to maintain reaching all the users inside of the building with a minimum error distance as well. While the theories in the current study are implemented inside a buildings environment, there are multiple parameters that might influence the system performance. In this section, two of these elements—controlled and uncontrolled influence—will be discussed in order to evaluate the system and how much these domains impact the IPS.
3.2 Control Influences

At the indoor localization, there are some conditions that have a direct influence on the system operation and controlling it allows researchers to obtain the impact on the IPS. Some primary characteristics will be proposed.

There’s no doubt, it takes more time and effort when more RPs are added. Adding more RPs minimizes the distance error rate, enhances the location estimation process at the area of interest, and gives better location results. In other words, collected reference points as samples of the RSS will be used later for localization procedure steps. This means that more verification samples are available, which makes the model moving closely to the actual point. Building the radio map grid with functional perspective, more efficient responsiveness, and the RPs selection process are appropriate to the real time system.

While working for indoor locations and inside building environments, there are propagation characteristics that have been known to be better for understanding the IPS. Access Point has a significant role to the IPS in serving the connection between the user and his/her area location. Therefore, increasing the number of the APs may improve the system responsiveness and accuracy, widening the coverage signal deployment inside the building. The more APs available at the position target, the more allowance for reference points and to expand the radio map. In this project, the Wi-Fi AP role is investigated in propagation signal deployment by providing three non-overlapping channels with a 2.4 GHz band and one with a 5 GHz. The modern technologies allow each AP to send more than one virtual connection multi SSID [25]. Probability using SSID signal is more trusted and reliable to verify the RSS and avoid the fluctuating issue that might happen
for one of them. On the other hand, adding more APs means more cost and maintenance effort are needed. Reducing the number of the APs is always desirable.

The layout distribution network of this paper is that the IPS has been explicated more efficiently and accurately based on the RSS technique. This system builds a radio map grid by using K-means clustering and the fingerprinting approach to estimate the position. While the Wi-Fi IEEE 802.1, 2.4 GHz band is used, there are a couple of devices that use the same range, such as mobile, microwave, and other electronics. There is one obstacle that the system faces and that causes interference because it’s working within the same RSS scale waves spectrum. The latest research confirms that the human body could influence the straight-line connection between the transmitter AP and the mobile user cell phone [18]. The engineer-designer is advised to divide the hall inside the building to the zones of non-overlapping channels, and they, in turn, distribute the APs. It could be a benefit on both sides. First, the signal deployment propagation can reach all the users; and second, it will be easier to cluster these zones and radio map building.

3.3 Uncontrolled Influences

Another concerning point considered is that the saving stability of the signal cannot be maintained with the IPS. Each RP takes a procedure of sequence steps before being assigned. Thus, the RP might be sensitive to any changes in the internal structure of the indoor environment [29]. Before a radio map can be made, some important things have to be set: the distance among the RPs, channels, band selection, the density and the quiet zone at the experiment field. These properties allow for surveying and making performance simulation with fieldwork. One of the important steps is to take a
concentration of the location and measure the influence characteristics that might affect the IPS. The goal is to provide an efficient system and minimize the effects of these burdens.

3.3.1 RSS Variance

According to Diversity of the Devices, a wide range of the mobile models and brands is considered one that the IPS is tackling in the indoor environment. Wi-Fi technologies are becoming more available inside buildings, so RSS techniques become built-in with most of the mobile devices. The main challenge comes from the different configurations between the mobile user and the network access [30]. This brand diversity makes for an RSS bias and might pass the 25 dBm, which will increase the error rate. Some approaches uses signal strength ratio as a fingerprint called Hyperbolic Location Fingerprinting (HLF) to overcome the variance devices issue. This method saves fingerprinting of the RSS ratio measurements for the numbers of the APs instead of precisely measuring the RSS normal values. This relieves the effects of the heterogeneous devices during the computational estimate process [31]. Spatial Mean Normalization (SMN) is another solution for the different devices. This method presents an IPS more efficiently by cutting the spatial mean of the RSS that exists because of the hardware diversity. In addition, the method offsets the RSS shifting by devices user brands [32].

3.3.2 Number of the Users per Cell

At the Wi-Fi distribution network, there are different interest areas that depend on the number of the users for each cell on the AP. The quality of the radio signal starts
lower than other points. With an exaggerated number of the users per cell, a high density band per cell may be faced [25]. In that case, an override connection might be happening to the AP at this density cell. Some of these characteristics (e.g. speed, duty cycle, radio type, band, signal, and SNR) could be influenced by this situation. AP transmission/receiver is influenced due to the number of the users per cell and that has a direct impact on the IPS.

3-3-3 Indoor Environment and RSS Attenuation

![Figure 3.1 The variation of RSS signal with time.](image)

Wi-Fi signal, as any radio wave signal, could be affected by physical objects, including building structure and floors. The unstable input parameters impair the IPS. It’s recommended to calibrate the RPs of the system due to the changes that could happen to the workspace. The Wi-Fi signal is not straight propagation spreading, and this signal could encounter some objects during the traveling time. When this signal is moving
through, it may hit these obstacles. Moreover, it may either lose or gain some of its power and that might generate a multipath situation. This status happens during the traveling time from the transmitters to the receivers. There are more phenomena that the signal deployment inside the building could face, including scattering, refraction, and interface. These drawbacks create a challenge for our measurements processes. Figure 3.1 shows the mechanism of the RSS, which suddenly goes down to -100 dB during a certain part of the time sample scanning. This is a big concern for all research work on the IPS.

One of the practical ways to measure the system efficiency is a Signal to Noise Ratio (SNR). For inside sites, the noise floor is considered a signal quality assessment. Mobile users and laptops are not designed to collect the floor noise, but a network engineer can log in to the APs and use SNR value to compute the RSS for each particular user.

Regression of the SNR might affect the clients more than other parameters and that could influence both co-channel and neighboring channels as well. Designers tend to reduce this effect by maximizing the circle of the APs coverage area to a certain extent, while it’s still not feasible at the high density of the mobile user's field. In many other studies, the SNR variable depends on the attenuation signal during the traveling time [35, 36]. While the message is becoming weaker by increasing the distance between the user and the source, the researcher uses the SNR to evaluate the system performance. The drawing below shows that the noise rate can change the SNR amount. For example, if RSS is (-45dBm) and the signal with a location that has a noise floor of (-70dBm), then the SNR is (25dBm). Mostly, SNR of 20 dB is preferred to the data network and 25 dB for the voice application [25].
CHAPTER 4

SYSTEM DESIGN AND EXPERIMENT RESULTS

In this chapter, the IPS will be proposed, and all procedure steps for implementation of the study case will be provided for both offline and online phase detection. Also, the main improvements of this system will be discussed, and the unit step results will be displayed. All the tools and devices that were used in this project will be presented. Figure 6 below includes the estimate location procedure steps in the system-built design.

Figure 4.1 The proposed frame work.
4.1 System Design

The test was conducted on the first floor of the main building of Waldo Library / Western Michigan University. Figure 7 below shows the main floor of the building with 15 APs dual bands (2.4GHz and 5GHz) indications. These APs send the main signals from the work site main floor, but 411 radio signals come from different distributed APs in the building, as shown as figure 4.2.

A in-depth site survey was made and helped to build a more functional, efficient system. There are a couple of things that have to be taken into account, which are signal statistics, radio properties, frequency bands, and channels. These elements assisted in the identification of the highly beneficial area of the interest and divided it into clusters and scalar distance to the RPs.

Figure 4.2 Waldo Library at Western Michigan University with APs coverage area.
4.2 Framework and Implementation

This project makes use of the existing WLAN service in the main library building of the Western Michigan University campus in order to verify the approach of the current work. The building consists of four floors, but the density area was chosen on the main first floor. Wi-Fi signal features that have been used in the IPS are provided the details could utilize at the IPS. The time spent collecting the data and constructing the system was 15 days, working 6-10 hours per day. To organize and arrange the References Points (RPs), a density environment area was selected as the work spot in the building. The study provides substantial information that allows monitoring the controllable and uncontrollable parameters to inspect the system performance. These elements present the features needed to develop a good system and verify fingerprinting more accurately. In addition, the elements make the IPS ready for network testing.

This building supplies the internet service with Cisco Series Access Point. This model offers an 802.11n standard that increases the network performance up to six times better than the typical 802.11a/b/g model networks [25]. This brand operates with dual-band frequencies 2.4 GHz & 5 GHz 3500 series. The brand improves the quality of the network detecting of radio frequency. The network that works with this model is able to make an auto adjustment to optimize the wireless coverage area and the signal with less attenuation for 5GHz frequency band [25]. Figures 4.3 and 4.4 show the difference between 2.4 GHz and 5GHz deployment in the indoor environment.

Another significant character for the dual band 3500 series is that the 5GHz frequency band can address the interference that might happen to the 2.4 GHz band.
Other application works may face interference with the same brand, such as door openers and high-density spot of small devices [25]; this is the advantage of 5GHz frequency band. On the other hand, the 5GHz band is propagated with less than 2.4GHz as seen in the Figures 4.3 and 4.4. This property can be utilized for small- and large-scale controlling clusters.

Fieldwork building at the Waldo Library was of a full space and a high-density zone chosen for testing the plan. A Dell laptop with regular Network Interface Card (NIC) was used to read both radio frequencies bands 2.4 GHz and 5 GHz. RSS has been collected in this area using software Wireless Net View v1.72. This program is able to detect and monitor the numbers of the following feature, e.g., SSID, RSS quality, average signal quality, MAC Address, channel, and frequency. In the working site spot, 103 Reference Points with 142 Test Point of 411*80 RSS were placed there for the purpose of maintaining the difference of APs in the area. Two kinds of the RPs metering measurement scale were employed, with 1x1 meters dedicated for the crowded zones and 2x2 meters for the rest.
Figure 4.3 Signal Deployment inside the building for the 5 GHz Band.

Figure 4.4 Signal Deployment inside the building for the 2.4 GHz Band.
4.3 Offline Phase (Training Phase)

In this project, the RSS fingerprinting-based model is employed; this technique is useful in offering both accuracy and acceptance to build the radio grid. In this model, the APs’ locations are not needed, as non-LOS requirements are available; plus the model provides and expresses the accuracy of the superior localization. This leads to remarkable comparison with other methods.

The IPS radio map consists of two phases—offline phase and online phase. For the offline phase, the data of the RSS have been collected at the indoor environment to build the radio map. This data will be used later for discriminating and classifying the purposes at the online phase. The optimal system shows the closest point to the actual position, which represents the error distance. These valuable methods can utilize the data collected in the online phase by using fingerprinting methods to estimate the location with minimum error rate. Finding an accurate position depends on the radio map resolution and efficiency of the matching rate. Increasing the reference points at the training phase enhances the system accurately.

During the data collecting process, the RSS may be affected by indoor interior design. In some cases, human movement causes more RSS deviation. Therefore, each building has a certain signal scale behavior that is different from others. Sometimes, the RPs’ grid has to be calibrated according to the changes of the interior environment inside the building. Labor and time consumption become two of the offline phase’s challenges, especially when there is a big area with a high number of the APs.
During offline procedures, RPs measurements have to be considered for the high density area. Organizing the distribution of RPs in the attractive area supports the IPS in the sorting selection process. Another option is to divide the radio map grid into numbers of clusters and groups, which makes the estimation process more precise and restricts the maximum error rate limit that may come from the neighbor’s clusters. Figure 4.5 shows the radio map point with reference points and clustering zones for two options—2x2 meters and 1x1 meters.

In the offline phase, all the radio signals that come from APs and are counted in the calculations of the current study have been scanned. The reason for that is to take all of the measurement probability of the signal statistic dual bands. The points have already been assigned on the building map, then the RPs are associated with the actual floor. For radio map RPs distribution plan, there are two metric scales: 1x1 meters, which is for the high density area; and 2x2 meters for the rest of the area. For each RP, a signal receiver is being computed from four directions so as to provide redundant RSS, which is essential
Radio map data information is an M×N matrix that transforms the vector of radio map to its equivalent representation in the RSS magnitude. Where M is the number of APs and N is the number of the RPs. Figure 4.6 illustrates RSS-Based Fingerprint Position Estimation Process.

For each reference, four directions are implemented, and the RSS receives different angles (0°, 90°, 180°, 270°). The reason for changing the phase every 20 samples between each direction is to reduce the effects of the RSS behavior inside the building. The behavioral effects do usually come from the objects, obstacles, and even human bodies. The human body generates signal redundancy. Figure 4.7 shows the RSS’s up and
down change during the sample times. To solve this issue, the RSS utilizes the average of all the RPs. Figure 4.8 indicates the refined average RSS used at the metric scale.

Figure 4.6 Fingerprint Position Estimation Process RSS based.

Figure 4.7 The RSS fluctuating during the time measuring sample (0°, 90o,180o, 270o)
4.3.1 Clustering Method

The mean idea for the clustering technique is that the radio map is divided into a number of the groups or clusters. Each one of these groups has to share common features or characteristics based on unique identification, such as AP location, frequency band, RSS, frequency channel, and Media Access Control Address (MAC Address). This technique helps to reduce the time consuming finding estimation process. In some cases, it’s difficult to find the correct cluster when the number of the APs is high. This technique saves more time; hence, the entire radio map does not get browsed since it may take too much time. Only the equivalent cluster is selected from the whole grid [26]. For the full range indoor area space, adding extra clusters is unnecessary and makes the model not feasible. The reason is due to the opposite effects in the indoor environment behavior. Whereas, ignoring a large number of RPs from the selection process may not
quite give an optimal location and make APs ineffective inside the buildings [6,9]. Using clustering groups aims to overcome many drawbacks that could withhold the RSS deployment and reduce the time consumption and effort. A clustering model enhances the estimation position time and relieves the computational complexity. Figure 4.9 shows the clustering technique at the building field work.

![Figure 4.9 The clustering zones at Waldo Library.](image)

### 4.3.2 K-Means Clustering RSS-Based K-Means

Clustering methods are a model designed to divide the existing area in an indoor environment. This model allows splitting the work site into small clusters with each group having a centroid already defined on the offline radio map. Each cluster has one delegate exemplar, which is considered the cluster identifier. For this exemplar, the group or the cluster for the RPs can be reached quickly. In other words, if there are sample
spaces and different buildings, this model can be used by setting one delegate for each building, then it may be moved to another exemplar for the small cluster as well. This leads to a closer RP compared to the actual one. This figure shows the division of the area into small groups, with the centroid then set for each cluster.

Each cluster has a corresponding weight. The cluster weight secures the minimum distance among the centroids of each set of the RPs and for each cluster. Then all the minimum square distance between the exemplar and each RP are calculated. At the online phase, the batch is received by taking rough selection of the RSS then comparing the minimum distance among the user, the closest RPs, and the Euclidean distance. The Euclidean distance is applied to obtain a fairly good result. The following non-linear system equation with the Euclidean equation below expresses the estimation location calculation:

\[
A = \begin{bmatrix}
AP_1, RP_1 & AP_1, RP_2 & \ldots & AP_1, RP_X \\
AP_2, RP_1 & AP_2, RP_2 & \ldots & AP_2, RP_X \\
AP_n, RP_1 & \vdots & \ddots & \vdots \\
AP_n, RP_1 & \vdots & \ddots & AP_n, RP_X 
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
AP_1, TP_1 \\
\vdots \\
AP_n, TP_1
\end{bmatrix}
\]

\[
d = |A - B| = \sqrt{\sum_{i=0}^{n} |A_i - B_i|^2}
\]

Where \{A\}:- Represent the reference points radio map (offline phase)

\{B\}:- Online phase and test point

\(d\):- the (x,y) minimum distance
Figure 4.11 shows the final result for the K-means clustering Cumulative Distribution Function (CDF). The K-means model is run with the radio map built with 143 TPs, 104 RPs, and 411 APs, respectively. The result is around 2 meters at the 83% error rate, which are more feasible preliminary results. RSS signals have a tremendous diversity of features, e.g. SSID, RSS quality, and dual band channels 2.4GHz & 5GHz.

Figure 4.10 The K-means clustering with virtual delegation exemplar indicated at (k=7).
4.3.3 C-Means (Fuzzy Method) Clustering RSS-Based Model

To increase the efficiency of the K-means function, C-means (Fuzzy Method) clustering Fuzzy model is applied to improve the performance of the existing K-means system. The model is different from the regular K-means clusters as shown as in figure 4.12; that is, RPs are shared with a percentage rate called membership value for each cluster. In that case, each RP is represented with the rate per group unlike a K-means 0/1 system. Figure 4.13 below points out the mechanism of dividing the same working area into the small clusters.

While these techniques give more efficient results, there are more calculations of processing steps that will be considered. Those steps are adaptive to gain the minimum distance because each RP is presented on each cluster depending on the membership value. Figure 4.14 shows the CDF plot by using C-means.

Figure 4.11 CDF plots of error K-means clustering model.
The union of the two fuzzy sets

\[ C = 0.5/x1 + 0.6/x2 + 0.3/x3 \]
\[ D = 0.7/x1 + 0.2/x2 + 0.8/x3 \]

is given by

\[ C \cup D = 0.7/x1 + 0.6/x2 + 0.8/x3 \]

Where:
- \( C \): Area of the interest
- \( D \): Intersection area
- \( x \): Reference points
- \( \mu_A \): Membership value

**Figure 4.12** The mutual RPs between the clusters using the Fuzzy Logic function.
Figure 4.13 Fuzzy Membership Value with RPs clustering distribution

Figure 4.14 CDF plots of error by using C-means.
4.4 Online Phase (Localization Phase)

The online phase is the second stage of the IPS measurement process. The real-time RSS is collected at the mobile user. The real-time RSS sends/receives a signal from/to APs at the building Wi-Fi access. This signal has a consistent identity from the name, AP address, channel, and frequency band. These specifications can be used to find the estimate location. Choosing the appropriate model to address the collected data plays a significant role in the estimate position processing. Many theories can help to select the refining signal, which leads to a minimum error rate distance. K-Nearest Neighbor (KNN) and K-means clustering are the most common models dedicated to data solution algorithms using the IPS [6,27].

Choosing formative APs is necessary. The correct APs selection and arrangement of RSS data help to be close to the actual position, increase the system efficiency, and obtain a more accurate location.

Figure 4.15 Estimate location process online and offline phases.
The spatial mean of RSS is considered to be one of the online stage drawbacks. RSS fluctuation comes during the traveling time when the signal hits the obstacles. The system in this study is proposed to process the steps and display all the models and algorithms that are used to design the IPS. Figure 4.15 shows the processing steps as they move from the offline phase to the real-time phase.

4.4.1 Sampling Verification Model

As mentioned in section 3-3-3, the challenge of IPS occurs when the indoor objects and obstacles of the indoor environment allow the RSS to go up or down during the time of travel. The Sampling Verification Model is suggested to resolve the fluctuated RSS at online detection. Figure 5 shows how the RSS can be obtained by affecting and dropping down from -50 dB to -90dB during the scanning time. The defective sample is used at the online phase to measure the location, with a high chance of failure with the wrong position. It’s difficult to match the RSS in an online phase to the closest one at the offline phase. The unreal RSS allows for a high error in distance to occur. Therefore, another modification is added to the model to solve the fading issue and to enhance the IPs. The Sampling Verification Model supports the online phase; instead of taking one sample RSS at the online time, six or more samples of RSS are taken, which confirms the stability of the signal and takes the average of the samples in case the RSS goes down. Figure 4.16 shows the Sampling Verification Model, which presents the modification solution for the up/down RSS at the real-time orientation.
Where:

\[ B = \frac{S_{p1} + S_{p2} + \cdots + S_{pn}}{n} \] 

\( S_{p} \): RSS online sample/sec, \( B \): the average of the Samples/Sec

It is necessary to mention that the Sampling Verification Method spends more time on sending/receiving the sampling per second. While progress can be gained by using this technique, saving time is not as high a priority in the IPS as the functions of the
tracking and navigation systems. A K-means clustering algorithm can’t solve the RSS fluctuation by itself.

Employing the current design, things can be controlled are the number of the samples/secs and the number of the clusters as well. The system will become more efficient in matching the radio map, then will estimate the location more accurately and adaptively. Figure 4.17 shows the improved performance of the Sampling Verification comparing.

![CDF plots of Error](image)

**Figure 4.17** CDF plots of error compare the K-means vs. Sampling Verification Model

4.4.2 Access Point Selection by Strongest and Random Methods

The Wi-Fi signal inside the site building was initially scanned. Significant RSS data were calculated for both radio frequencies 2.4GHz & 5GHz, which indicated more time consumption.
In the last site survey, more than 400 APs signals had been captured. The necessity of adding all of these RSS signals to the IPS algorithm is questionable. In other words, there are many APs signals that appear only one time. Thus, it is essential to reduce the RSS estimation process and cut all the unnecessary RSS of the selection scale radio map.

Figure 4.18 shows the variometer for all of the signals inside of the building. On the other hand, more than 120 unnecessary RSSs come with the overall training phase. The highest RSS values are selected and the smallest ones are ignored. It is recommended to remove RSS small variance values from the APs’ selection process. This technique is useful when a large number of APs exist. Minimizing the numbers of the APs lead the reduction of the RSS coverage area, which is necessary for choosing the formative APs. Figure 23 presents the K-means clustering by using Strongest Method.

![Figure 4.18 APs variance overall of the area of interest](image-url)
There are some drawbacks that may come when this theory is applied, especially for any cluster that has more than one Strongest APs. In the same cluster, many APs monitored the RSS diversion values coming from two or more APs. That may be not realized in the online phase in the same cluster in the same time frame. With a high number of APs, a little change in RSS leads to the wrong destination. No big changes have been recognized in some of the RSS values for the different clusters, which leads to the online phase estimation process facing significant complexity.

The Random Model is presented to address the RSS fluctuation at the closest clusters. Figure 4.19. Shows how the IPS becomes more feasible and functional. Random selection process is highly common to work. Two important considerations when using the Random technique are 1) that the RSS changes over time and is important to balance all of the APs’ signals in the computation process, and 2) this method requires auto-selection for a nonlinear system. RSS selection from the 411 signals comes from the diversity of APs. Figure 4.19expresses the Random selection process gathering K-means and Strongest versus Random Selection

49
4.5 Small-Scale Area

For this work, a testing section was added to confirm the small area resolution. A small area is made as shown in Figure 4.23 below. In this area, a specific plan is conducted for the radio map; it runs 1x1 scale meter for the RPs distribution grid with the following blue area. The result indicates high accuracy of the model fingerprinting for the small-scale area. When environment is divided into very close RPs of no more than 1m, an error rate will be no more than 0.7m as shown in Figure 4.24 and Table 4.1 Table 2 includes the mean error rate comparison between the normal and the small metric scale. However, adding more RPs leads to spending more time and effort and reduces time consumption, which is always requested for the IPS.
Table 4.1 The main error rate comparison between three different scalers.

<table>
<thead>
<tr>
<th></th>
<th>Scale 4x4m</th>
<th>Scale 2x2m</th>
<th>Scale 1x1m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.4368</td>
<td>1.8006</td>
<td>0.7044</td>
</tr>
</tbody>
</table>

Figure 4.20 Scale 1x1m RPs small area clustering
Figure 4.21 CDF plots of error comparing 1x1m, 2x2m, 4x4m.
CHAPTER 5

CONCLUSION

5.1 Summary and Contributions

This thesis work aimed to investigate IPS accuracy of positioning estimates using affordable techniques and while minimize the execution time. The K-means clustering method was applied to divide the overall radio map of the area of interest into a smaller regions, and followed by using C-means clustering (Fuzzy Method) to further divided the RPs clusters depending on the RPs’ membership value and taking advantage of the dual band. This matching area for the online testing was minimized. The use of clustering groups helped to reduce the execution time and overcome many drawbacks that would have been present if the entire radio map was used instead. In addition, this thesis work helped to validate the RSS dependence on signal orientation and how the averaging algorithm may help, and hence, a solution for the RSS fluctuation and attenuation problems has been proposed. Moreover, truly, the essence of this thesis is the approach that was built on taking advantage of the WLAN dual bands for an RSS fingerprint-based IPS and therefore, a more meaningful clustering approach can bring higher positioning accuracy.
5.2 Future Work

IPS system can be improved via several future investigations that include the following:

1. New techniques can be designed to be more practical in the way to collect the RSS data at the offline phase. It should not be as labor intense to reduce time and effort in map updates. That might be done using a smart data acquisition application/device; especially for the vast major area.

2. Design a system that has the ability to use more meaningful utility of the AP dual frequencies 2.4GHz and 5GHz in creating radio map in both small and long distance scales of the RPs.

3. Investigate how can a multi-floor building IPS system can be implemented within wide areas.


